

Nº 53
REV.



UNIVERSIDAD NACIONAL
AUTONOMA DE MEXICO

Facultad de Ciencias

PREDICCIÓN ESPACIAL A TRAVÉS DEL
MÉTODO KRIGING ORDINARIO

T E S I S

Que para obtener el título de:

ACTUARIO

Presenta:

JOAQUÍN PÉREZ BRITO

México, D. F.

1992

FALLA DE ORIGEN



Universidad Nacional
Autónoma de México



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.

PROLOGO

El tema de Geoestadística se puede considerar como una rama de los procesos estocásticos o como una rama de la Estadística que estudia los fenómenos relacionados con las Ciencias de la Tierra. Este tema no se ha difundido adecuadamente en la literatura Estadística en México, ya que poco se ha escrito al respecto, por ello se decidió abordar este, como tema de mi tesis de licenciatura. Aunque fue un poco difícil encontrar bibliografía, después de recopilar algunos artículos, me pareció interesante el hecho de poder interpolar un grupo de datos utilizando la correlación espacial con objeto de predecir valores en localidades del espacio de muestreo.

En general la Geoestadística trata del estudio de fenómenos con correlación espacial en las observaciones. Dicha correlación puede ser con respecto al tiempo o al espacio. El tema de este trabajo es interpolación espacial de fenómenos con correlación espacial.

Por ejemplo: Si se desea saber el volumen de un depósito mineral utilizando como información un grupo de n perforaciones, el objetivo es interpolar el grupo de datos para poder inferir en que forma se distribuye el volumen mineral y cual es su valor estimado. Esta tarea no es trivial, ya que se necesita conocer la estructura de correlación entre los datos para identificar los puntos que influyen más en la estimación.

También se pueden hacer predicciones en un espacio de una dimensión; por ejemplo: si se tiene en el horizonte cierta distribución de diferentes tipos de árboles y se hacen mediciones de una determinada clase de ellos a ciertos intervalos, no necesariamente intervalos iguales, se obtiene un grupo de n observaciones. El objetivo es hacer interpolaciones en localidades nuevas para estimar la cantidad de árboles en dichas localidades.

Un ejemplo de interpolación espacial en cuatro dimensiones, podría ser la medición del monóxido de carbono en cierta localidad a través de interpolar un grupo de n observaciones en el espacio, obtenidas sobre una área determinada durante diferentes intervalos de tiempo.

El problema que dio lugar al origen y desarrollo de la Geoestadística fue, el problema de estimar el volumen mineral de un depósito. Este enfoque se desarrolla en detalle en la sección 1. En general para poder interpolar un grupo de datos, lo que se necesita en primer lugar es: encontrar una medida de correlación entre las observaciones, y en segundo lugar, se utiliza dicha estructura de correlación para interpolar un valor que se quiere conocer en una localidad dada.

El método de interpolación espacial que se va a utilizar es el método Kriging Ordinario, ya que este trabajo es introductorio. El método Kriging es el más utilizado en estudios relacionados con el medio ambiente, y por esta razón, se acerca más al objetivo de realizar una aplicación a un problema real.

A continuación se mencionan a los profesores e investigadores que contribuyeron en la realización de este trabajo.

Quiero dar las gracias al Dr. Ruben Hernández Cid por proponerme este tema de tesis y por asumir la dirección de la misma, así como al Dr. Manuel Mendoza por su valiosa crítica y revisión de éste trabajo, como también al profesor Jose Antonio Florez Diaz.

También, quiero agradecer a la Maestra en Ciencias Eloisa Díaz Francés Murgia por su ayuda y por permitirme bibliografía que fuera necesaria para completar este trabajo.

Finalmente, quiero agradecer al Dr. Marcos Leykoo por permitirme utilizar el equipo de computo del IIE, e incluso por mostrar interés en futuras investigaciones sobre predicción espacial.

INDICE

Introducción

1. Desarrollo Histórico de la Geostatística
 - 1.1 Introducción al método Kriging
 - 1.2 Estimación de la Covarianza
 - 1.3 Insuficiencia de los conceptos estadísticos clásicos
 - 1.4 Noción de Variable Regionalizada.

2. Hipótesis sobre el modelo
 - 2.1 Procesos Estocásticos
 - 2.2 Anisotropía en la Variable Regionalizada
 - 2.3 Estacionaridad
 - 2.3.1 Estacionaridad Fuerte
 - 2.3.2 Estacionaridad de Segundo Orden
 - 2.3.3 Hipótesis Intrínseca

3. Cuantificación de la Interdependencia Espacial
 - 3.1 El Correlograma
 - 3.2 El Variograma
 - 3.2.1 Propiedades del Variograma
 - 3.2.2 El Variograma Experimental en diferentes direcciones
 - 3.2.3 Variograma Isótropo
 - 3.3 Parámetros del Variograma
 - 3.3.1 Efecto Pepita
 - 3.3.2 Techo
 - 3.3.3 Rango
 - 3.4 Dispersión Absoluta o Ley Intrínseca
 - 3.5 Estimación de la Varianza
 - 3.6 Varianza de la dispersión

4. Estimación del variograma
 - 4.1 Estimación del variograma por medio de rezagos
 - 4.2 Ajuste de Modelos Variograma
 - 4.2.1 Variograma Potencia
 - 4.2.2 Variograma Esférico
 - 4.2.3 Variograma Exponencial
 - 4.2.4 Variograma Lineal
 - 4.2.5 Variograma Gaussiano
 - 4.3 Problemas comunes Presentados en Variogramas

- 5. Ajuste del modelo Geoestadístico
 - 5.1 El sesgo y el Modelo Geoestadístico
 - 5.2 El error de Varianza y el Modelo Geoestadístico
 - 5.3 Técnica Lagrange
 - 5.4 Minimización del error de la Varianza
 - 5.5 Validación Cruzada
 - 5.6 Kriging por Bloques

- 6. Aplicación del Método Kriging Ordinario
 - 6.1 Diseño 1 del estudio Geoestadístico
 - 6.2 Análisis Geoestadístico de Plot-1.dat
 - 6.2.1 El Archivo de datos
 - 6.2.2 Estadísticas Univariadas
 - 6.2.3 Análisis del Variograma
 - 6.2.4 Cálculo de los pares muestrales
 - 6.2.5 Parámetros del Variograma
 - 6.2.6 Modelo Variograma
 - 6.2.7 Método Kriging Ordinario
 - 6.2.8 Predicción Espacial
 - 6.2.9 Varianza de la dispersión
 - 6.2.10 Interpretación de los resultados de Plot-1.dat
 - 6.3 Resultados de los archivos Plot-2.dat, Plot-3.dat, Plot-4.dat, y Plot-5.dat
 - 6.4 Diseño 2 del estudio Geoestadístico
 - 6.4.1 Análisis Geoestadístico del archivo level1.dat
 - 6.4.2 Análisis Geoestadístico del archivo level2.dat
 - 6.4.3 Análisis Geoestadístico del archivo level3.dat

Conclusiones

Bibliografía

INTRODUCCION

Usualmente la Estadística supone que las observaciones de un fenómeno en estudio son independientes e idénticamente distribuidas, sin embargo dentro de un contexto de tiempo o espacio dicha suposición no se ajusta con las propiedades físicas del fenómeno, ya que un fenómeno espacial es aquel que exhibe correlación en las observaciones.

En la teoría de series de tiempo se supone que existe una dependencia entre las observaciones separadas a un intervalo de tiempo determinado. En forma similar, un fenómeno con correlación espacial se refiere a que sus observaciones mantienen una estructura de correlación en el espacio.

Por otro lado, la teoría de las Ciencias de la tierra se dedica al estudio de los fenómenos relacionados con el medio ambiente, y trata de dar una respuesta a la incógnita de estimar:

- La correlación entre las muestras que se toman en distintos puntos del espacio.
- El valor de un punto en donde no se obtuvo medición alguna.
- La forma y cantidad del volumen del fenómeno en estudio.

A continuación se da una definición de Geoestadística (Cressie, 1990):

"La Geoestadística es la aplicación del formalismo de las funciones aleatorias al reconocimiento y estimación de ciertas características espaciales de los fenómenos naturales".

Esto quiere decir que no solo se utiliza un modelo aleatorio para describir la propiedad física de variabilidad del fenómeno, sino que se toman en cuenta la correlación entre los datos.

El propósito de este trabajo es introducir el método de predicción espacial Kriging Ordinario y presentar una aplicación utilizando un programa de análisis Geoestadístico.

En la práctica, se obtiene un grupo de datos, sobre una área determinada; este grupo de datos contiene las propiedades físicas del fenómeno espacial. Ejemplos de estos fenómenos podrían ser:

- La distribución de variables meteorológicas; lo que se quiere en este tipo de fenómenos es encontrar la correlación en tiempo y espacio de las diferentes variables. Sin embargo, en este trabajo no se toma en cuenta la correlación entre variables, ya que se trataría de un tema avanzado.
- La distribución de concentraciones de contaminantes en un espacio de cuatro dimensiones (el espacio tridimensional más la dimensión tiempo).
- La distribución de los volúmenes minerales de un depósito en un

espacio de tres dimensiones (Minería).

- La distribución de ciertas especies de árboles en un espacio de dos dimensiones (Silvicultura).

La muestra que se obtiene está formada por un grupo de n datos, donde cada observación $Z(x)$ es considerada como una variable aleatoria, la cual representa el volumen en una localidad x de un fenómeno que tiene la característica de tener un volumen variable en el espacio.

Gráficamente, una variable regionalizada se puede ver como una función $Z(x)$, donde x es la localidad de muestreo en un espacio de dimensión n . Por ejemplo, $Z(x)$ podría representar el volumen mineral que se mide en un depósito centrado en el punto x .

La variable aleatoria es considerada como el resultado de una función aleatoria definida en un espacio determinado. Esto implica la necesidad de utilizar un proceso estocástico.

La Geoestadística considera a un proceso estocástico como un conjunto de variables aleatorias correlacionadas formadas por una función aleatoria, cuyos valores $Z(x)$ se evalúan en diferentes localidades espaciales.

Una vez que se obtenga la muestra, el objetivo es medir que tanto varía una observación con respecto a otra que esté separada a una distancia h . Una forma de hacerlo es a través de calcular la diferencia entre las variables $Z(x)$ y $Z(x+h)$. El grupo de Dichas diferencias $\{Z(x)-Z(x+h)\}$ es un indicador de la variación del volumen del fenómeno.

Una forma de medir la correlación entre los distintos pares $Z(x)$ y $Z(x+h)$ es a través de una función llamada Variograma, denotada por la letra griega $\gamma(\cdot)$, (sección 3.2).

Después de haber calculado el Variograma, se está en la posibilidad de determinar que datos influyen más en la estimación del valor $Z(x_0)$ en un punto dado. Es por eso que antes de implementar la interpolación, se requiere saber la estructura de correlación de los datos.

El punto central del método Kriging se basa en el hecho de que las observaciones del fenómeno espacial mantienen un nivel de correlación alto si los datos se encuentran cerca y un nivel de correlación bajo si los datos se encuentran retirados. De esta forma se puede obtener una estimación de $Z(x_0)$ en una nueva localidad x_0 .

A continuación se presentan algunas disciplinas en donde el método Kriging se ha desarrollado:

En Minería:

Hace 40 años, esta disciplina utilizó una herramienta estadística muy rudimentaria con objeto de tomar una decisión en establecer o no una explotación minera. Originalmente esta decisión fue basada por la simple visualización del mineral expuesto en la superficie por lo que implicaba que habría posibilidad de encontrar algún depósito.

Matheron (1963a) y (1963b) publicó su "Tratado de la Geoestadística aplicada", formado por dos volúmenes, los cuales desarrollan toda una teoría para resolver el aspecto espacial de los problemas que surgieron en minería. En particular el volumen II, Matheron (1963a) trata enteramente del método Kriging. En cambio en el volumen I, Matheron (1963b) habla del origen de la Geoestadística, analiza el Variograma, da técnicas de muestreo y presenta formalmente el método Kriging Ordinario.

En el mismo año, Matheron (1963) publicó un artículo titulado "Principios de la Geoestadística", en donde describe y resume en forma muy general los puntos más importantes de sus dos anteriores volúmenes.

El origen de la predicción espacial se debe a Matheron (1962), quien utilizó un modelo estocástico nombrado "Esquemas de Wijs", aplicado a la minería.

En Meteorología:

Mientras Matheron estaba desarrollando la teoría del Mejor Estimador Lineal Insesgado (MELI) en Francia, el meteorólogo soviético L.S. Gandin estaba logrando notables avances de investigación en el mismo campo. En su libro publicado en 1963 y traducido al inglés en 1965; Gandin llama al variograma función estructural homogénea (Capítulo 2, sección 1), a Kriging simple lo llamó interpolación óptima (capítulo 3, sección 2), a Kriging ordinario lo llamó interpolación óptima a través de pesos normalizados (capítulo 3, sección 5), y a Cokriging simple lo nombró mapeo óptimo de campos (capítulo 5, sección 2).

Los investigadores en meteorología no sólo estaban interesados en interpolación, sino que también, en índices para determinar la circulación de la atmósfera, corrientes totales, y patrones periódicos.

En Fenómenos Forestales:

Matheron (1960) desarrolló una estructura para modelar procesos espaciales, y aplicó estas ideas a grupos de datos tomados espacialmente. Sin embargo, él utilizó como predictor el promedio de las n observaciones:

1.- Cokriging es un interpolador que estima el valor de una variable utilizando la correlación de un grupo de variables.

$$\bar{Z} = \frac{\sum_{i=1}^n Z(x_i)}{n}$$

Aunque este predictor \bar{Z} es lineal e insesgado, no es óptimo. El único caso en que este predictor es óptimo es cuando la correlación espacial no tiene datos aberrantes ni anisotropías.

En Física:

La teoría de turbulencias se basó en contribuciones de N. Kolmogorov 1941. El supone la existencia de un variograma definido en el espacio tridimensional más la dimensión tiempo para caracterizar la estructura de la turbulencia local.

En Geografía:

El interés fue dibujar mapas a partir de datos espaciales, modelando las observaciones de un proceso aleatorio. Esto empezó con Gauss (1809), el cual asumió errores independientes y desarrolló el método de mínimos cuadrados para minimizar el cuadrado de los errores. Este método apareció anteriormente en una publicación de Legendre (1805).

Más tarde esta teoría fue publicada en el artículo de Wiener (1949). Originalmente el objetivo de la Geostatística fue la búsqueda del óptimo en exploración minera. La Geostatística provee una medida exacta de la predicción de valores y utiliza la estimación de la varianza de los errores.

Actualmente se han desarrollado varios métodos de interpolación espacial lineal y no lineal considerados dentro de la familia Kriging, pero éste trabajo sólo se inclina en dar una introducción al método de interpolación espacial "Kriging Ordinario".

A continuación se describen los contenidos de los capítulos:

En el capítulo 1, se da una introducción al tema Kriging Ordinario, como también se da un resumen histórico de las diferentes etapas por las que pasó la Geostatística.

En el capítulo 2, se definen conceptos básicos que se utilizarán en el resto del trabajo y que serán vitales para el entendimiento de la teoría.

1.- Zona anisótropa: Área donde las propiedades físicas del fenómeno cambian drásticamente, ya que existen direcciones para las cuales sus volúmenes aumentan y en otras direcciones sus volúmenes disminuyen.

En el capítulo 3, se habla de la función variograma, la cual es una parte muy importante en cualquier estudio Geoestadístico, como también se mencionan sus principales características.

En el capítulo 4, se presentan varios modelos de variogramas conocidos, los cuales sirven para ajustar al variograma experimental.

En el capítulo 5, se presenta el método de interpolación espacial Kriging Ordinario, y se incluye el método Kriging por Bloques, el cual será utilizado en el capítulo 6.

En el capítulo 6, se desarrolla un ejemplo de análisis Geoestadístico aplicado a un problema referente a Bienes Raíces.

CAPITULO 1

Historia de la Geoestadística

La Geoestadística es la ciencia concerniente con la predicción espacial, a continuación se hará un breve recuento de los principales factores que contribuyeron al desarrollo del método Kriging.

El problema original surgió cuando los mineros y estadísticos trataron de evaluar el volumen mineral en los depósitos, Matheron (1963).

La forma original de estimar el volumen de un depósito mineral fue asignando a cada observación un peso de acuerdo a la cantidad de mineral obtenida. Una vez obtenido el grupo de datos con su correspondiente ponderación, lo que se utilizó fue el Cálculo y la Probabilidad para obtener un valor estimado.

Actualmente, la teoría Geoestadística en lugar de desechar dichas técnicas, las toma como punto de partida para poder desarrollar su propio formalismo matemático hasta alcanzar el nivel científico actual.

Es importante hacer notar que las perforaciones sobre un depósito mineral contienen información esencial que describe las propiedades físicas de los mismos, Noel Cressie (1990), como por ejemplo:

- El volumen mineral en el punto x ; Matheron (1963), no sólo provee información concerniente con la forma y dimensión del volumen mineral, sino también, estudia las anisotropías y las zonas de influencia (áreas con alto contenido mineral). Por esta razón, el volumen mineral de cada muestra contiene información acerca de la dirección hacia donde se desplazan los volúmenes minerales.

El contenido del volumen mineral de cada perforación será diferente de un lugar a otro dependiendo de las propiedades físicas del depósito. Así que al tomar muestras en una área determinada donde sus volúmenes no varían mucho, implica que existe una relación de dependencia entre las muestras. En términos Geoestadísticos se dirá que existe una continuidad en la distribución del mineral en esta zona.

Por ejemplo, en un mismo depósito se pueden obtener muestras formadas por medio de pepitas minerales, debido a la granulación mineral de algunas áreas del depósito, o se pueden extraer muestras con un nivel alto de volumen mineral.

Esta propiedad de granulación de los depósitos minerales generalmente forma una discontinuidad del volumen mineral con respecto a las muestras cercanas. Además, es difícil determinar el volumen en áreas de granulación mineral, ya que se pueden encontrar perforaciones que contengan cierto volumen mineral, cerca de observaciones con casi o virtualmente no volumen mineral. Este hecho de incertidumbre, donde no se puede asegurar el volumen mineral, es interpretado por uno de los parámetros del variograma el cual es llamado precisamente efecto pepita.

El segundo cambio apareció cuando la probabilidad y el cálculo clásico fueron insuficientes. Por tal motivo, fue necesario desarrollar nuevos métodos para introducir las características espaciales de los fenómenos espaciales, que son simplemente propiedades físicas.

Fue entonces como la Geoestadística empezó a desarrollar su propio formalismo matemático y su propio lenguaje.

Originalmente, la teoría Geoestadística no consideró a la variabilidad del fenómeno, ni a la dependencia entre las observaciones. Por esta razón, se dice que la Geoestadística no tomó en cuenta a los resultados aleatorios de las observaciones como una expresión física del fenómeno en estudio, Matheron (1963).

De aquí en adelante la Geoestadística va a considerar una de las propiedades más importantes de los fenómenos espaciales, la cual es la variabilidad del fenómeno.

1.1 LOS ORIGENES DE KRIGING

En esta sección se trata de explicar como surgió el término "Kriging" y como se fue involucrando con la predicción espacial. Originalmente, la Geoestadística se preocupó por resolver el problema de encontrar el contenido total de un depósito mineral (volumen de tierra y roca) que podría ser minado si este fuera suficientemente rico.

Las ideas fueron desarrolladas por D.G. Krige, (Ord, 1990), en las décadas de los 50's y 60's. Sin embargo, no fue hasta 1963 cuando Matheron introdujo un modelo de interpolación lineal que él llamó "predicción espacial por el método Kriging".

Sin embargo, existe otro enfoque de la palabra Kriging como: "El procedimiento de regresión Múltiple para llegar al mejor predictor lineal insesgado de un volumen o de un bloque mineral de cualquier tamaño", (Cressie Noel, 1990).

La predicción espacial se hace posible a través de asignar un grupo de pesos para todos los datos disponibles dentro y fuera de un bloque mineral. Además, estos pesos sirven para interpolar valores en un bloque fijo, e incluso de pesos apropiados para extrapolar datos que estén afuera del bloque.

No hay duda que Matheron trajo la palabra "Kriging" dentro de la terminología Geoestadística Anglosajona, ya que el término original es francés "Krigeage". Este término fue registrado por el Francés Pierre Carlier. Matheron describe esto en una carta dirigida a Noel Cressie fechada el 12 de abril de 1989. Aparentemente este término fue usado en Francia entre 1958 y 1959 por la comisión de energía atómica Francesa; lo anterior lo describe en una carta dirigida a Matheron en 1962 por F. Blondel, (Cressie Noel, 1990).

1.2 INTRODUCCION AL METODO KRIGING

En esta sección se presentan las bases del método de predicción espacial Kriging Ordinario.

Supóngase que se desea estimar el volumen promedio, denotado por Z_v sobre una área V , del contenido total de un depósito mineral. Dicho volumen promedio se expresa de la siguiente forma:

$$Z_v = \frac{1}{V} \int_V Z(x) dx \quad (1.1)$$

donde, $Z(x)$ es una función aleatoria, la cual describe el contenido mineral en la localidad con dirección y longitud del vector x . Los datos que se utilizan en la estimación de Z_v son las n perforaciones $Z(x_i)$, tomadas en las localidades x_i con $i=1, \dots, n$.

Aquí se supone que el valor esperado $Z(x_i) = \mu$, es constante; el cual representa el valor promedio de los datos $Z(x_i)$, para todas las localidades x_i y además se supone que la estructura de covarianza entre los datos es conocida (Cressie Noel, 1989).

El objetivo consiste en encontrar un grupo de pesos $(\lambda_i | i=1, \dots, n)$ de tal forma que la varianza del estimador lineal $\hat{Z}_v(x) = \sum \lambda_i Z_v(x_i)$ sea mínima bajo la condición de que $\sum \lambda_i = 1$. } (1.2)

El método Kriging estima el valor de Z_v usando los datos que se encuentran localizados en el área V , más adelante se verá que el valor estimado Z_v da pesos apropiados para estimaciones fuera del área V (extrapolación), pero requiere de una estructura de covarianza detallada (Cressie, 1989).

La importancia de la covarianza se debe a que provee un valor de dependencia entre grupos de datos separados a cierta distancia, de esta forma es posible interpretar la propiedad intrínseca, en donde las muestras cercanas se parecen y las muestras alejadas no se parecen (Journel y Huijbregts, 1978).

1.3 ESTIMACION DE LA COVARIANZA

Quando se habla de la covarianza entre las observaciones se está haciendo referencia al grado de dependencia entre el grupo de datos disponibles. Con objeto de estimar la covarianza entre las observaciones, éstas se agruparon de acuerdo a la distancia de separación entre ellas. Por lo que se calcularon las distancias de separación de cada muestra contra todas las restantes para formar varios grupos de parejas muestrales. Cada grupo es llamado rezago y una vez encontrados los distintos rezagos se puede proceder a evaluar la covarianza entre ellos (Cressie Noel, 1989).

La covarianza también depende de las anisotropías del fenómeno. Es decir que la covarianza entre grupos de datos no solo depende de la distancia entre ellos sino también de la dirección del vector de separación. En este trabajo solamente se consideran fenómenos cuya covarianza depende de la distancia y no de la dirección, éste es un ejemplo de un fenómeno isotrópico.

A continuación se muestra la covarianza entre las observaciones $Z(x_1)$ y $Z(x_2)$ separadas a una distancia d_{ij} :

$$C_r = \frac{1}{N_r} \sum_{i,j} [Z(x_1) - \bar{\mu}] [Z(x_2) - \bar{\mu}] \quad (1.3)$$

Donde, C_r es la covarianza del rezago r . N_r es el número de parejas muestrales que están separadas a una distancia d_{ij} con una cierta tolerancia. La tolerancia se refiere a que se da una amplitud $d_{ij} \pm c$, para incluir a más muestras en el intervalo $[d_{ij} - c, d_{ij} + c]$.

El conjunto N_r es el número de parejas muestrales que están en el rezago r , las cuales están formadas por el conjunto:

$$\{ (x_1, x_2) \mid x_1 - x_2 = d_{ij} \}$$

Debido a que es muy difícil encontrar muestras separadas a una distancia fija o exacta, se necesita definir un nivel de tolerancia c .

En el siguiente ejemplo, N_r es el conjunto de parejas separadas d_{ij} unidades, con una tolerancia de $\pm c$:

$$N_r = \{ (x_1, x_2) \mid x_1 - x_2 = d_{ij} \in [r+c, r-c] \}$$

donde, r es la distancia para la cual se eligió el rezago y $x_1, x_2 \in D$.

El valor $\bar{\mu}$ es el estimador por mínimos cuadrados de la función aleatoria $Z(x)$ (i.e. $E[Z(x)] = \mu$), y la suma se toma sobre las N_r parejas de localidades x_1, x_2 que están separadas a una distancia d_{ij} .

Con objeto de estimar la covarianza, el grupo de parejas muestrales se dividió en subgrupos, en donde cada subgrupo da un valor de correlación.

1.- Un fenómeno isotrópico se refiere a que las propiedades físicas de un cuerpo son las mismas para cualquier dirección.

Supóngase que se quiere determinar la distribución del volumen mineral de un depósito. Los resultados que se obtengan tendrán una aplicación general para cualquier fenómeno con caracteres espaciales.

Desde un enfoque estadístico clásico los volúmenes muestrales de un depósito pueden ser clasificados en un histograma. Pero este procedimiento no toma en cuenta la localización de dichas muestras dentro del depósito. No es suficiente considerar la frecuencia de aparición de un volumen mineral dado, sino también en qué sentido los diferentes volúmenes se desplazan. Además, el objetivo es descubrir cual es la posición y el tamaño de los bloques minerales, Matheron (1963).

El primer problema que se presenta es la inhabilidad de la Estadística clásica de tomar en cuenta el aspecto espacial en sus estudios, el cual es precisamente la característica más importante de los fenómenos con correlación espacial. A continuación se dará un ejemplo del punto de vista de la probabilidad y el cálculo clásico sobre el concepto que se tenía de las variables aleatorias.

Supóngase que se lanza una moneda en donde se registra +1 si cae sol y -1 si cae águila. Antes de lanzar la moneda lo único que se sabe es que se tiene una posibilidad de dos de obtener, ya sea águila o sol. Ahora bien, desde un punto de vista clásico, una variable aleatoria tiene dos propiedades:

Propiedad 1) Teóricamente, existe la posibilidad de repetir el evento de lanzar una moneda tantas veces como se quiera para asignarle un valor numérico a la probabilidad del evento.

Propiedad 2) Se supone que existe independencia entre cada suceso; ya que si en los primeros 100 intentos se obtiene sol, esto no implica que el 101 intento va a ser sol. En cada suceso de lanzar una moneda, siempre habrá un resultado de dos de obtener sol o águila. Es decir, los eventos son independientes.

En cambio, desde un punto de vista Geoestadístico, el volumen mineral de una muestra no puede tener estas dos propiedades. El volumen del contenido de una observación en determinada localidad es única y solamente se puede extraer una vez. Ahora bien, supóngase que se toman varias muestras en un bloque bien determinado. Se sabe que si se obtiene la primer muestra su valor es único. Sin embargo supóngase que se obtienen varias muestras cerca de la primera. Los volúmenes minerales de estas últimas observaciones muestran una posibilidad de repetir el suceso pero con un aparente cambio. De hecho el resultado no es el mismo pero, si existe cierta similitud dependiendo si el fenómeno presenta o no anisotropías.

La segunda propiedad tampoco se cumple, ya que dos muestras vecinas no son independientes. La practica nos dice que si se realizan dos perforaciones adyacentes, en donde en una de ellas se obtiene un nivel

de mineral alto y denso, entonces es probable que la otra perforación muestre casi las mismas propiedades. Esto quiere decir, que el bloque donde fueron extraídas estas perforaciones contienen un nivel de mineral alto. Esta tendencia expresa el menor o mayor nivel de continuidad en la variación de los volúmenes minerales de un depósito.

En la década de los 50 existía la creencia de que al obtener muchas observaciones en una área determinada se obtendría una estimación del valor de $Z(x)$, más acercada a la realidad. Ya que se pensaban que la multiplicidad del número de muestras implicaría mejores estimaciones. Esta mala interpretación de la teoría llevó a una serie de resultados absurdos. Por ejemplo, si se tenía un depósito el cual se iba a explorar por medio de perforaciones, pensaban que sería mejor tomar las muestras en localidades separadas a 5 mm en lugar de 50 cm para obtener 100 veces más muestras y por lo tanto, 10 veces más precisión. Esto es incorrecto, ya que la multiplicidad en el número de muestras no necesariamente implica mejores estimaciones.

Lo único que se estaría haciendo es repetir tantas veces como se quiera, el mismo o casi el mismo valor, sin aportar más que tiempo y dinero en evaluaciones adicionales.

1.5 NOCIÓN DE VARIABLE REGIONALIZADA

En la práctica, frecuentemente se encuentran fenómenos naturales que pueden ser caracterizados por la distribución en el espacio de ciertas cantidades medibles llamadas variables regionalizadas.

La variable regionalizada es una función que explica el aspecto espacial del fenómeno. Dicha función toma un valor determinado en cada región en el espacio. Por esta razón, la variable regionalizada representa un valor local y toma valores diferentes a lo largo del espacio de estudio. Las variables regionalizadas ocurren en el espacio mineral del depósito y llevan asignadas características físicas importantes, las cuales se presentan a continuación (Matheron, 1963):

a) El depósito mineral es llamado campo geométrico de regiones. La función aleatoria $Z(x)$ está definida sobre el volumen de la muestra, la cual es llamada soporte geométrico. Dicho volumen representa el grueso del depósito en ese punto. Comparando los volúmenes muestrales en diferentes áreas, se pueden detectar las direcciones en las cuales los volúmenes de las muestras aumentan y hacia que direcciones disminuyen. A las áreas donde existen variaciones del volumen mineral para una dirección con respecto a otra dirección, se le llamará zonas anisótropas. Estudiando el comportamiento de este fenómeno se puede identificar hacia donde se desplazan los volúmenes minerales altos y de igual forma se pueden detectar áreas isotrópicas, zonas de influencia, áreas de granulación mineral y bloques minerales puros.

b) Supongase que se tienen dos vecindades inmediatas sobre un campo geométrico. Divídase este depósito por regiones; tómese una muestra en cada región. Estas observaciones muestran un cierto nivel de continuidad en su variación espacial, debido a que el volumen de cada

observación muestra un patrón de dependencia con respecto a las muestras de localidades vecinas.

Algunas variables regionalizadas pueden mostrar una relación de dependencia muy fuerte en los volúmenes muestrales, a esta relación se le llamará *continuidad estricta*; sin embargo, habrá casos en los cuales los volúmenes de las observaciones muestren una incertidumbre, la cual se llamará *continuidad en promedio*, Matheron (1963).

En algunas circunstancias la continuidad en promedio no existe y entonces se hablará de un *efecto pepita*, Matheron (1963), el cual está formado por una información pobre del contenido mineral registrado por las muestras. Esta información deficiente mostrada en las observaciones, es debida a que en las áreas donde se practicó el muestreo se encontraron pepitas minerales en lugar de un volumen mineral bien definido. En resumen cuando no se puede confirmar la continuidad se dice que existe un efecto pepita.

c) La variable regionalizada también puede mostrar diferentes tipos de anisotropías. Posiblemente existan direcciones para las cuales los volúmenes minerales no varían mucho en una dirección mientras que varían rápidamente en otra dirección. Por ejemplo, se puede encontrar anisotropías verticales, horizontales y laterales, Matheron (1963).

d) Matemáticamente, una variable regionalizada es una función $z(x)$ de un punto con coordenadas x (en un espacio de dimensión n); el punto x representa las n coordenadas u_1, u_2, \dots, u_n . Por ejemplo, $z(x)$ puede representar el volumen mineral de un depósito centrado en el punto x . La característica de las variables regionalizadas en las Ciencias de la Tierra es que su variación en el espacio es local. Esto excluye la posibilidad de utilizar un estudio determinístico sobre la función $z(x)$. La alternativa que se tiene es considerar a la medida $z(x)$ como una realización de una variable aleatoria $Z(x)$.

e) La Geoestadística interpreta a la variable regionalizada $z(x)$ como la obtención de una función aleatoria $Z(x)$, donde x toma cualquier punto en el dominio de localidades. Una función aleatoria se puede ver como un grupo de variables aleatorias correlacionadas.

De igual forma si el fenómeno en estudio está caracterizado por la distribución conjunta de un grupo de variables regionalizadas $z_k(x)$, $k = 1, \dots, K$; el enfoque geoestadístico es interpretar a estas variables como una realización particular de k funciones aleatorias correlacionadas $\{Z_k(x); k=1, \dots, K\}$. Todos éstos diferentes aspectos concernientes a las distribuciones espaciales de las variables regionalizadas deben ser consideradas por la Geoestadística. Y ésto es posible gracias al concepto matemático expresado por el Variograma.

-
- 1.- La continuidad estricta se refiere a que los volúmenes muestrales de un fenómeno en estudio muestran una correlación alta.
 - 2.- Continuidad en promedio se refiere a que los volúmenes muestrales de un fenómeno en estudio muestran una correlación no muy alta.
 - 3.- Efecto Pepita es un parámetro del variograma e indica el grado de incertidumbre en la continuidad del volumen de fenómeno.

CAPITULO 2

Conceptos fundamentales de la Geoestadística

2 HIPOTESIS SOBRE EL MODELO GEOESTADISTICO

En este capítulo se necesitan algunas conceptos fundamentales para que el método de predicción espacial se pueda utilizar.

Primero, se le pide a los datos que formen parte de un proceso estocástico. Es decir, que la variable regionalizada $Z(x)$ tome sus valores aleatoriamente conforme se desplaza de un punto a otro, en el espacio de localidades.

La otra condición se refiere a que los valores de los volúmenes de las observaciones sean estacionarias, es decir, que las variaciones de los volúmenes del depósito mineral sean suaves o que no varíen mucho a lo largo de todo el depósito.

Otra condición, es suponer que el fenómeno en estudio sea isotrópico. La isotropía se refiere a que las propiedades físicas del depósito mineral son las mismas a lo largo de todos sus ejes.

En la práctica es muy difícil encontrar fenómenos físicos que sean isotrópicos. En los depósitos minerales es muy frecuente encontrar lugares en donde las propiedades físicas del fenómeno en estudio sean diferentes de una dirección a otra, es decir el depósito contendría áreas anisótropas. La anisotropía del fenómeno provoca que la variabilidad en el volumen mineral disminuya hacia determinada dirección y aumente hacia otra dirección; éste hecho es reflejado en los volúmenes de las observaciones.

Los modelos estocásticos se aplican a cualquier fenómeno o sistema que comprenda variabilidad al azar en el transcurso del tiempo o del espacio.

Para considerar estas propiedades se construye un modelo que considere a los datos como parte de un proceso estocástico, donde los datos $Z(x)$'s son observaciones en localidades espaciales x que varían continuamente sobre el espacio de muestreo D , donde $D \subset R^2$ o R^3 .

2.1 PROCESO ESTOCASTICO

La palabra estocástico es sinónimo de aleatorio. "Un proceso estocástico es un sistema que se desarrolla en el tiempo o espacio mientras pasapor fluctuaciones al azar" (Journel y Huijbregts, 1978). También un proceso estocástico se puede ver como una familia de variables aleatorias $\{Z_t\}$, correlacionadas, donde t es un punto en el espacio T llamado espacio parametral. Además, para cada $t \in T$, Z_t es un punto en un espacio S llamado espacio de estados (Rodney Coleman, 1987).

Supóngase que se tiene una familia $\{Z_t\}$, la cual se interpreta como una trayectoria de una partícula que se mueve "al azar" en el espacio S , siendo Z_t su posición en el instante t . Un registro de una de éstas trayectorias se conoce como realización de un proceso estocástico.

El objetivo es considerar a un proceso estocástico que sea estacionario, ya que de ésta forma se puede calcular el valor esperado y la varianza de las variables regionalizadas $Z(x)$, describiendo mejor al fenómeno espacial (Statistics Encyclopedia, 1989).

Se puede describir un sistema, en donde las observaciones formen parte de una familia de variables aleatorias $\{Z(x) \mid x \in D\}$, donde $Z(x)$ mide en la localidad x el valor de la función aleatoria Z .

Sean $\{x_i\}$ las localidades espaciales las cuales varían continuamente sobre el espacio $D \subset R^2$ o R^3 . A continuación se presenta un modelo Geostatístico, Cressie (1988), el cual aborda una clase particular de problemas, tales que los datos se pueden modelar de acuerdo a un proceso estocástico.

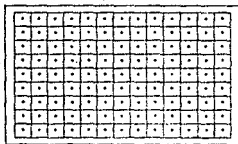
$$\{ Z(x) : x \in D \} \text{ con } D \subset R^2 \text{ o } R^3. \quad (1)$$

Los datos $Z(x)$ son observaciones en las localidades espaciales conocidas $\{x_i : i = 1, \dots, n\}$.

Como se sabe una función aleatoria está definida por el grupo de funciones de distribución conjunta acumuladas; sin embargo, lo que se necesita es asegurar la existencia del valor esperado y la varianza de $Z(x)$. Es decir, se necesita que los dos primeros momentos existan y no dependan de la localidad en donde se evalúa la variable regionalizada (sección 2.3).

La muestra está formada por n observaciones $\{Z(x_i)\}$, evaluadas $x_i \in D$ en una área D . Donde las observaciones se pueden obtener sobre localidades separadas a una misma distancia, a este tipo de muestreo se le llamará muestreo regular. Entonces, utilizando el muestreo regular se obtuvieron las muestras ya sea en el centro de cada cuadro o en la intersección de las líneas horizontales con las verticales (Fig. 9).

Entonces, el área D es dividido por regiones cuadrículas en donde se obtienen las muestras en el centro de cada región a diferencia del muestreo regular estratificado, Richard McArthur (1987), en donde generalmente se toma un promedio de varias muestras de cada área. De cualquier forma la variable regionalizada toma el valor correspondiente en la localidad elegida.



Esta es el área de muestreo D en donde se obtienen las observaciones en el centro de cada sección. A este tipo de muestreo se le llama muestreo regular, donde cada región no implica que sea un promedio.

Figura 9.

2.2 ANISOTROPIA EN LA VARIABLE REGIONALIZADA

En la sección anterior se habló de la variable regionalizada y se vió como cambia de valor conforme se desplaza de un punto muestral a otro. Dicho cambio es originado por la propiedad física de variabilidad del fenómeno.

Se necesita imponer una condición de estacionaridad a la función aleatoria con objeto de suponer que dichas anisotropías tengan un nivel de suavidad (Myers, 1989).

2.3 ESTACIONARIDAD

En la teoría de procesos estocásticos frecuentemente se encuentran procesos $Z(x)$ que tienen distribuciones en equilibrio cuando $t \rightarrow \infty$, en el sentido de que éstas distribuciones dependen del parámetro t (Rodney Coleman, 1967). Lo anterior implica que el proceso es estacionario en el tiempo t .

En la práctica de la Geoestadística, la estacionaridad en una forma u otra es una característica esencial de la función aleatoria, la cual sólo se aplica a la función aleatoria y no a los datos (Myers, 1989).

Un proceso estacionario tiene la propiedad de que la distribución conjunta de $Z(t)$, $Z(t+h)$ es una función que no depende del tiempo t , sino de h (Warrich y Myers, 1986).

En particular, el proceso es estacionario fuerte si la función aleatoria no sufre variación alguna ante cualquier traslación. Esto quiere decir que si se evalúa un grupo de variables aleatorias en ciertas localidades (x_i) , la función de distribución conjunta es la misma que si se evaluara el mismo grupo de datos en localidades trasladadas (x_i+h) . En la práctica existen algunas dificultades

analíticas con los datos al utilizar esta definición, ya que en la realidad esto sería equivalente a suponer que los valores de las variables regionalizadas son iguales a lo largo de todo el espacio muestral, lo cual no es cierto (Myers, 1987).

Por esa razón se recurrió a un enfoque más sencillo de estacionariedad, el cual es a través de los momentos. Un proceso es estacionario hasta el orden k , si todos los momentos hasta el orden k tienen la misma propiedad de estacionariedad. Por ejemplo; el segundo orden de estacionariedad asegura la existencia de los dos primeros momentos y además estos no dependen de la localidad x . Se asumirá que $Z(x)$ es una función aleatoria con dominio en todas las localidades $\{x_i\}$ y con rango un espacio ya sea en los reales, en R^1 o en R^n . En la realidad la forma de estacionariedad fuerte no explica el aspecto físico del fenómeno espacial.

Entonces se requiere de una forma de estacionariedad menos fuerte, al menos se necesita estacionariedad hasta el segundo momento, Myers (1989). La estacionariedad de segundo orden es la más usada en la Geoestadística, sin embargo, muchas veces se requiere una forma de estacionariedad menos fuerte. Por esa razón, se utiliza la hipótesis intrínseca ya que esta forma de estacionariedad explica mejor la variabilidad del fenómeno (sección 2.3.3).

2.3.1 ESTACIONARIDAD FUERTE

La estacionariedad fuerte no es muy utilizada en Geoestadística, sin embargo, es esencial para aplicar técnicas no lineales: tales como kriging universal, kriging disyuntivo, kriging indicador y kriging probabilístico, ver Zimmerman y Zimmerman (1989). La función aleatoria $Z(x)$ cumple con la estacionariedad fuerte, si para cualquier número finito de puntos x_1, x_2, \dots, x_n y cualquier vector $h \in D$, la función de densidad conjunta de $Z(x_1), \dots, Z(x_n)$ es la misma que la función de densidad conjunta de $Z(x_1+h), \dots, Z(x_n+h)$. Esta forma de estacionariedad no asegura la existencia de los valores esperados, las varianzas y covarianzas. De todas formas, los datos son representados como una muestra no aleatoria de una función aleatoria y por lo tanto no se puede probar por estacionariedad. La estacionariedad que se usa en Geoestadística es la llamada hipótesis intrínseca (2.3.3).

2.3.2 ESTACIONARIDAD DE SEGUNDO ORDEN

$Z(x)$ es de segundo orden estacionario si con las siguientes dos condiciones:

- (i) $Cov[Z(x+h), Z(x)]$ existe y depende sólo de h , esto implica que la $Var [Z(x)]$ existe y no depende de x .
- (ii) $E[Z(x)]$ existe y no depende de la variable x (detalles en Cox y Miller, 1965 pg 277).
- (ii) El segundo momento asegura la existencia del valor esperado y de la varianza.

La estacionariedad fuerte no implica la estacionariedad de segundo orden. Es decir, una función estacionaria fuerte no necesariamente es de segundo orden de estacionariedad, ya que sus dos primeros momentos podrían no estar definidos. Sin embargo, en la estacionariedad fuerte no siempre existen los valores esperados, varianzas y covarianzas. Por lo tanto no se puede asegurar la existencia de sus momentos.

2.3.3 HIPOTESIS INTRINSECA

La forma de estacionariedad implicada por la hipótesis intrínseca es esencialmente de segundo orden estacionario no por la función aleatoria $Z(x)$, sino por la diferencia de primer orden $Z(x+h) - Z(x)$. La hipótesis intrínseca es el tipo de estacionariedad que se necesita ya que considera la estacionariedad de las diferencias entre las muestras (Journel y Huijbregts, 1978). De esta forma se pueden clasificar los rezagos con respecto a la dirección y longitud del vector de separación entre los grupos de muestras.

Matheron (1963), define a la hipótesis intrínseca de la siguiente forma:

- (i) $E[Z(x+h) - Z(x)] = 0$ para todo $x, h \in D$
- (ii) $\gamma(h) = 1/2 \text{ var}[Z(x+h) - Z(x)]$ existe y depende sólo de h

La primera condición, pide que las diferencias varíen alrededor de un valor constante. La segunda condición asegura la existencia de una función que depende de h , la cual mide la correlación entre los datos (Capítulo 4).

CAPITULO 3

Determinación de la dependencia espacial

4. CUANTIFICACION DE LA INTERDEPENDENCIA ESPACIAL

Existen varias formas de medir la dependencia espacial entre los datos de un fenómeno, una de ellas es a través de una función llamada variograma la cual se denota por la letra griega $\gamma(\cdot)$. El variograma es una función que caracteriza las propiedades de dependencia espacial de segundo orden de un proceso estocástico definido en R^n . Su uso más importante es en la predicción espacial o Kriging.

Para construir el modelo Geostatístico es indispensable considerar a la variabilidad del fenómeno como una función aleatoria, y ésta a su vez como una realización de un proceso estocástico el cual definimos a continuación:

Sea $\{Z(s): s \in D \subset R^n\}$ un proceso estocástico de valor real definido en un dominio D de R^n , y supongamos que:

$$\text{Var}(Z(s+h) - Z(s)) = 2\gamma(h), \text{ para toda } s, s+h \in D \quad (3)$$

La cantidad $2\gamma(h)$ es una función de la diferencia entre los grados de volumen del fenómeno en estudio evaluados en las localidades espaciales $s, s+h$, esta función ha sido llamada *variograma* (Matheron, 1963).

Las condiciones del variograma son las siguientes:

1.- Se requiere una suposición débil de estacionaridad, la cual es llamada hipótesis intrínseca (sección 2.3.3).

El variograma y el correlograma son funciones que sirven para medir la interdependencia espacial entre las observaciones de un fenómeno.

A continuación se presenta otra forma de medir la dependencia entre las observaciones, la cual es una función llamada correlograma. El correlograma es una función $\rho(H)$, la cual depende de la distancia de las variables regionalizadas $Z(x), Z(x+h)$; las condiciones del correlograma son las siguientes:

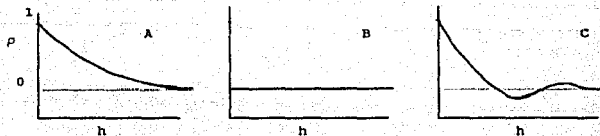
1.- La función correlograma requiere estacionaridad de segundo orden, (sección 2.3.2).

3.1 EL CORRELOGRAMA

El Correlograma $\rho(h)$ de la variable regionalizada Z está definida por:

$$\rho(h) = \frac{\text{Cov} [Z(x), Z(x+h)]}{\sigma^2} \quad (3.1)$$

donde, la covarianza se calcula para cualquier par de muestras de la variable regionalizada $Z(x)$, $Z(x+h)$, $x \in D$. La varianza de $Z(\cdot)$ es σ^2 , Matheron 1963, (sección 3.1.2). El correlograma está formado por una serie de correlaciones de una variable donde cada pareja dista de una longitud h una de la otra. En general, x y h son vectores, el valor ρ de 3.2 dependerá de la distancia h que existe entre las muestras como también de la dirección, $\rho(h)$ toma valores entre $(-1, 1)$.



Se presentan tres tipos de Correlogramas:

- A.- Representa un buen comportamiento, en el sentido que la correlación entre los datos, separados a distancias cortas de h , es alta.
- B.- Corresponde a una función que es independiente o que es aleatoria.
- C.- Corresponde a una función periódica.

Fig 3-1.

En la figura 3-1 se presentan tres tipos de Correlogramas etiquetados con las letras A, B, y C. En la gráfica A cuando $h=0$, el valor de $\rho(h)$ alcanza su máximo en 1.

En la gráfica B, a medida que h se incrementa el valor de $\rho(h)$ decrece gradualmente hasta que eventualmente para distancias grandes de h , $\rho(h)$ es cero, y por lo tanto la correlación no existe. Esta última gráfica corresponde a una variable aleatoria, donde $\rho(0) = 1$, y $\rho(h) = 0$, $\forall h > 0$.

La gráfica C representa un fenómeno cíclico. Cuando la distancia h se incrementa el correlograma se transforma negativo y positivo periódicamente hasta que finalmente para distancias muy grandes se hace cero.

Para estimar la función correlograma se supondrá que $N(h)$ es el número de parejas muestrales que están separadas a una distancia h . En general una distancia de clase τ se define como el grupo de parejas muestrales que están separadas a una distancia h dentro de una cierta tolerancia. Todas las parejas de datos que se encuentren dentro de esta tolerancia formarán un conjunto, el cual se denotará como h . Entonces, se hablará de todas las parejas de puntos que estén contenidas en este tipo de clase.

La función covarianza la definiremos por grupos (ver sección 1.4),

$$C(h) = \left[\frac{1}{N(h) - 1} \right] \sum_{i=1}^{N(h)} [Z(x_i) - \bar{Z}] [Z(x_{i+h}) - \bar{Z}]$$

y el correlograma muestral es:

$$r(h) = C(h) / s^2$$

Donde \bar{Z} y s^2 son estimadores del valor esperado y de la varianza respectivamente. Lo que se quiere es tener un número de parejas muy alto para rezagos pequeños, (Myers, 1987); pero en la práctica no sucede así, especialmente a distancias extremadamente grandes y extremadamente pequeñas. Esto es debido a que solamente se tienen n datos disponibles y no existe información adicional de observaciones separadas a distancias menores, ya que no hay más muestras.

3.2 EL VARIOGRAMA

El variograma es una función que da un valor de dependencia espacial entre los datos e indica qué tan parecidas son las muestras que están separadas a una distancia h . Es decir, el variograma da un valor al grado de variación entre las muestras separadas a una distancia h .

A continuación el variograma se expresa como "el valor esperado del cuadrado de la diferencia entre los volúmenes de las observaciones", Myers (1987). Esto es posible gracias a la existencia de los dos primeros momentos aplicados a la diferencia $Z(x) - Z(x+h)$ de las variables regionalizadas.

$$2\gamma(h) = E[Z(x+h) - Z(x)]^2 \quad (3.2)$$

3.2.1 PROPIEDADES DEL VARIOGRAMA

Matheron G. (1971), muestra que el variograma debe satisfacer la condición negativa semidefinida:

$\sum_{i=1}^n \sum_{j=1}^n a_i a_j 2\gamma(s_i - s_j) \leq 0$, para cualquier número finito de localidades espaciales $(s_i; i=1, \dots, n)$ y números reales $(a_i; i=1, \dots, n)$ que satisfagan la condición $\sum_{i=1}^n a_i = 0$.

El variograma tiene la propiedad de que $\gamma(-h) = \gamma(h)$, debido a que está en función de la longitud del vector $|h|$, además $\gamma(h) = 0$, para $h = 0$.

Si $Z(x)$ cumple con la hipótesis intrínseca entonces,

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x) - Z(x+h)] \quad (3.3)$$

El variograma es una función que depende de la distancia h , ya que en promedio mientras más separadas estén las observaciones más diferentes serán sus volúmenes muestrales¹. Este concepto explica el término llamado "Zona de influencia"² (Myers, 1987); es decir para cada muestra x_i , existe una área alrededor de la misma en donde los volúmenes muestrales son similares dependiendo de su proximidad con la muestra.

1.- Los volúmenes muestrales que se obtengan alrededor de un punto x , mantienen un valor similar al volumen del punto x , por las propiedades físicas del fenómeno, la cual es la definición de zona de influencia.

2.- La zona de influencia se refiere a que los valores muestrales son similares en cierta área.

Esto es una consecuencia de las propiedades físicas de los depósitos minerales que en términos Estadísticos se refiere a la dependencia que existe entre los datos. En ingeniería minera se define al variograma en tercera dimensión y se le da el nombre de semi-variograma o ley de dispersión, Matheron (1963). $\gamma(h)$ está definido sobre un vector h a través de la siguiente expresión:

$$\gamma(h) = \frac{1}{2V} \iiint_V \{Z(s+h) - Z(s)\}^2 dV \quad (3.4)$$

donde $Z(s)$ es el valor de la muestra la cual es la variable regionalizada tomada en el punto s del campo geométrico V , y dV se puede considerar como una densidad uniforme. La variable regionalizada $Z(x)$, está definida sobre un soporte geométrico V , el cual es el volumen mineral de la observación evaluada en un punto (Matheron, 1963). Con respecto al gráfico de la función variograma, a medida que la función se incrementa, la zona de influencia se deteriora y la dependencia entre las variables disminuye.

Esto se puede observar en la gráfica del variograma, donde la distancia τ_1 representa el valor de la covarianza entre las muestras separadas a una distancia h_1 (figura 3.2). Se nota cómo la covarianza $\tau_1 = \text{Cov}\{Z(x_0), Z(x_0+h_1)\}$ va disminuyendo conforme se incrementa el vector h_1 .

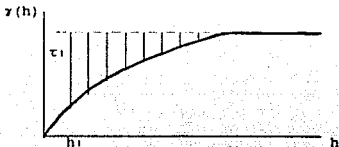


Figura 3.2

Las características cualitativas de las variables regionalizadas son expresadas satisfactoriamente a través del variograma.

a) Analizando el nivel de correlación entre las observaciones es posible determinar cuatro principales tipos de gráficas de Variogramas, los cuales se presentan a continuación.

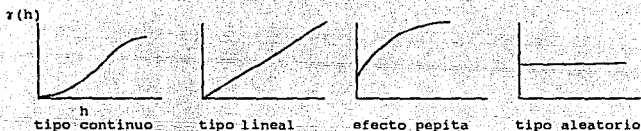


Figura 3-3.

En el primer tipo la gráfica muestra un correlación alta llamada variograma tipo continuo. Se le llama continuo debido a que las observaciones muestran un nivel de isotropía, es decir que el volumen de las muestras son similares.

En el segundo tipo, o tipo lineal, el variograma está caracterizado por una tangente oblicua en el origen, y representa una variable la cual tiene una continuidad "en promedio".

El tercer tipo, revela una discontinuidad en el origen y corresponde a una variable que no presenta continuidad en promedio pero sí un efecto pepita.

En el cuarto tipo el variograma corresponde a la noción clásica de variable aleatoria. Entre el tipo uno y el tipo cuatro existe toda una gama de variogramas intermedios los cuales son parte del objetivo de estudio de la Geostatística.

b) Si el variograma no es el mismo a lo largo de diferentes direcciones en el espacio entonces existen anisotropías en tales direcciones. La función $\gamma(h)$ no solo depende de la longitud del vector h , sino también de su dirección.

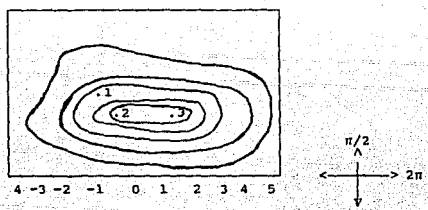
Muchas veces veremos que si modificamos el ángulo de dirección detectaríamos tendencias que no se habían encontrado antes e incluso identificaríamos diferentes tipos de anisotropías. En otras palabras el variograma sufre distorsiones cuando se altera la dirección y la longitud del vector h . De esta forma se obtiene información útil sobre la estructura de variación y localización de las anisotropías de las variables regionalizadas.

c) Las características estructurales del depósito son también reflejados por el variograma. Por ejemplo, el grueso del panel mineral aparece en la curva experimental como el rango a , en donde $\gamma(h)$ alcanza su máximo, es decir, el rango es igual al valor esperado del diámetro del volumen del depósito mineral, Matheron (1963).

El hecho de que los rangos son diferentes para varias direcciones, hace posible determinar en promedio, las direcciones de elongación del volumen del depósito mineral. Esto se puede lograr analizando la curva y el rango del variograma de esta forma podemos tener una idea de cómo y hacia dónde se desplaza el volumen del depósito (Fig. 3-4).

Por ejemplo: Supongamos que tenemos un depósito mineral y el objetivo es analizar como se desplaza físicamente su volumen mineral.

La gráfica 3-4 representa el área de muestreo, y supóngase que de acuerdo a las curvas de nivel de la gráfica existe un cerro en el centro del área de muestreo D, el cual tiende a alargarse en la dirección 2π , es decir hacia la derecha. Es por eso que al analizar la gráfica del variograma, se nota una tendencia en la dirección $\pi/2$. Por lo tanto, se calculó el Variograma éstas dos direcciones, figura 3-5A y 3-5B.



La relación de dependencia entre los puntos 1 y 2 en la dirección es menor que la relación de dependencia entre los puntos 2 y 3 en la dirección 2π .

Figura 3-4.

Analizando las gráficas de las figuras 3-5A y 3-5B se nota que el rango a_1 es menor que el rango a_2 . Esto quiere decir que el volumen mineral del depósito representado por la gráfica del Variograma de 3-5A aumenta más rápidamente hacia la dirección $\pi/2$ que el volumen del depósito representado por el Variograma de la gráfica 3-5B, en la dirección 2π .

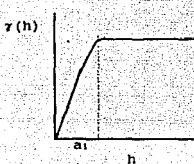


figura 3-5A
Gráfica del variograma
en la dirección $\pi/2$.

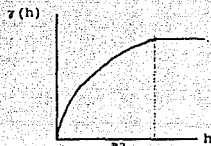


figura 3-5B
Gráfica del variograma
en la dirección 2π .

De esta forma se deduce hacia dónde se desplazan las anisotropías del depósito mineral. De la gráfica 3-5A se puede apreciar que hacia la dirección $\pi/2$ y para $h = a_1$, la covarianza alcanza el valor de cero más rápidamente que en la gráfica del Variograma 3-5B.

Para valores $h > a_1$ los volúmenes muestrales no mantienen correlación alguna. En cambio para valores de $h < a_1$ el variograma muestra una covarianza entre los datos muy alta para distancias h pequeñas.

3.2.1 VARIOGRAMA ISOTROPO

Algunas veces el variograma es una función solamente del vector h y no de la dirección. En este caso el variograma es llamado "isótropo".

El variograma es isótropo, si $2\gamma(h) = 2\gamma^\circ(\|h\|)$. Es decir el variograma es el mismo a lo largo de diferentes direcciones. Supongamos que el vector h se divide en d rezagos $\{h_i; i=1, \dots, d\}$, con objeto de evaluar el variograma en cada rezago.

Sea $h = (h_1, h_2, \dots, h_d) \in D \subset R^d$, y la norma de h es la usual

$$\|h\| = (h \cdot h)^{1/2}$$

El variograma isótropo mantiene el mismo valor a lo largo de todas las direcciones del vector h . Si el variograma mantiene los mismos parámetros, cuando este se evalúa hacia diferentes direcciones. Entonces el fenómeno en estudio es isótropo. Esto implica que las propiedades físicas de la distribución del volumen mineral son iguales a lo largo de todas sus direcciones, o sea que se trata de un fenómeno suave.

3.3 PARAMETROS DEL VARIOGRAMA

El variograma tiene asociados tres parámetros los cuales sirven para ajustar una función conocida a la descripción de correlación, entre los datos, obtenida por la gráfica del variograma. Los parámetros son nombrados efecto pepita, meseta y rango. Ellos describen en forma sintética los caracteres estructurales del fenómeno en estudio como son:

- El mayor o menor nivel de continuidad en el volumen mineral.
- La localización de zonas de influencia y anisotropías para las cuales el variograma sufre distorsiones cuando se altera el ángulo del vector h.
- El rango es un parámetro muy importante ya que representa en promedio el diámetro del grueso del depósito en estudio hacia determinada dirección, Matheron (1963).
Al calcular el rango en diferentes direcciones se pueden determinar las elongaciones o cambios en los volúmenes y la forma de la distribución del fenómeno en estudio.

Los parámetros del variograma se fijan al momento de ajustar un modelo variograma la relación de dependencia que existe entre los diferentes grupos de parejas muestrales, esto se hace de acuerdo a la distancia y dirección de separación entre los grupos de datos.

La estimación del variograma es mejor si el número de parejas muestrales es alto para distancias h's pequeñas. Lo anterior no quiere decir que se tomen muchas muestras separadas a distancias cortas, ya que entonces, se estarían repitiendo tantas veces como se quisiera el mismo o casi el mismo valor.

3.3.1 EFECTO PEPITA

El efecto pepita es un concepto que en Geostatística es sinónimo de incertidumbre, ya que no se puede asegurar el volumen del fenómeno en cierta área; dicho término se aplica cuando se obtienen muestras en cierta área donde su volumen asociado no exhibe ningún tipo de correlación.

La forma de interpretar al fenómeno efecto pepita dentro del modelo Geostatístico, es considerándolo como un parámetro del Variograma. Gráficamente, el efecto pepita es representado como una discontinuidad en el origen y corresponde a una variable regionalizada que no tiene "continuidad en promedio".

1.- Según Matheron 1963, la continuidad en promedio se refiere a la pobre correlación que existe entre los datos de un fenómeno espacial donde se presenta el efecto pepita.

Su nombre se debe a la minería, ya que en las perforaciones de los depósitos minerales frecuentemente se encuentran gránulos minerales llamado pepitas minerales, como en los depósitos de oro y diamante.

La presencia local de este efecto marca una incertidumbre en los volúmenes minerales de las observaciones, ya que no se puede asegurar el volumen mineral en esta área del depósito. También cuando se obtienen muestras con un cierto volumen mineral próximas a muestras con casi no volumen significaba la presencia de un efecto pepita.

3.3.2 MESETA

El parámetro Meseta representa el valor máximo del variograma para el cual la covarianza entre los datos se transforma cero. Cuando $|h| \rightarrow \infty$, el variograma tiende a un valor fijo, $\gamma(h) \rightarrow C(0)$. El valor $C(0)$ se le llama meseta, Cressie (1989).

La estimación de este parámetro está compuesta por la estimación de un meseta parcial más un estimador del efecto pepita, Matheron (1963). El parámetro meseta parcial pertenece a un proceso continuo, el cual es el valor esperado del cuadrado de la diferencia entre $Z(x)$ y $Z(x+h)$. Las modificaciones que se tengan que hacer al parámetro meseta se deben de aplicar a la meseta parcial en lugar de a la meseta total, debido a que este último incluye al efecto pepita.

3.3.3 RANGO

Para cualquier vector a , para el cual $\gamma(a) = C(0)$ y además cumpla con $\gamma(a(1-\epsilon)) < C(0)$, para cualquier $\epsilon > 0$. El vector " a " es llamado rango en la dirección $a/\|a\|$. El rango en la dirección $a/\|a\|$ especifica el rezago más pequeño para el cual las variables regionalizadas $Z(x)$ y $Z(x+a)$ no son correlacionadas.

Este parámetro frecuentemente es interpretado erróneamente, ya que el rango permite determinar en forma precisa las distancias entre los datos que deben ser incluidos en el predictor kriging,

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad (3.5)$$

El Rango no tiene una influencia directa en la varianza del valor estimado $Z(x)$, ya que indica la distancia para la cual los datos dejan de ser correlacionadas.

Solamente los datos dentro del rango de s_0 necesitan ser retenidos. El rango no indica que variables deben de ser incluidas en el predictor Kriging. Pero sí muestra las distancias en donde las correlaciones se hacen cero.

3.4 ESTIMACION DE LA VARIANZA

Cuando estimamos un valor desconocido $Z^*(x)$ a través de una combinación lineal de un grupo de n datos, la diferencia entre el valor observado menos el estimado forma un error, el cual definimos a continuación:

$$e(x) = [Z(x) - Z^*(x)]. \quad (3.6)$$

Este error es interpretado como una variable aleatoria, la cual es el resultado de la diferencia entre la variable regionalizada $Z(x)$ y su correspondiente valor estimado $Z^*(x)$.

Un buen estimador debe asegurar insesgamiento, es decir, que el valor esperado de la diferencia entre el observado y el estimado sea cero, además, que la estimación de la varianza del error sea mínima:

$$\left. \begin{aligned} (i) \quad E[Z(x) - Z^*(x)] &= 0. \\ (ii) \quad E[(Z(x) - Z^*(x))^2] & \end{aligned} \right\} \quad (3.7)$$

A continuación se considera un estimador de $Z^*(x)$ considerado como una combinación lineal de los datos:

$$Z^*(x) = \sum \lambda_i Z(X_i), \quad \text{con } i=1, \dots, n. \quad (3.8)$$

La forma que utiliza para minimizar la estimación de la varianza representada en la ecuación (3.11), es a través de normalizar los pesos:

$$\sum \lambda_i = 1.$$

La hipótesis intrínseca asegura que $E[Z(x+h)] = \mu$, $\forall x, h \in D$. Donde μ es el valor esperado de $Z(x+h)$. Esto implica que el valor esperado de $Z(x)$ es:

$$\begin{aligned} E[Z^*(x)] &= E\left[\sum \lambda_i Z(X_i)\right] \\ &= \sum \lambda_i E[Z(X_i)] \end{aligned}$$

ya que las λ_i 's son constantes, utilizando la condición (i) de la hipótesis intrínseca,

$$= \sum \lambda_i \mu$$

por hipótesis $\sum \lambda_i = 1$, $\forall i = 1, \dots, n$. Entonces,

$$E[Z^*(x)] = \mu.$$

2.- La estimación de la varianza se puede considerar como una función lineal del modelo variograma $\gamma(h)$, Journel y Huijbregts (1978):

$$\begin{aligned} \sigma_z^2 &= E \{ [Z(x) - Z^*(x)]^2 \} \\ &= 2 \sum_{i=1}^n \lambda_i \gamma(x - x_i) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(x_i - x_j) \end{aligned}$$

Bajo la hipótesis intrínseca (2.3.3) y aplicando la condición de sesgo se estima el volumen del fenómeno en estudio, ya sea en un punto o en un bloque, a través de una combinación lineal de los datos disponibles:

$$Z^*(x) = \sum_{i=1}^n \lambda_i Z_i(x)$$

Es importante mencionar que no es suficiente conocer el valor esperado y la varianza de la función aleatoria $Z(x)$, para caracterizar el error de estimación. Ya que también se necesitaría conocer los intervalos de confianza de la estimación, es decir se necesita saber, que nivel de confiabilidad se tiene en la predicción de un valor $Z^*(x_0)$, en una localidad conocida x_0 .

La práctica nos ha mostrado que cuando se utilizan datos referentes a algún fenómeno espacial, la distribución de los errores experimentales aunque no se distribuyen como una normal tienden a ser simétricos alrededor del valor esperado cero.

Afortunadamente, el intervalo $[0, \pm 2\sigma_z]$ contiene aproximadamente el 95% de los errores observados (Delfiner, 1976). La expresión de la estimación de la varianza solamente se refiere a la geometría de la información y más no a los datos. Ya que la estimación de la varianza se podría calcular antes de tomar cualquier muestra lo cual permite diseñar un buen método de muestreo.

3.5 VARIANZA DE LA DISPERSION

Al estudiar un fenómeno espacial el primer problema que se presenta es como medir la variabilidad del fenómeno, una forma de lograrlo es utilizando en concepto de variable regionalizada.

Sea V una área o volumen, en donde se cumple la hipótesis intrínseca. El área o volumen V es la unión de N unidades del mismo tamaño $v(x_i)$:

$$V = \sum_{i=1}^N v(x_i). \quad (3.9)$$

Sea $z(x_i)$ definida por (3.12), la cual es la característica promedio sobre cada unidad $v_i(x_i)$.

Una medida de la variabilidad de las unidades $z_v(x_i)$ dentro del área V es el valor promedio del cuadrado de las diferencias:

$$s^2 = \frac{1}{N} \sum [z(x_i) - z_v(x)]^2$$

Donde, $Z(x_i)$ es el valor observado y $Z_v(x)$ es el valor estimado. Esta medida s^2 la llamaremos varianza, la cual es generalmente desconocida, ya que los valores $z_v(x_i)$ son también desconocidos. Pero una vez que s^2 es interpretado como la obtención de una variable aleatoria S^2 , el valor esperado $E[S^2]$ se puede calcular. Por definición, "este valor esperado es la varianza de la dispersión $D^2(v|V)$, de v unidades dentro del área V " (Statistics Encyclopedia, pg 843).

CAPITULO 4

Estimación del Variograma

4. ESTIMACION DEL VARIOGRAMA

Todas las diferentes formas de Kriging utilizan el variograma para determinar un peso a cada observación durante el proceso de interpolación espacial. Los métodos de estimación del variograma son ineficientes debido a que la distribución del variograma se estima a partir de un número finito de datos y bajo suposiciones restrictivas. Por ejemplo, el negativo del variograma debe de ser positivo definido.

Otro problema que se presenta al estimar el variograma es la presencia de anisotropías en los datos. Aquí vamos a suponer que los fenómenos que analizamos son isotrópos. Por lo tanto, aquí se presentan modelos de variogramas isotrópos.

El estimador del variograma insesgado simple es el siguiente: Supóngase que $Z(x)$ satisface la hipótesis intrínseca entonces,

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

el estimador no paramétrico de $\gamma(h)$ es $\hat{\gamma}^*(h)$. Según Cressie (1988), se define como:

$$\hat{\gamma}^*(h) = \left[\frac{1}{2N(h)} \right] \sum_{i=1}^{N(h)} [Z(x_i+h) - Z(x_i)]^2 \quad (4.1)$$

donde $N(h)$ es el número de parejas $Z(x_i+h)$, $Z(x_i)$.

Debido a que $\hat{\gamma}^*(h)$ es esencialmente un valor esperado muestral tiene la desventaja de ser no robusto. Ya que se considera a la variable aleatoria $Z(x)$ como una obtención de un proceso estocástico, no se pueden hacer suposiciones respecto a la distribución de $Z(x)$. Por esta razón la distribución muestral de $\hat{\gamma}^*(h)$ no se conocerá.

Además de la naturaleza no robusta de $\hat{\gamma}^*(h)$, en la práctica pueden surgir otras dificultades. En un espacio de 2 o 3 dimensiones, h es un vector, esto implica que $\hat{\gamma}^*(h)$ está en función de la dirección e y de la longitud del vector h_r .

Esto es necesario para poder identificar posibles anisotropías. Entonces, el variograma estimado $\hat{\gamma}^*(h)$ debe de calcularse para diferentes distancias y ángulos.

Desafortunadamente $N(r, e)$ es muy pequeño para cualquier elección de h_r y e . $N(r, e)$ representa el número de parejas $Z(x)$ separadas a distancias de clase r y dirección e . Este inconveniente se presenta aun cuando las muestras se obtienen a iguales intervalos de separación de muestras consecutivas. El programa que se emplea en el ejemplo del capítulo 6 utiliza rezaqos en distintas direcciones para estimar el variograma experimental $\hat{\gamma}^*(h)$.

4.1 ESTIMACION DEL VARIOGRAMA POR MEDIO DE REZAGOS

El estimador propuesto por Myers (1987), calcula el variograma experimental $\gamma^*(\bar{r}, \bar{\theta})$ utilizando los diferentes rezagos y ángulos.

$$\gamma^*(\bar{r}, \bar{\theta}) = \frac{\iint G(|x-y|, x-y, u) [Z(x) - Z(y)]^2 dx dy}{2 \iint G(|x-y|, x-y, u) dx dy} \quad (4.2)$$

donde

$$G(|x-y|, x-y, u) = [h_1(|x-y|) - h_2(|x-y|)] [l_{\theta_1}(\theta) - l_{\theta_2}(\theta)]$$

$$\theta = \sin^{-1} \frac{\langle x-y, u \rangle}{|x-y|}$$

el vector u es unitario con dirección $\bar{\theta}$. Las parejas muestrales del variograma $\gamma(r, \theta)$ son todas las $Z(x)$, $Z(y)$, donde $h_1 < |x-y| < h_2$ y $\theta_1 < \theta < \theta_2$. Estos parámetros son utilizados para calcular $\gamma(r, \theta)$.

Supondremos que θ_1 , θ_2 , h_1 , h_2 son elegidos de tal forma que $h_1 < r < h_2$, y $\theta_1 < \theta < \theta_2$. Esta condición ayuda a la estimación del variograma porque aumenta el número de parejas usadas para calcular $\gamma(r, \theta)$.

Una desventaja de este método es que $\gamma^*(\bar{r}, \bar{\theta})$ es un estimador sesgado de $\gamma(\bar{r}, \bar{\theta})$.

En general el sesgo no es uniforme con respecto a \bar{r} , $\bar{\theta}$ y depende de los patrones muestrales,

$$E[\gamma^*(\bar{r}, \bar{\theta})] = \frac{\iint G(|x-y|, x-y, u) \gamma(|x-y|, \theta) dx dy}{\iint G(|x-y|, x-y, u) dx dy} \quad (4.3)$$

El sesgo será mínimo, donde γ es más suave y será mayor donde γ es no diferenciable.

Cuando los rezagos y las ventanas de ángulo¹ son usadas, frecuentemente es necesario experimentar con el ancho de cada rezago y con el ángulo para obtener un ajuste más suave. Este hecho complica el problema de diseñar un patrón muestral para optimizar el variograma muestral.

1.- Ventanas ángulo.- Se refiere a la amplitud de la tolerancia que se le da a un lado y otro del vector ángulo de dirección para determinar las muestras que entran en cada rezago, figura 6-8.

Para patrones regulares o casi regulares de muestreo, el número de parejas es pequeño para distancias muy cortas, grandes para distancias intermedias y decrece conforme la distancia es mayor que la mitad del diámetro de la región en estudio. Lo que se desea es tener un número grande de parejas a distancias cortas en el caso de que el fenómeno sea anisótropo, ya que se tendría una descripción más detallada de la dependencia entre los datos. En el caso de que el fenómeno sea isotropo no se necesitan tener muchos datos, ya que se estarían repitiendo casi los mismos valores.

Zimmerman y Zimmerman (1989), resume y compara varios métodos de estimación paramétrica del variograma. Ellos concluyen que ningún método domina, pero sus simulaciones indican que el método de mínimos cuadrados ponderados generalmente da buenos resultados. Noel Cressie (1985), propone minimizar la suma de cuadrados ponderados del siguiente modelo.

$$\sum_{k=1}^k \left[\frac{2\gamma(h(k))}{2\gamma(h(k); \tau_0)} - 1 \right]^2 |N(h(k))|, \quad (4.4)$$

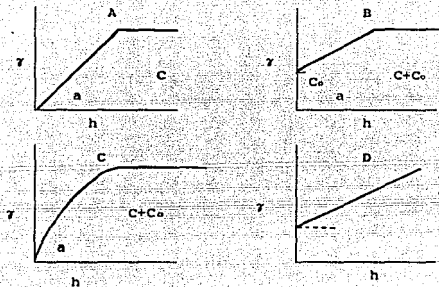
Esta minimización está hecha con respecto al parámetro del modelo variograma τ_0 . El vector h está dividido en k rezagos $h(1), h(2), \dots, h(k)$, que fueron usados para calcular el estimador no paramétrico (4.1), el cual satisface las condiciones dadas por Journel y Huijbregts.

4.2 AJUSTE DE MODELOS VARIOGRAMA

Se han desarrollado varios modelos de variograma y elegir el modelo variograma que describa mejor la dependencia espacial entre los datos.

En la figura 3-5A se muestra un variograma lineal el cual toma el valor de $\gamma(0) = 0$ y alcanza su máximo en $\gamma(h) = C$, $\forall h \geq a$.

El valor $h = a$ es la distancia máxima entre los datos, para las cuales las parejas muestrales son todavía correlacionadas. En algunos casos, existe un efecto pepita $\gamma(h) = C_0$, cuando C_0 es diferente de cero conforme h tiende a cero (figura 3-5B).



4-1 Ejemplos de variogramas típicos; A es un modelo lineal con rango a y techo C ; B es un modelo lineal con rango a , techo C y efecto pepita C_0 ; C es un modelo esférico con rango a , techo $C+C_0$; y D es un modelo lineal sin meseta, es decir no está acotado.

Figura 4-1.

Los anteriores ejemplos muestran que el variograma $\gamma(h)$ alcanza un valor constante para distancias grandes h . Pero no siempre sucede lo mismo, la figura 4-1D muestra que $\gamma(h)$ continua creciendo al menos para parejas muestrales separadas a una distancia h .

Para tales casos, la hipótesis intrínseca (sección 2.3.3) se cumple pero las condiciones de estacionaridad fuerte (sección 2.3.1) no se cumplen. Consecuentemente en este caso no podría existir el correlograma, ya que la varianza no tiene limite al menos para la escala de h . Sin embargo, la función variograma está definida y toma un valor determinado.

La elección de un modelo variograma válido es muy limitado ya que el negativo de $\gamma(h)$ debe de ser una función positiva definida, Journel y Huijbregts (1978), Armstrong (1989). Es mejor elegir modelos que sean conocidos por comportarse adecuadamente para ajustarlos al variograma experimental. Cuando la meseta existe se recomienda utilizar el modelo esférico o un modelo exponencial, Myers (1987).

4.2.1 VARIOGRAMA POTENCIA

El modelo potencia es el siguiente:

$$2\gamma(h, C_0, B, p) = 2 [C_0 + b|h|^p], \quad \text{para } h > 0. \quad (4.5)$$

Donde C_0 es el efecto pepita, h es un vector, y $p > 0$ es la potencia.

4.2.2 VARIOGRAMA ESFERICO

El modelo esférico es ilustrado en la figura 4-1C y está expresado por,

$$\gamma(h) = \begin{cases} C_0 + C[(3/2) - (1/2)(h/a)^3], & \text{para } 0 < h < a \\ C_0 + C, & h \geq a \end{cases} \quad (4.6)$$

donde, C_0 es el efecto pepita, a es el rango, y $C_0 + C$ es la meseta.

4.2.3 VARIOGRAMA EXPONENCIAL

El modelo exponencial es el siguiente:

$$\gamma(h) = \begin{cases} 0, & \text{para } h = 0 \\ C_0 + C_1 [1 - \exp(-r/a)], & \text{para } h > 0, \end{cases} \quad (4.7)$$

donde, $C_0 \geq 0$, $C_1 \geq 0$, $a > 0$, y C_0 es el efecto pepita, a es el rango, y $C_0 + C_1$ es la meseta.

4.2.4 VARIOGRAMA LINEAL

El modelo lineal es el siguiente:

$$\gamma(h) = \begin{cases} C_0 + C(h/a), & \text{para } h < a \\ C_0 + C, & \text{para } h \geq a \end{cases} \quad (4.8)$$

donde, C_0 es el efecto pepita, a es el rango, y $C_0 + C$ es la meseta.

4.2.5 VARIOGRAMA GAUSSIANO

En el caso particular de la función covarianza de un proceso Gaussiano, el proceso es muy suave debido a que es infinitamente diferenciable. Se ha demostrado que pequeñas modificaciones a la anisotropía pueden resultar en predicciones con valores muy diferentes en los promedios de las diferencias del cuadrado de los errores. Según Stein y Handcock (1983), la función covarianza es un modelo pobre para la mayoría de los procesos espaciales, ya que existen propiedades físicas que influyen sobre ella como por ejemplo la anisotropía.

$$\gamma(h) = \begin{cases} 0, & \text{para } h = 0 \\ C_0 + C_1[1-\exp(-h/a)], & \text{para } h > 0 \end{cases} \quad (4.9)$$

donde, $C_0 \geq 0$, $C_1 \geq 0$, $a > 0$

4.3 PROBLEMAS COMUNES ENCONTRADOS EN VARIOGRAMAS

Algunos textos de Geostatística dan la impresión de que es muy fácil calcular el variograma experimental y ajustar un modelo variograma, pero aquellos que lo han intentado se han dado cuenta que no es así. Frecuentemente el variograma experimental es aberrante. En esta sección se presentan algunos de los problemas más comunes vistos en variogramas.

Matheron (1965) destaca el hecho de que el variograma experimental es muy variable para longitudes de h muy grandes. El variograma experimental podría ser muy diferente para distintos subgrupos de datos.

A continuación presentamos los problemas más comunes que se presentan en los variogramas.

Elección pobre de distancias de clase

Una elección pobre en los rezagos causa un comportamiento muy errático en la gráfica del variograma. Por ejemplo. Supongamos que graficamos un histograma, donde el número de apariciones de parejas muestrales en cierto rezago es muy grande y para otros rezagos el número de parejas es muy pequeño.

Si no se toma en cuenta este hecho, el comportamiento del variograma sería no robusto. En la figura 4-2a se muestra un variograma de contaminación de carbón cuyos niveles se han calculado a 40 metros de separación. La gráfica 4-2b muestra el histograma donde se ven dos módulos con una frecuencia de aparición de parejas muy alta, cada una a 160 y 320 metros de separación.

En este caso el variograma debe de ser recalculado utilizando un incremento distinto, el cual no debería de ser un múltiplo de 160 (por ejemplo, 100). De esta forma las parejas muestrales se distribuirían entre otros rezagos y el variograma sería más suave que el anterior.

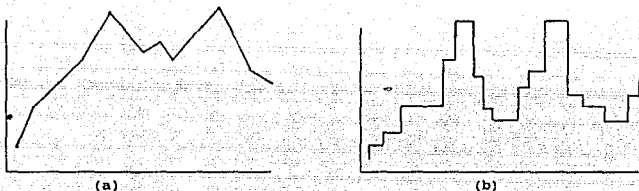


Figura 4-2

Poblaciones mezcladas

El término poblaciones mezcladas se refiere a que existen dos grupos de datos que vienen de dos poblaciones estadísticamente diferentes. Este tipo de problemas podrían surgir al efectuar dos muestreos en zonas distintas y mezclar los datos, o también este problema se podría presentar por la unión de dos grupos de datos provenientes de muestreos efectuados con diferentes métodos pero sobre la misma área.

CAPITULO 5

El estimador Kriging Ordinario

5. EL MODELO GEOESTADISTICO

La necesidad de obtener mejores predicciones a partir de un grupo de observaciones se puede encontrar en cualquier disciplina científica. Aquellas áreas que han tomado en cuenta nociones estadísticas referentes a variaciones aleatorias para predecir un valor en un espacio de dimensión n pueden llegar a dicho objetivo. Esto se lleva a cabo a través de encontrar la dependencia existente entre los datos.

Nuestro modelo consiste en una función aleatoria estacionaria formada por un grupo de n variables aleatorias correlacionadas $\{Z(x_i)\}$.

Estas variables aleatorias satisfacen la misma ley probabilística la cual se refiere a que el valor esperado de cualquiera de los n puntos es el mismo:

$$E [Z(x_i)] = E [Z(x)].$$

5.1 EL SESGO Y EL MODELO GEOESTADISTICO

Primeramente, consideremos el valor estimado como una combinación lineal de las n variables aleatorias.

Como vimos en la sección 3.5, cuando estimamos un valor desconocido $Z(x)$ a través de una combinación lineal de variables aleatorias, ecuación 3.7, se forma un error el cual es a su vez una variable aleatoria, Goldgerg (1962):

$$c_1 = Z^*(x_1) - Z(x_1).$$

El valor promedio m_r es también aleatorio:

$$m_r = \frac{1}{k} \sum_{i=1}^k c_i = \frac{1}{k} \sum_{i=1}^k [Z^*(x_i) - Z(x_i)]. \quad (5.1)$$

Desafortunadamente, se puede utilizar esta ecuación para obtener un promedio de la medida de los errores, ya que se necesitarían conocer los valores reales $Z(x_i)$, y estos no se conocen.

La solución al problema de encontrar una medida del error, consiste en considerar a los valores reales desconocidos $Z(x)$, como parte de un proceso estocástico y resolver el problema para el modelo teórico, aquí se va a utilizar un grupo de datos los cuales se suponen que representan la realidad.

Según el capítulo 3, cualquier par de variables aleatorias tienen una distribución conjunta que sólo depende de la separación entre ellas, y no de la localidad.

Sea C_h la covarianza entre las variables aleatorias $Z(x_i)$, $Z(x_j)$ las cuales están separadas a una distancia h .

El error asociado a la estimación de una variable aleatoria evaluada en un punto x_0 es el siguiente:

$$R(x_0) = Z^*(x_0) - Z(x_0). \quad (5.2)$$

Si se sustituye $Z^*(x_0)$ definido en 5.1 se obtiene la siguiente expresión:

$$R(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) - Z(x_0). \quad (5.3)$$

ahora se puede asegurar que el valor esperado del error en cualquier punto es cero, ya que se supone que el valor observado es igual al valor estimado.

$$\begin{aligned} E[R(x_0)] &= E\left[\sum_{i=1}^n \lambda_i Z(x_i) - Z(x_0)\right] \\ &= \sum_{i=1}^n \lambda_i E[Z(x_i)] - E[Z(x_0)] \end{aligned}$$

Utilizando el supuesto de que la función aleatoria es estacionaria, entonces se tiene:

$$E[R(x_0)] = \sum_{i=1}^n \lambda_i E[Z] - E[Z].$$

El valor esperado del error $E[R(x_0)]$ en cualquier punto es el sesgo en ese punto. Si se iguala a cero ese valor, entonces el valor estimado $Z(x)$ en ese punto es insesgado, por lo que se obtiene la siguiente expresión:

$$\begin{aligned} E[R(x_0)] = 0 &\Rightarrow E[Z] \sum \lambda_i - E[Z] = 0 \\ &\Rightarrow E[Z] \sum \lambda_i = E[Z] \\ &\Rightarrow \sum_{i=1}^n \lambda_i = 1. \end{aligned}$$

5.2 EL ERROR DE VARIANZA Y EL MODELO GEOESTADISTICO

El método Kriging Ordinario tiene la característica de producir un grupo de estimadores cuya varianza de error es mínima.

$$\sigma_R^2 = \frac{1}{K} \sum_{i=1}^n [c_i - m_r]^2 \quad (5.4)$$

Donde, σ_R^2 es la varianza del error y m_r es el valor promedio de los errores cuya ecuación la definimos en 5.1.

Sustituyendo el valor de c_i y el de m_r en la ecuación 5.4 se obtiene lo siguiente:

$$\sigma_R^2 = \frac{1}{K} \sum_{i=1}^n [Z^*(x_i) - Z(x_i)] - \frac{1}{K} \sum_{i=1}^n [Z^*(x_i) - Z(x_i)]^2$$

Donde, los $\{Z(x_i)\}$ son los valores reales y $\{Z^*(x_i)\}$ son los correspondientes valores estimados.

Si se supone que el valor esperado del error es cero, entonces la varianza del error representada por la ecuación anterior se reduce a la siguiente:

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

la razón de suponer que el valor esperado del error es cero es simplemente para suponer que los valores observados $Z(x)$ son iguales que sus correspondientes valores estimados $Z^*(x)$.

Desafortunadamente, esta expresión contiene los valores reales los cuales se desconocen; y por lo tanto, no se puede calcular la varianza del error. Para poder resolver este problema se va a recurrir a los modelos de funciones aleatorias.

Primeramente, se forma un sistema de $n+1$ ecuaciones, de las cuales n variables modelan el comportamiento del fenómeno alrededor de las localidades en donde se tomaron las muestras, y una variable modela el comportamiento del fenómeno en la localidad por estimar.

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

Ahora se va a transformar el problema original de minimizar la varianza en su correspondiente modelo $R(x_0)$. Entonces, se resuelve el problema equivalente utilizando el siguiente modelo.

El procedimiento de minimización de la varianza del error se lleva a cabo al encontrar una expresión para la varianza del error en el modelo σ_R^2 , para luego igualar a cero las derivadas parciales de primer orden.

En resumen, primero se encuentra una expresión para σ_R^2 , la forma de hacerlo es recurrir al teorema de la varianza de una combinación lineal de variable aleatorias, Edward Issaaks (1989), la cual es:

$$\text{Var} \left[\sum_{i=1}^n \lambda_i Z_i \right] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov} [Z_i Z_j].$$

Utilizando esta ecuación junto con la expresión 5.2, se puede escribir la varianza del error como:

$$\begin{aligned} \text{Var} [R(x_0)] &= \text{Cov} [Z'(x_0) Z'(x_0)] - \text{Cov} [Z(x_0) Z'(x_0)] \\ &\quad - \text{Cov} [Z'(x_0) Z(x_0)] + \text{Cov} [Z(x_0) Z(x_0)] \\ &= \text{Cov} [Z'(x_0) Z'(x_0)] - 2 \text{Cov} [Z'(x_0) Z(x_0)] \\ &\quad + \text{Cov} [Z(x_0) Z(x_0)]. \end{aligned}$$

El primer término es igual a:

$$\begin{aligned} \text{Cov} [Z'(x_0) Z'(x_0)] &= \text{Var} [Z'(x_0)] = \text{Var} \left[\sum_{i=1}^n \lambda_i Z(x_i) \right] \\ &= \text{Var} \left[\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \bar{C}_{ij} \right]. \end{aligned}$$

El segundo término es equivalente a:

$$\begin{aligned}
 \text{Cov} [Z^*(x_0) \cdot Z(x_0)] &= \text{Cov} \left[\sum_{i=1}^n \lambda_i Z(x_i) \cdot Z(x_0) \right] \\
 &= E \left[\sum_{i=1}^n \lambda_i Z(x_i) \cdot Z(x_0) \right] - E \left[\sum_{i=1}^n \lambda_i Z(x_i) \right] \cdot E [Z(x_0)] \\
 &= \sum_{i=1}^n \lambda_i E [Z(x_i) \cdot Z(x_0)] - \sum_{i=1}^n \lambda_i E [Z(x_i)] \cdot E [Z(x_0)] \\
 &= \sum_{i=1}^n \lambda_i \left[E [Z(x_i) \cdot Z(x_0)] - E [Z(x_i)] \cdot E [Z(x_0)] \right] \\
 &= \sum_{i=1}^n \lambda_i \left[E [Z_i \cdot Z_0] - E [Z_i] \cdot E [Z_0] \right] \\
 &= \sum_{i=1}^n \lambda_i \text{Cov} [Z_i \cdot Z_0]
 \end{aligned}$$

El tercer término es la varianza de $Z(x_0)$:

$$\text{Cov} [Z(x_0) \cdot Z(x_0)] = \text{Cov} [Z_0 \cdot Z_0] = \text{Var} [Z_0] = \bar{\sigma}^2.$$

Finalmente se suman estos tres términos, y se obtiene una expresión para el valor estimado de la varianza del error σ_c^2 :

$$\bar{\sigma}_c^2 = \bar{\sigma}^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \bar{c}_{ij} - 2 \sum_{i=1}^n \lambda_i \bar{c}_{i0} \quad (5.5)$$

Ahora bien, el objetivo es minimizar 5.5 sujeta a la restricción 5.1. La forma de hacerlo es a través de obtener las derivadas parciales de primer orden e igualarlas a cero.

Este proceso produce un sistema de n ecuaciones con n incógnitas. Sin embargo, el proceso de solución debe de considerar la restricción 5.1, la cual implica introducir una ecuación más.

Esto significa que se tiene un sistema de $n+1$ ecuaciones con n incógnitas. Sin embargo, el modelo no acepta cualquier solución, ya que se está utilizando el sesgo aplicado al vector de pesos λ .

Es decir, la solución se restringe a aquellas soluciones cuya suma sea igual a 1. Tales problemas se pueden resolver por medio de la técnica Lagrange.

5.3 TECNICA LAGRANGE

Este método es un procedimiento que convierte un problema de optimización con restricciones en uno sin restricciones. La condición 5.4 aumenta una ecuación sin aumentar una variable. Por lo que se tiene, un sistema de $n+1$ ecuaciones con n incógnitas. Para evitar dificultades a continuación se muestra cómo se aumentó el parámetro μ a la ecuación 5.5 con el objeto de extender el sistema original a otro equivalente de $n+1$ ecuaciones con $n+1$ incógnitas (Edward Issaaks, 1989):

$$\bar{\sigma}_E^2 = \bar{\sigma}^2 + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \bar{C}_{ij} - 2 \sum_{i=1}^n \lambda_i \bar{C}_{i0} \left[2\mu \left(\sum_{i=1}^n \lambda_i - 1 \right) \right]. \quad (5.6)$$

Ahora hemos transformado la ecuación 5.1 en un sistema de optimización sin restricciones. El número que se aumentó en 5.6 es cero, el cual se formó de la ecuación 5.1, en donde se asegura el sesgo:

$$\begin{aligned} \sum_{i=1}^n \lambda_i &= 1 \\ \sum_{i=1}^n \lambda_i - 1 &= 0 \\ 2\mu \left(\sum_{i=1}^n \lambda_i - 1 \right) &= 0. \end{aligned}$$

Al igualar a cero las $n+1$ derivadas parciales de primer orden, con respecto a cada variable λ_i . Se obtiene un sistema de $n+1$ ecuaciones con $n+1$ incógnitas. además, si se iguala a cero la primer derivada parcial, de 5.6 se obtiene la condición de sesgo.

$$\frac{\partial (\bar{\sigma}_E^2)}{\partial \mu} = \frac{\partial (2\mu (\sum \lambda_i - 1))}{\partial \mu} = 2 (\sum_{i=1}^n \lambda_i - 1) = 0 \Rightarrow \sum_{i=1}^n \lambda_i - 1.$$

Ahora el sistema tiene $n+1$ variables y $n+1$ incógnitas.

5.4 MINIMIZACION DEL ERROR DE LA VARIANZA

El objeto de esta sección es minimizar la varianza del error representada en la ecuación 5.6, la cual se minimiza a través de calcular las $n+1$ derivadas parciales de primer orden e igualarlas a cero.

Primero, se deriva el segundo término de la ecuación 5.6. La derivada parcial de primer orden con respecto a la primer variable λ_1 es la siguiente:

$$\frac{\partial \left(\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \bar{C}_{ij} \right)}{\partial \lambda_1} = \frac{\partial (\lambda_1^2 \bar{C}_{11}) + 2\lambda_1 \sum_{j=2}^n \lambda_j \bar{C}_{1j} + \sum_{j=2}^n \lambda_j^2 \bar{C}_{jj} + \sum_{i=2}^n 2\lambda_i \sum_{j=1}^n \lambda_j \bar{C}_{ij})}{\partial \lambda_1}$$

$$= \lambda_1 \bar{C}_{11} + 2 \sum_{j=2}^n \lambda_j \bar{C}_{1j} = 2 \sum_{j=1}^n \lambda_j \bar{C}_{1j}$$

El tercer término de 5.6 contiene sólo un valor de λ_1 y su derivada parcial de primer orden es la siguiente:

$$\frac{\partial \left(\sum_{i=1}^n \lambda_i \bar{C}_{i0} \right)}{\partial \lambda_1} = \frac{\partial (\lambda_1 \bar{C}_{10})}{\partial \lambda_1} = \bar{C}_{10}$$

El último término de 5.6 solamente tiene un valor de λ_1 y su derivada parcial de primer orden es la siguiente:

$$\frac{\partial \left(-2\mu \left(\sum_{i=1}^n \lambda_i - 1 \right) \right)}{\partial \lambda_1} = \frac{\partial (-2\mu \lambda_1)}{\partial \lambda_1} = -2\mu$$

Finalmente, la derivada parcial de primer orden de la ecuación 5.6 es:

$$\frac{\partial (\bar{\sigma}_E^2)}{\partial \lambda_1} = 2 \sum_{j=1}^n \lambda_j \bar{C}_{1j} - 2 \bar{C}_{10} + 2\mu$$

Igualando esta última expresión a cero tenemos:

$$2 \sum_{j=1}^n \lambda_j \bar{c}_{1j} - 2 \bar{c}_{10} + 2\mu = 0.$$

De donde se obtiene la siguiente expresión:

$$\sum_{j=1}^n \lambda_j \bar{c}_{1j} + \mu = \bar{c}_{10}. \quad (5.7)$$

De igual forma se obtienen las derivadas restantes:

$$\frac{\partial (\sigma_E^2)}{\partial \lambda_1} = 2 \sum_{j=1}^n \lambda_j \bar{c}_{1j} - 2 \bar{c}_{10} + 2\mu = 0 \Rightarrow \sum_{j=1}^n \lambda_j \bar{c}_{1j} + \mu = \bar{c}_{10}$$

$$\frac{\partial (\sigma_E^2)}{\partial \lambda_2} = 2 \sum_{j=1}^n \lambda_j \bar{c}_{2j} - 2 \bar{c}_{20} + 2\mu = 0 \Rightarrow \sum_{j=1}^n \lambda_j \bar{c}_{2j} + \mu = \bar{c}_{20}$$

$$\vdots$$

$$\frac{\partial (\sigma_E^2)}{\partial \lambda_n} = 2 \sum_{j=1}^n \lambda_j \bar{c}_{nj} - 2 \bar{c}_{n0} + 2\mu = 0 \Rightarrow \sum_{j=1}^n \lambda_j \bar{c}_{nj} + \mu = \bar{c}_{n0}$$

Por lo tanto, el grupo de pescos que minimiza el error de la varianza bajo la condición de que su suma sea 1, satisface el siguiente sistema de $n+1$ ecuaciones:

$$\sum_{j=1}^n \lambda_j \bar{c}_{ij} + \mu = \bar{c}_{i0}, \text{ para toda } i = 1, \dots, n \quad (5.8)$$

$$\sum_{j=1}^n \lambda_j = 1.$$

Este sistema se puede escribir en notación matricial, el cual

es llamado sistema Kriging Ordinario:

$$\begin{bmatrix} C_{11} & \dots & C_{1n} & 1 \\ \vdots & & \vdots & \vdots \\ C_{n1} & \dots & C_{nn} & 1 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{C}_{10} \\ \vdots \\ \bar{C}_{n0} \\ 1 \end{bmatrix} \quad (5.9)$$

$(n+1) \times (n+1)$ $(n+1) \times 1$ $(n+1) \times 1$

$$C \cdot \lambda = B.$$

Ahora, se tiene que despejar el vector de pesos λ de la expresión anterior:

$$\begin{aligned} C \cdot \lambda &= B \\ C^{-1} \cdot C \cdot \lambda &= C^{-1} \cdot B \\ \lambda &= C^{-1} \cdot B \end{aligned} \quad (5.10)$$

Este resultado permite encontrar una expresión para el valor mínimo de la varianza del error.

$$\begin{aligned} \bar{\sigma}_E^2 &= \bar{\sigma}^2 + \sum_{i=1}^n \lambda_i \bar{C}_{i0} - \mu - 2 \sum_{i=1}^n \lambda_i \bar{C}_{i0} \\ &= \bar{\sigma}^2 - \left(\sum_{i=1}^n \lambda_i C_{i0} + \mu \right) \end{aligned} \quad (5.11)$$

Esta última ecuación se puede expresar en términos matriciales como:

$$\bar{\sigma}_E^2 = \bar{\sigma}^2 - \lambda \cdot B \quad (5.12)$$

donde, $\bar{\sigma}^2$ es la varianza Kriging Ordinaria, $\bar{\sigma}^2$ es la varianza de $Z(x_0)$, λ es el vector de pesos que se quiere estimar, y B es el vector de covarianzas del punto x_0 con respecto a todos los puntos restantes.

5.5 VALIDACION CRUZADA

Este es un procedimiento para probar que tan exactas fueron las estimaciones hechas por el método Kriging al usar el modelo variograma hipotético y los parámetros del variograma.

El método kriging ordinario es un estimador lineal exacto ya que si se estima un valor en una de las localidades muestradas, se obtiene exactamente el mismo valor.

Para probar que tan exactas fueron las predicciones, se propone el método de validación cruzada definido por Donald E. Myers (1987), Matheron (1965) y Journel y Huijbregts (1978).

El proceso de validación cruzada es el siguiente:

Supóngase que se tienen n observaciones, el objetivo es eliminar una observación x_i para estimar ese valor utilizando las $n-1$ observaciones restantes. Este proceso se repite n veces, una para cada observación. De esta forma se tienen n valores estimados, donde cada uno corresponde al valor observado.

La idea es calcular ciertas estadísticas para determinar qué tan bueno fué el modelo variograma elegido para la estimación.

Si el modelo variograma refleja adecuadamente la correlación espacial entre el grupo de datos: entonces, los valores kriging estimados deben de estar cerca de los valores observados. Esta cercanía puede ser cuantificada por varios caminos. Generalmente, las siguientes estadísticas son las más usadas:

$$(a) \quad \frac{1}{n} \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]$$

Teóricamente el valor esperado de (a) debería de ser cero, ya que el valor observado sería igual al valor estimado.

$$(b) \quad \frac{1}{n} \sum_{i=1}^n [Z(x_i) - Z^*(x_i)]^2$$

El valor esperado de (b) debería ser muy pequeño.

$$(c) \quad \frac{1}{n} \sum_{i=1}^n \left[\frac{Z(x_i) - Z^*(x_i)}{\sigma_i} \right]^2$$

Este valor debería ser uno ya que para cada muestra la diferencia entre el valor observado y el estimado se divide entre su desviación estándar σ_i .

5.6 KRIGING POR BLOQUES

Kriging Ordinario es un método de interpolación espacial puntual, pero muchas veces es necesario estimar un bloque o un promedio de una área o volumen determinado, Edward Issaaks (1989).

Supóngase que se quiere estimar el valor promedio de algún fenómeno sobre una área A.

$$V_A = \frac{1}{|A|} \sum_{j \in A} V_j.$$

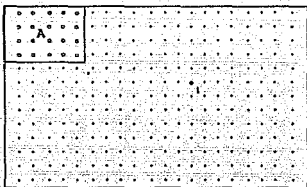
Donde V_A es una variable aleatoria que corresponde al valor esperado de una Área A y V_j son variables aleatorias correspondientes a puntos dentro del espacio de localidades D, pero que no están dentro de A (figura 5-1).

Se necesitan transformar las covarianzas puntuales a covarianzas por bloques. Al realizar este único cambio, se puede convertir el sistema Kriging Ordinario puntual a un sistema Kriging Ordinario por bloques:

$$\begin{aligned} C_{1A} &= \text{Cov} (V_A \cdot V_i) \\ &= E (V_A \cdot V_i) - E (V_A) \cdot E (V_i) \\ &= E \left\{ \frac{1}{|A|} \sum_{j \in A} V_j \cdot V_i \right\} - E \left\{ \frac{1}{|A|} \sum_{j \in A} V_j \right\} \cdot E (V_i) \\ &= \frac{1}{|A|} \sum_{j \in A} E \{ V_j \cdot V_i \} - \frac{1}{|A|} \sum_{j \in A} E \{ V_j \} \cdot E \{ V_i \} \\ &= \frac{1}{|A|} \sum_{j \in A} \left\{ E \{ V_j \cdot V_i \} - E \{ V_j \} \cdot E \{ V_i \} \right\} \\ &= \frac{1}{|A|} \sum_{j \in A} \text{Cov} (V_j \cdot V_i) \end{aligned}$$

V_A es una variable aleatoria que corresponde al valor esperado, o valor promedio dentro de una área A. V_i también es una variable aleatoria y corresponde a los valores puntuales dentro del área A. V_j es otra variable aleatoria que corresponde a los valores puntuales fuera del área A.

D



La covarianza entre el bloque A y el punto i fuera de A es igual a la promedio de la covarianza entre todos los punto j con el punto i fuera de A.

Figura 5-1.

La covarianza entre el punto i y el bloque A se representa por la siguiente ecuación:

$$C_{iA} = \frac{1}{|A|} \sum_{j|j \in A} C_{ij}$$

El sistema Kriging 5.8 se transforma en el siguiente sistema:

$$\begin{bmatrix} \bar{C}_{11} & \dots & \bar{C}_{1n-1} \\ \vdots & \ddots & \vdots \\ \bar{C}_{n1} & \dots & \bar{C}_{nn-1} \end{bmatrix} \cdot \begin{bmatrix} \omega_1 \\ \vdots \\ \omega_{n-1} \\ \mu \end{bmatrix} = \begin{bmatrix} \bar{C}_{10} \\ \vdots \\ \bar{C}_{n0} \\ 1 \end{bmatrix} \quad (5.13)$$

$(n+1) \times (n+1)$ $(n+1) \times 1$ $(n+1) \times 1$

$$C \cdot \omega = B.$$

La varianza Kriging por bloques es:

$$\sigma_{ok}^2 = \bar{C}_{AA} - \left(\sum_{i=1}^n \omega_i \bar{C}_{iA} + \mu \right) \quad (5.14)$$

La covarianza promedio entre los pares muestrales dentro de A es:

$$\bar{C}_{AA} = \frac{1}{|A|^2} \sum_{i \in A} \sum_{j \in A} C_{ij}.$$

El sistema Kriging Ordinario por bloques tiene la ventaja de utilizar el promedio de una área y calcular la covarianza con otras áreas o con puntos fuera de A. El caso más común se presenta cuando el fenómeno en estudio contiene áreas en donde sus puntos no varían mucho entonces se obtiene un valor promedio que representa dicha área para evitar calcular las covarianzas de todos los puntos dentro de A con los puntos fuera de A.

CAPITULO 6

Ejemplo de un análisis Geostatístico

6. APLICACION DEL METODO KRIGING ORDINARIO

Lo que se pretende en este capitulo es desarrollar y mostrar el procedimiento de análisis Geoestadístico aplicado a un ejemplo con objeto de:

- Predecir un valor $Z(x_0)$ en una localidad conocida x_0 , en donde no se había obtenido una observación inicialmente.
- Encontrar una medida del error de la estimación.
- Estimar una estructura de correlación entre los datos.

El procedimiento general de predicción espacial consiste en:

- 1.- Encontrar el predictor $Z(x_0)$ definido en (3.10) conociendo la localidad x_0 .
- 2.- Determinar el estimador de la varianza, $\sigma^2(x_0)$ definida en (5.5), conforme x_0 varía a lo largo de la superficie de localidades.
- 3.- Calcular las ecuaciones kriging (5.9), una vez que el Variograma (4.2) se conoce.

Este ejemplo tiene el enfoque práctico de encontrar cómo se distribuyen los precios por metro cuadrado de terrenos en la ciudad de Cuernavaca. Este estudio permite localizar y hacer deducciones acerca de la distribución de las áreas comerciales, residenciales y de zonas marginadas.

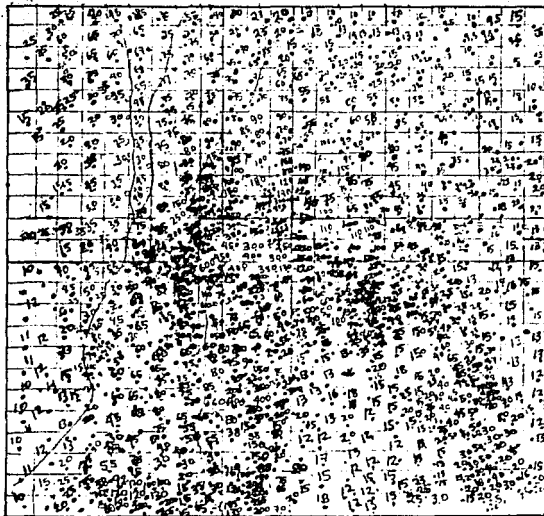
Los datos

Se tomó una muestra de 693 terrenos distribuidos uniformemente sobre toda la superficie de la ciudad de Cuernavaca. Debido a que el tamaño y forma del terreno influye en la calidad de la interpolación, se decidió tomar las muestras en el punto sur-oeste de cada terreno. La figura 6-1 exhibe el área total en donde se efectuó la medición y su respectivo precio en dólares por metro cuadrado.

Es importante mencionar que el tamaño y la forma de los terrenos influyen en el variograma experimental, y por lo tanto en las ecuaciones Kriging. Esto se debe a que el método kriging es un estimador puntual, sin embargo aquí se consideran los terrenos como bloques de áreas y un metro cuadrado en cada terreno se considerará como un punto muestral.

El problema que se presenta al elegir el metro cuadrado como punto muestral dentro del terreno es que el tamaño y la forma del terreno afecta la distancia entre las muestras, y por esta razón el variograma y las ecuaciones Kriging cambian. Lo que se propone en esta sección es estandarizar la forma de elegir el punto muestral dentro del terreno, en este caso se eligió el metro cuadrado que se encuentra en el extremo de la posición sur-oeste de cada terreno de acuerdo a la dirección Norte señalada en la figura 6-3. Esto implica que las observaciones tendrán como coordenadas las correspondientes a la parte inferior izquierda de cada terreno, figura 6-2.

Se realizaron dos estudios Geoestadísticos en forma independiente. El primer diseño consistió en dividir el área total en 5 partes, ver la figura 6-4, y efectuar estudios Geoestadísticos independientes en cada partición. El segundo diseño consistió en dividir el área total en 3 partes, ver figura 6-16.



Este es el mapa de la ciudad de Cuernavaca donde se obtuvieron las 643 observaciones.
Figura 6-1

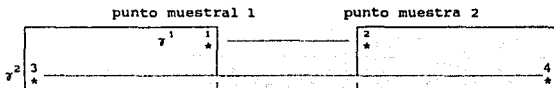
Supóngase que se tienen los terrenos A y B, y se eligieron los puntos muestrales de cada terreno de tal forma que la distancia entre las muestras es la mínima, figura 6-2. Se puede asegurar el hecho de que la correlación espacial entre los puntos 1 y 2 es diferente a la correlación entre los puntos 3 y 4. Esto se debe a que la distancia entre los puntos 1 y 2 es menor que la distancia entre los puntos 3 y 4.

Este hecho afecta la dependencia entre los datos, por lo tanto el Variograma sería distinto; por ejemplo:

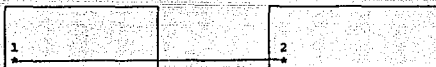
Supóngase que se estima el variograma γ^1 , el cual considera las distancias más grandes entre cada pareja de terrenos; de igual forma se estima el variograma γ^2 utilizando la misma información pero en este caso se considera la distancia mínima entre cada pareja de terrenos. La siguiente gráfica muestra cómo la interpolación espacial cambia debido a la variación de la distancia entre los puntos.

El variograma γ^2 es distinto al variograma γ^1 , ya que el efecto pepita asociado a γ^1 es más grande en proporción a la meseta y el efecto pepita del variograma γ^2 . Esto implica que la interpolación espacial realizada por el método Kriging utilizando el variograma γ^1 es distinta a la interpolación espacial utilizando el variograma γ^2 .

En la práctica esto quiere decir que el volumen estimado $Z(x_0)$ es distinto utilizando el patrón muestral de los puntos 1 y 2 (muestras cercanas), que el patrón muestral de los puntos 3 y 4 (muestras alejadas).



La dependencia entre los terrenos varía de acuerdo a la elección del metro z elegido como punto muestral.
Figura 6-2.



En cada terreno se eligió como punto de muestreo, la esquina inferior izquierda en el mapa de la ciudad o el metro que se encuentre al suroeste de los terrenos.
Figura 6-3.

Programa Geoestadístico

Se utiliza el programa Geo-EAS, el cual es un paquete de análisis Geoestadístico escrito en Fortran 77, para encontrar la función variograma, para resolver el sistema de las ecuaciones Kriging y para encontrar la medida del error de la técnica validación cruzada.

Restricciones del programa Geo-EAS

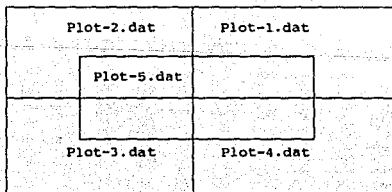
Debido a que la subrutina Prevar del programa Geo-EAS solamente acepta un máximo de 180 datos, el área total de estudio formado por 693 datos se dividió en 5 partes analizar posibles anisotropías en la intersección de las particiones como se muestra en la figura 6-4. Cada archivo se analizará en forma independiente.

La rutina Prevar produce un archivo de salida formado por el número total de parejas muestrales de n datos. El máximo número de parejas que puede formar es 16384 debido a que la siguiente subrutina llamada Vario no puede analizar más de este número máximo de parejas. El límite impuesto al número de datos del archivo de Prevar es: $(N^2 + N)/2 = 16384$. De donde $n = (32,768.25)^{1/2} - 1/2 = 180.53$. Por esta razón se dividió el área total de 693 datos en 5 archivos para su estudio.

En la tabla 6-1 se muestra el nombre de cada archivo de datos con su respectivo número de datos y la parte que corresponde del área total.

nombre del archivo	No. de partición	No. de datos	área de partición
plot-1.dat	1	179	nor-oeste
plot-2.dat	2	172	nor-este
plot-3.dat	3	178	sur-oeste
plot-4.dat	4	221	sur-este
plot-5.dat	5	194	centro

Tabla 6-1.



El área de estudio se dividió en 4 partes mas una parte central.
Figura 6-4.

Hipótesis sobre el Modelo

Los datos se consideran como obtenciones de un proceso estocástico estacionario de segundo orden, en particular utilizamos la hipótesis intrínseca aplicada a $Z(x)$. Se supone que los datos $\{Z(S_1), Z(S_2), \dots, Z(S_{179})\}$, son observaciones tomadas "sin ninguna medida de error".

Análisis Geoestadístico utilizando Geo-EAS

A continuación se presenta las rutinas del paquete Geo-EAS que se utilizan para desarrollar un estudio Geoestadístico simple.

Programa Geo-EAS (menú principal):

- 1) Rutina Postplot: Gráfica de la distribución de la información
 - 2) Rutina Stat1: Estadísticas univariadas
 - 3) Rutina Prevar: Reducción de pares muestrales
 - 4) Rutina Varic: Estimación de las distancias de clase, dirección y ajustes de los parámetros del variograma
 - 5) Rutina Krige: Cálculo de las ecuaciones kriging
 - 6) Rutina Conrec: Gráfica de contornos, curvas de nivel, y gráficas de la medida del error
 - 7) Rutina Xvalid: Estimación de los valores en cada localidad muestral utilizando la técnica validación cruzada.
-

6.1 DISEÑO 1 DEL ESTUDIO GEOESTADISTICO

Como se vió anteriormente existen algunas restricciones las cuales obligan a hacer particiones del grupo de datos. El procedimiento Geoestadístico que se utilizará es el siguiente.

- 1.- Desarrollar un análisis Geoestadístico independiente para cada archivo de datos correspondiente.
 - 2.- Describir en forma detallada el análisis Geoestadístico para el archivo de datos plot-1.dat.
 - 3.- Presentar los resultados de plot-2.dat, plot-3.dat, plot-4.dat y plot-5.dat con sus respectivas interpretaciones.
 - 4.- Obtener los resultados de la bondad de ajuste a través de la técnica validación cruzada para cada archivo de datos con objeto de encontrar las zonas anisótropas.
 - 5.- Graficar las curvas de nivel del área estimada.
-

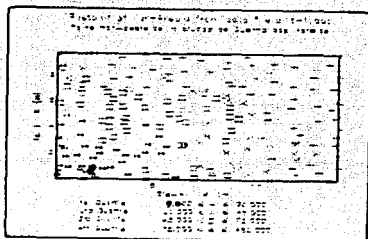
6.2 ANALISIS GEOESTADISTICO DEL ARCHIVO PLOT-1.DAT

El análisis Geoestadístico consiste en estudiar el grupo de 179 datos del archivo Plot-1.dat, que corresponden al área 1 de la figura 6-1. Para desarrollar el estudio Geoestadístico a continuación se muestra como se utiliza el programa Geo-EAS.

6.2.1 EL ARCHIVO DE DATOS

Primero se tiene que conocer como se distribuye el grupo de datos Plot-1.dat. En este caso, se utiliza la subrutina llamada Postplot, la cual permite graficar las localidades de las muestras, y entonces se puede apreciar la distribución espacial en el área de estudio del grupo de datos. Esta rutina clasifica los datos por cuartiles.

El Objetivo de la rutina Postplot es analizar por medio de una gráfica el grupo de datos para localizar posibles datos aberrantes, patrones de dispersión, o zonas de altas de concentración de datos.



Gráfica Postplot del archivo Plot-1.dat
Figura 6-5.

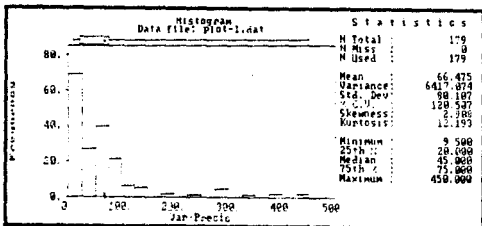
Las coordenadas (x,y) están dadas en kilómetros y el centro de la ciudad se considera como el origen de un sistema cartesiano de puntos en $D \subset R^2$. El área de muestreo es el espacio de localidades D es: $\{s_i : i = 1, \dots, n\}$, con $s_i \in D \subset R^2$.

En la figura 6-5, se puede apreciar el comportamiento de la dispersión de los datos de Plot-1.dat, se nota la acumulación de valores altos en la parte sur-oeste y de valores bajos a lo largo de la parte norte y norte-este.

El Variograma se calculó en la dirección oeste-este a 0° con respecto al eje X. Los valores más altos están marcados con *, los cuiles llegan a ser hasta 5 veces más grandes que los valores σ . Estos dos grupos están relativamente cercanos e incluso existen áreas en donde los valores de ambos grupos se mezclan. Esto va a implicar una varianza mayor en esta zona.

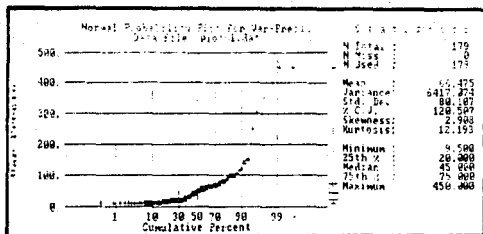
6.2.2 ESTADÍSTICAS UNIVARIADAS

Ahora se van a generar algunas estadísticas del grupo de datos. La rutina Stat1 contiene opciones para graficar histogramas y acumulación de probabilidades.



Histograma del archivo Plot-1.dat
Figura 6-6.

Como se puede apreciar, el histograma tiene una frecuencia muy alta de valores $Z(x)$'s. menores que μ . Este hecho también se puede apreciar en la gráfica de probabilidades.



Gráfica de las probabilidades acumuladas con
distribución Normal.

Figura 6-7.

Aquí, se grafica sobre la abscisa el porcentaje acumulativo de aparición de cada observación contra el valor precio que toma cada observación. Se puede apreciar fácilmente la acumulación de los valores menores que la media $\mu = 66.47$. El 80% de los datos son menores que 100. El valor máximo es 450 y el mínimo es 9.5.

6.2.3 ANALISIS DEL VARIOGRAMA

La estimación e interpretación del modelo variograma es la parte esencial de cualquier estudio Geostatístico.

Todos los métodos de interpolación espacial necesitan algún tipo de correlación entre los datos para explicar las propiedades físicas del fenómeno. El modelo variograma define la estructura de correlación espacial del grupo de datos y sirve para que el estimador Kriging asigne pesos a las muestras durante el proceso de interpolación. Si la asignación de los pesos es correcta entonces existirá calidad en los resultados, ya que la interpolación será adecuada.

Lo que se quiere saber es qué tan confiable es la medida de dependencia entre los datos disponibles para representar una nueva localidad hacia otra distancia y dirección.

Se intentarán utilizar Variogramas experimentales distintos hasta encontrar el que sea apropiado a este caso. A continuación se utiliza la rutina Prevar, la cual genera un archivo llamado Plot-1.pcf. Este archivo está formado por las distancias entre parejas de datos, donde se clasifican por grupos de acuerdo a la distancia de cada muestra contra todas las restantes.

6.2.4 CALCULO DE LAS PAREJAS MUESTRALES

El programa Prevar acepta el archivo de entrada llamado Plot-1.dat y genera el archivo de salida llamado Plot-1.PCF. Este último es el archivo de comparación de parejas y contiene información concerniente al número total de parejas muestrales separadas a ciertas distancias.

Esta rutina calcula el número de parejas de acuerdo a las distancias de separación para después grabarlas en un archivo de salida que será usado por la rutina Vario.

6.2.5 PARAMETROS DEL VARIOGRAMA

La rutina Vario modela y analiza un variograma bi-dimensional isotrópico. El objetivo de Vario es encontrar los parámetros del Variograma, los cuales asignan un peso a cada muestra en el proceso de interpolación. La forma de encontrar a los parámetros efecto pepita, techo y rango es la siguiente:

Primero.- Se especifica el vector h , es decir tenemos que definir cómo se divide el vector h de tal forma que dejemos muchas parejas muestrales a distancias cortas.

Segundo.- Se Tiene que especificar un criterio de selección de la orientación de los pares muestrales; es decir, la dirección del vector h y su tolerancia. Para mayor información consulte el manual de usuarios del programa Geo-EAS.

Tercero.- Se desea ajustar un modelo variograma isotrópico conocido a la descripción de la gráfica del variograma. De esta forma se pueden encontrar los parámetros efecto pepita, meseta y rango.

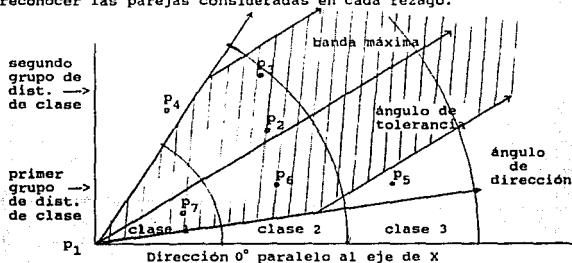
La dirección determina el criterio de selección de las parejas muestrales, Myers (1987). Para este caso se utilizó una dirección de h de 0° paralela al eje X.

En la figura 6-8 se ilustra la forma en que se eligen los rezagos. En donde la dirección tiene un valor de cero grados. La dirección es un parámetro numérico muy importante en la determinación de las anisotropías, la dirección se mide en grados (o radianes) y aquí se va a considerar que la dirección tiene un rango que oscila entre 0 y 180 grados.

La tolerancia es un valor trigonométrico en grados que tiene un rango entre cero y 90°. Esta opción sirve para definir el ancho de la banda que incluye los datos dentro de este rezago.

La tolerancia es positiva y negativa. Es decir, la tolerancia va a tener una amplitud del doble de su valor con respecto a la línea central de dirección del vector h . Por ejemplo; si la dirección es de 50° y la tolerancia es 10°, la amplitud que cubre la zona de elección de parejas estará entre los ángulos 40° y 60°. Para el caso del archivo plot-1.dat la tolerancia que se eligió es de 90°.

El valor banda máxima es la máxima distancia perpendicular entre la línea central marcada por la dirección y una línea paralela a ella que intersecta al segundo punto de la pareja. La figura 6-8 ilustra cómo se eligen los rezagos, la tolerancia y banda máxima. Es fácil reconocer las parejas consideradas en cada rezago.



En esta figura se muestra la forma de como se incluyen las parejas muestrales en cada rezago.

Figura 6-8.

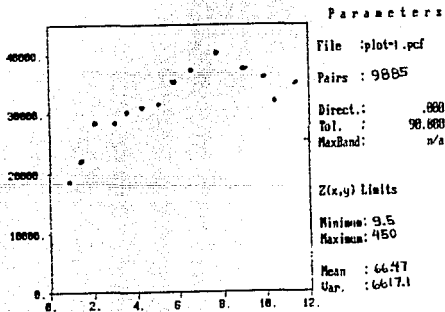
Es necesario experimentar con las distancias de clase para obtener una gráfica más suave. Myers (1989) explica este hecho. Warrick y Mayers (1986), han demostrado que es posible intentar varios ángulos e incluso forzar el número de parejas para cada distancia de clase de tal forma que se aproxime a la distribución real. En otras palabras, se puede ajustar los rezagos en forma individual para que la distribución se aproxime a la requerida. Lo que se desea es tener un número alto de parejas muestrales rezagos pequeños; Myers (1987).

En la tabla 6-2 se muestran las distancias de clase, llamados también rezagos y la dirección de vector h .

Estadística		Valor	
Media		66.47	
Desviación Estándar		66.17	
Tamaño de muestra	9885		
Intervalo de confianza			
Intervalo Superior		132.64	
Intervalo Inferior		-0.10	
Media Superior		132.64	
Media Inferior		-0.10	
Max. Sup.	9885		
Max. Inf.	1		
Incremento	50		

Resaca, dirección y tolerancia
Tabla 6-2.

La siguiente figura 6-9 muestra el variograma $2\gamma(h)$ para 10 grupos de distancias de clase $\{h(1)e, h(2)e, \dots, h(10)e\}$; donde el vector e tiene una longitud de $|e| = 1$ (kilómetro), y una dirección de 0 grados (eje-X).

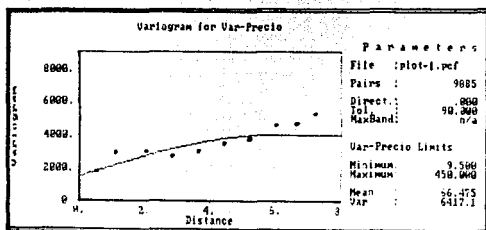


Variograma del Archivo Plot-1.dat
Figura 6-9.

6.2.6 MODELOS VARIOGRAMA

Se experimentó con un modelo lineal y un modelo esférico, sin embargo no hubo diferencias notables para elegir alguno de ellos. Se decidió utilizar un modelo variograma esférico.

El Variograma esférico es: $\gamma(h, \lambda) = C_0 + C [3/2 - 1/2 h/a]^3$,
donde $\lambda = (C_0, C, a) = (1500, 2500, 6)$, y $h = (h_1, h_2, \dots, h_{10})$.



Ajuste de un modelo Variograma a la gráfica del variograma experimental
Figura 6-10.

Ahora se pueden calcular las ecuaciones Kriging basadas en (5.4) y los mapas basados en las ecuaciones (5.3) y (5.5) ser graficadas.

Un rezago es un conjunto de datos que tienen la cualidad de mantener la misma distancia entre ellas, pero con cierta tolerancia, ver el área de tolerancia de la figura 6-8.

Los grupos de rezagos se eligieron según el siguientes criterio.

- 1.- La distancia mínima entre las parejas muestrales es 0.125 Km.
- 2.- La distancia máxima de los rezagos es 10 Kms.
- 3.- El incremento entre rezagos es de 1 Km.

La idea consiste en tratar de obtener el máximo de detalle en el variograma para distancias pequeñas. Sin embargo, Myers (1987) explica que siempre habrá una distancia mínima entre las muestras, en donde no se podrá estimar el variograma $\gamma(h)$ para distancias menores a este mínimo. Esto implica que γ no se estima en este intervalo. La varianza Kriging considera esta incertidumbre dentro del efecto pepita como también considera los errores de muestreo.

Para patrones de localidades regulares o casi regulares, el número de parejas separadas a una distancia corta es pequeña, grande para distancias medianas y decrece para distancias que excedan la mitad del diámetro de la región de interés. El variograma no evalúa las distancias de clase mayores a la mitad del diámetro del panel en estudio. La rutina Vario se basa en este hecho por default y divide el vector h en subintervalos iguales h_q con objeto de calcular el variograma en cada subintervalo.

$$\gamma^*(h_q) = \sum_{i=1}^n [Z(x_i+h_q) - Z(x_i)]^2$$

A continuación se ajustó un modelo variograma a la descripción de la dependencia entre los 179 datos de Plot-1.dat.

Parámetros del Variograma

Los parámetros del variograma que se encontraron son los siguientes:
Un efecto pepita de 1500.
Un meseta de 2500.
Un rango de 6.

6.2.7 METODO KRIGING ORDINARIO

La rutina Krige calcula las ecuaciones kriging (5.12) los parámetros del variograma obtenidos en Vario con objeto de asignar pesos a las muestras. Para predecir el valor de Z^* en una localidad nueva x_0 .

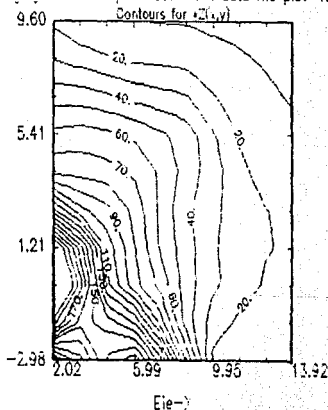
$$Z^*(x_0) = \sum_{i=1}^{129} \lambda_i Z(x_i)$$

Una vez que se calculen las ecuaciones kriging estamos en posibilidad de graficar los contornos de las curvas de nivel y de poder checar la áreas en donde la varianza de la dispersión es mayor.

6.2.8 PREDICCIÓN ESPACIAL

Conrec es una rutina que produce mapas de curvas de nivel de la variable aleatoria $Z(x)$. En la siguiente figura se aprecian las curvas de nivel que se desplazan en la parte sur-oeste, la cual es la parte que corresponde a observaciones con un volumen alto.

Kriging estimates produced from data file plot-1.gr



Gráfica de las curvas de nivel utilizando las rutina Conrec. y aplicado al archivo Plot-1.dat
Figura 6-11.

6.2.9 VARIANZA DE LA DISPERSION

La rutina Xvalid efectúa el proceso de validación cruzada, la cual es una técnica que elimina un valor de los datos y estima ese punto utilizando los datos restantes. Este proceso se repite n veces con objeto de comparar qué tan cerca está el valor observado del estimado.

Interpretación de Validación Cruzada

En la primer columna se tienen los valores de los cuartiles de las observaciones:

- 1.- Primer cuartil: $9.5 \leq X \leq 20.0$
- 2.- Segundo cuartil: $20.0 < X \leq 48.0$
- 3.- Tercer cuartil: $48.0 < X \leq 76.0$
- 4.- Cuarto cuartil: $76.0 < X \leq 450.0$

El valor mínimo es 9.5, el 25% superior es 20.0, la mediana es 48.0, el 75% superior es 76.0, y el máximo valor es 450.

Estos son los cuartiles de las estimaciones:

- 1.- Primer cuartil: $9.648 \leq X \leq 21.299$
- 2.- Segundo cuartil: $21.299 < X \leq 43.158$
- 3.- Tercer cuartil: $43.158 < X \leq 74.193$
- 4.- Cuarto cuartil: $74.193 < X \leq 331.190$

Ahora se analizará el error de estimación el cual es producido por la diferencia $Z(x) - Z(\hat{x})$. El objetivo ya no es minimizar el error de la estimación sino identificar las áreas de la varianza entre el observado y el estimado, como se verá en los siguientes dos casos.

Caso 1.- Cuando el error de estimación es mayor o igual a cero:

$$Z(x) \geq Z(\hat{x}).$$

Caso 2.- Cuando el error de estimación es menor que cero:

$$Z(x) < Z(\hat{x}).$$

El objetivo de las siguientes gráficas es localizar las zonas de varianza de la dispersión utilizando los resultados de las gráficas del error obtenidas por la validación cruzada. La gráfica de la figura 6-12B muestra con una cruz (+) el caso 1, y con una (x) el caso 2. El tamaño del error se representa en forma proporcional con el tamaño del símbolo. Si el error es grande el símbolo será grande y viceversa.

A continuación se muestra las gráficas de los errores de la predicción hecha por las ecuaciones kriging (figuras 6-12A y 6-12B).

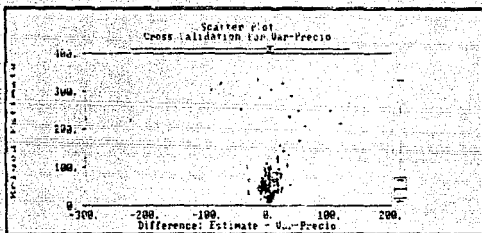


Gráfico de los errores del observado-estimado contra estimado

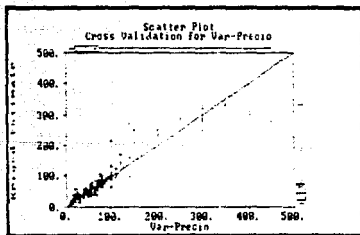
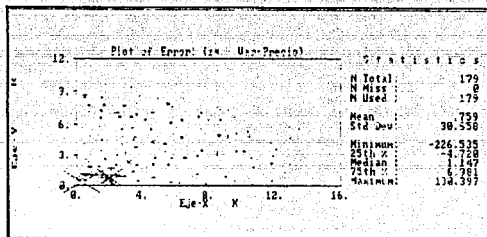


Gráfico de los errores, el observado (eje X) contra el estimado
Figura 6-12A.



Gráfica de los errores del archivo Plot-1.dat, la cual corresponde a la parte nor-este de la ciudad de Cuernavaca.
 Figura 6-120.

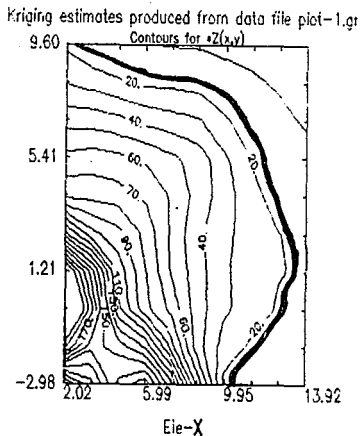
La \times indica que el valor observado es menor que el estimado y la $+$ indica el caso contrario. El tamaño de la $+/\times$ representa la dimensión del error.

Identificación de zonas de varianza de la dispersión

En esta gráfica se pueden apreciar las zonas en donde la varianza de la dispersión es mayor. Se sabe que existen anisotropías conforme se avanza de la parte norte-este en dirección sur-oeste. Al alcanzar esta porción de la gráfica las estimaciones registran un valor máximo de 331.190.

6.2.10 INTERPRETACION DE LOS RESULTADOS KRIGING

La interpretación de los resultados de las estimaciones del método Kriging se hizo analizando las curvas de nivel de la gráfica anterior. A continuación se muestra las curvas de nivel y se marca por donde pasa la autopista México-Acapulco.



Gráfica de las curvas de nivel y de la parte nor-este de Cuernavaca
Figura 6-13.

También se pueden distinguir las siguientes características:
 1.- La curva de nivel correspondiente a un valor de $Z(x) = 30$ se desplaza por encima de la línea que marca la autopista de México a Acapulco. Del lado derecho de dicha curva se encuentran terrenos de siembra cuyos valores por metro cuadrado son inferiores a los 15 dólares y cuyas dimensiones exceden a los 100, 000 metros cuadrados. En cambio del lado izquierdo el valor de los terrenos se incrementan entre 30 y 40 dólares el metro cuadrado.

Ahora se quiere saber qué tan dispersas están las estimaciones en relación a los valores reales, aunque los valores reales no se conocen. Se van a utilizar las observaciones como un punto de referencia. Bajo la suposición de que el muestreo se obtuvo sin ninguna medida de error. La técnica validación cruzada nos va a dar una idea de la bondad de ajuste del interpolador.

RESULTS				
Data File	c:\geog\data\plot-1.dat			
X Variable	Elev-X			
Y Variable	Elev-Y			
Ordering	Variable to	# data used	179	
Type	Ordinate	# missing data	0	
Variable	Estimate	Difference	Kriging Std	
Minimum	9.500	9.648	-226.535	42.687
25th %tile	20.000	21.299	-4.720	45.867
Median	45.000	43.158	1.147	44.736
75th %tile	75.000	74.193	6.981	47.575
Maximum	450.000	331.190	130.397	55.640
N	179	179	179	179
Mean	66.475	67.234	.759	46.870
Std. Dev.	80.107	72.230	30.658	1.658

Estos son los resultados de la rutina validación cruzada. Esta tabla muestra los cuantiles de las observaciones y estimaciones
 Tabla b-3.

Conclusiones de Plot-1. dat

Los resultados de esta comparación muestran que la variabilidad del fenómeno precios de terrenos tiene un patrón de concentración en donde los valores altos se agrupan en áreas comerciales y residenciales. Los valores máximos se agrupan en el centro de la ciudad.

Se pueden localizar las áreas de varianza de la dispersión en la figura 6-12A y 6-12B. Esta varianza de la dispersión se puede interpretar de varias formas ya que existen muchas razones por las cuales el valor observado sea muy diferente del valor estimado. Los errores pueden ser causados por varias razones, a continuación se mencionan algunas de ellas.

1.- El tamaño y la forma del terreno influye notablemente en el variograma y en el método Kriging. Además que los terrenos más chicos están en el centro de la ciudad y los terrenos más grandes están en las afueras de la ciudad.

2.- Es posible que existan errores en la captación de la información o que el método de muestreo que se utilizó no haya sido el adecuado.

3.- Las observaciones no necesariamente representan el valor real del precio del terreno, ya que el precio está sujeto a la oferta, y demanda entre otros factores.

4.- De la gráfica de Postplot se pueden detectar valores extremadamente altos en la parte sur-oeste. Evidentemente, existen anisotropías en esta dirección ya que la varianza de la dispersión es grande.

5.- Estas anisotropías provocan distorsiones del volumen de las muestras que a su vez son interpretadas por el efecto pepita en el variograma.

6.- No se tiene un número grande de parejas muestrales separadas a distancias pequeñas. Esto implica que la interdependencia espacial para este tipo de distancias no es muy fuerte. Además, el variograma no se puede estimar para distancias menores.

En general, se puede concluir que los resultados de las estimaciones que se presentaron en esta sección nos dan una idea de la fluctuación de los precios de terrenos en la parte nor-este de la ciudad de Cuernavaca.

6.3 RESULTADOS DE LOS ARCHIVOS PLOT-2.DAT, PLOT-3.DAT, PLOT-4.DAT Y PLOT-5.DAT

Ahora se va a ver los resultados del analisis Geoestadístico de los archivos 2,3,4, y 5 utilizando la rutina Xvalid y Conrec.

Interpretación del área total de muestreo

El programa Postplot puede graficar hasta 1000 datos y con un mínimo de 3 variables y un máximo de 48 variables. Para nuestro propósito, no tenemos problema alguno en graficar los 693 datos.

Los valores más altos se encuentran concentrados en las curvas de nivel de la parte central del mapa área total ver figura 6-14. Esta área alcanza valores entre 400 y 610 dólares por metro cuadrado estos están marcados con un asterisco (*) en la figura 6-5.

En la figura 6-15 se pueden localizar las áreas más altas correspondientes a zonas residenciales y comerciales como por ejemplo;

Alrededor de las coordenadas (0, 0) en un círculo con diámetro de 2 kilómetros se encuentra el centro de la ciudad; esta área tiene asignada la etiquetada A en la figura 6-15. En la zona sur del centro como a 6 kilómetros se encuentra el fraccionamiento residencial Tabachines localizado alrededor de las coordenadas (1.70, -8.55) y en un diámetro aproximado de un kilómetro, esta zona está etiquetada con la letra B. En la parte este se encuentra el centro comercial La Luna, localizada en las coordenadas (6.50, -3.75) y en un círculo aproximado de 1 kilómetro de diámetro y etiquetado con la letra C.

Existen otras zonas residenciales como son: El Club de Golf en las coordenadas (-2.30, -2.40), el Residencial Bondies, el residencial Palmira ubicados en una extensa superficie que pasa por (-1, -7) ubicada en la parte sur de la ciudad. La zona residencial Tabachines es una de las áreas más caras, la cual se encuentra ubicada en las coordenadas (-3, -3). El área residencial Lomas Cuernavaca al sur de Tabachines es una extensa área que cubre 3 kilómetros de diámetro y cuyo centro está ubicado en las coordenadas (-3.50, -13.30). El fraccionamiento Sumiya ubicado en las coordenadas (1.55, -12.70) se extiende en una área circular de 2 kilómetros de diámetro aproximadamente.

Variogramas

Los variogramas de las diferentes áreas se presentan en forma de tabla:

Nombre del archivo	número de datos	modelo variograma	efecto pepita	techo	rango
Plot-1.dat	179	esférico	1500	25000	6
Plot-2.dat	172	lineal	12000	17000	10
Plot-3.dat	178	exponencial	1900	7300	4
Plot-4.dat	221	lineal	2800	3200	10
Plot-5.dat	194	exponencial	4000	19000	6

Tabla 6-4

El siguiente paso es correr el programa Krige para resolver las ecuaciones kriging para cada uno de los archivos 2, 3, 4, y 5. Al finalizar esta rutina tenemos que checar qué tan buenas fueron las estimaciones para ello se corre la rutina de validación cruzada, Conrec.

Resultados de Xvalid

Los resultados de validación cruzada de los archivos 2,3,4, y 5 los presentamos a continuación:

RESULTS

DATA FILE	: C:\JPB\DAT\PLOT-2.DAT			
X VARIABLE	: EJE-X			
Y VARIABLE	: EJE-Y			
KRIGING	: Z(X,Y)			# DATA USED : 172
TYPE	: ORDINARY			# MISSING DATA : 0
	VARIABLE	ESTIMATE	DIFFERENCE	KRIGING STD
MINIMUM	15.000	26.151	-315.357	117.846
25TH %TILE	45.000	51.859	-9.817	120.787
MEDIAN	75.000	76.682	2.904	122.157
75TH %TILE	150.000	218.710	17.013	122.761
MAXIMUM	610.000	489.190	352.921	133.714
N	172	172	172	172
MEAN	135.599	142.057	6.458	122.201
STD. DEV.	145.790	124.805	75.664	2.189

Resultados de validación cruzada del archivo Plot-2.dat
Tabla 6-5

RESULTS

DATA FILE	: C:\JPB\DAT\PLOT-3.DAT			
X VARIABLE	: EJE-X			
Y VARIABLE	: EJE-Y			
KRIGING	: Z(X,Y)			# DATA USED : 178
TYPE	: ORDINARY			# MISSING DATA : 0
	VARIABLE	ESTIMATE	DIFFERENCE	KRIGING STD
MINIMUM	12.000	19.956	-546.509	58.803
25TH %TILE	35.000	49.318	-7.060	68.600
MEDIAN	65.000	69.345	5.714	72.216
75TH %TILE	100.000	109.621	17.901	76.653
MAXIMUM	610.000	340.949	229.986	89.261
N	178	178	178	178
MEAN	91.816	94.410	2.593	72.591
STD. DEV.	93.916	70.517	54.055	6.153

Resultados de validación cruzada del archivo Plot-3.dat
Tabla 6-6

RESULTS

DATA FILE	: C:\JPB\DAT\ PLOT-4.DAT			
X VARIABLE	: EJE-X			
Y VARIABLE	: EJE-Y			
KRIGING	: Z(X,Y)		# DATA USED	: 221
TYPE	: ORDINARY		# MISSING DATA	: 0
	VARIABLE	ESTIMATE	DIFFERENCE	KRIGING STD
MINIMUM	12.000	12.949	-107.347	57.262
25TH %TILE	15.000	21.302	-4.067	57.918
MEDIAN	30.000	35.837	2.762	58.167
75TH %TILE	60.000	58.140	10.030	58.545
MAXIMUM	400.000	300.092	129.654	64.424
N	221	221	221	221
MEAN	53.701	56.126	2.425	58.331
STD. DEV.	69.790	58.088	33.284	.772

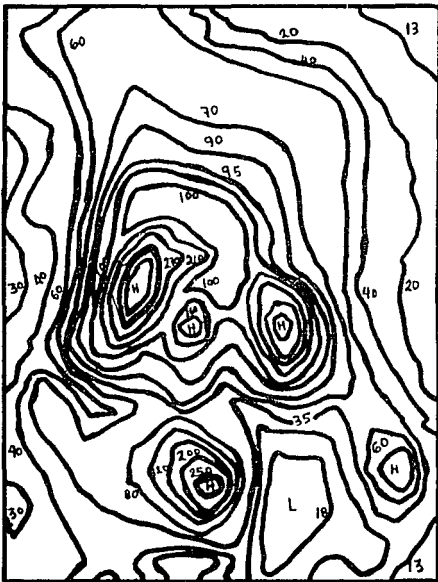
Resultados de validación cruzada del archivo Plot-4.dat
Tabla 6-7

RESULTS

DATA FILE	: C:\JPB\DAT\ PLOT-5.DAT			
X VARIABLE	: EJE-X K			
Y VARIABLE	: EJE-Y K			
KRIGING	: Z(X,Y)		# DATA USED	: 194
TYPE	: ORDINARY		# MISSING DATA	: 0
	VARIABLE	ESTIMATE	DIFFERENCE	KRIGING STD
MINIMUM	15.000	33.881	-297.789	78.250
25TH %TILE	75.000	83.383	-11.554	93.958
MEDIAN	100.000	126.851	4.727	98.078
75TH %TILE	200.000	218.289	24.497	102.975
MAXIMUM	610.000	558.211	247.283	122.750
N	194	194	194	194
MEAN	160.938	165.003	4.065	98.739
STD. DEV.	138.094	110.216	63.817	7.742

Resultados de validación cruzada del archivo Plot-5.dat
Tabla 6-8

Con objeto de predecir algún valor $Z(x_0)$ conociendo las coordenadas x_0 , se presentan las gráficas de Conrec. Estas gráficas muestran las curvas de nivel de los precios estimados. Las 5 gráficas de la partición mostrada en 6-4 se presentan unidas en la estimación de sus curvas de nivel en la figura 6-14.

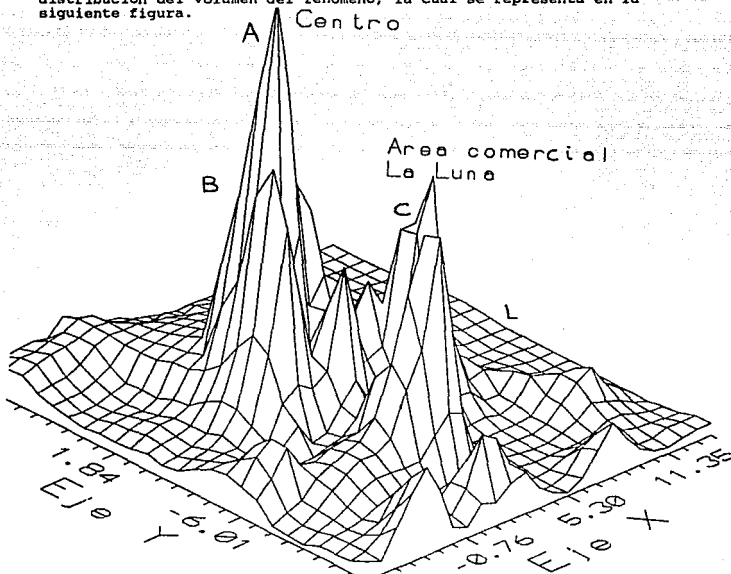


Gráficas de las curvas de nivel producidas por el programa Conrec de los archivos Plot-1.dat, Plot-2.dat, Plot-3.dat, Plot-4.dat y Plot-5.dat.
 Figura 6-14.

Si se conoce la posición de algún terreno en la superficie D se puede estimar el precio utilizando las gráficas de las curvas de nivel. En general las diferencias de $Z(x) - Z'(x)$ son en promedio las mismas, sin embargo existen algunas diferencias grandes en ciertas áreas, lo que tenemos que hacer es identificar las zonas donde la varianza de la dispersión es mayor. Las gráficas siguientes ayudan a localizar las áreas de varianza de la dispersión.

6.4 DISEÑO 2 DEL ESTUDIO GEOESTADISTICO

Vamos a utilizar el hecho de que se conoce la forma de la distribución del volumen del fenómeno, la cual se representa en la siguiente figura.



Estimación de la variabilidad del volumen precios por metro cuadrado de la ciudad de Cuernavaca utilizando la rutina Acrospin.

Figura 6-15.

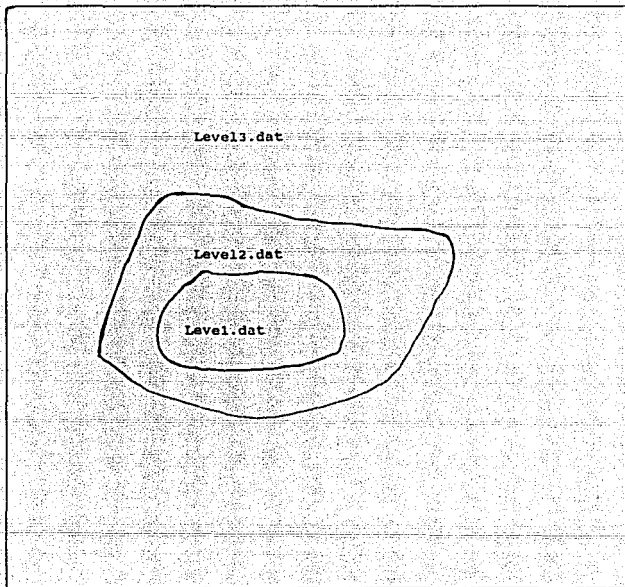
El diseño 2 consiste en elegir tres particiones del área total de la ciudad de Cuernavaca. La primer partición es la parte central de la figura 6-16, está etiquetada con la letra A, su archivo de datos lo llamaremos Level1.dat, en esta zona encontramos terrenos con dimensiones relativamente pequeñas entre 100 y 1500 metros cuadrados.

Generalmente, los terrenos más chicos se encuentran en esta zona de la ciudad y en su mayoría son considerados como locales comerciales ya que casi todos están bardados y techados. Este tipo de terrenos no se venden por metro cuadrado sino que se vende toda la propiedad, la cual depende de su ubicación. Los locales comerciales no se consideran en este estudio.

La siguiente zona corresponde al area residencial Tabachines, la cual está etiquetada con la letra B y su archivo de datos se llamará Level2.dat. La última partición es la zona correspondiente a las afueras de la ciudad y a esta zona se llamará Level3.dat, cuya área incluye la etiqueta C.

El archivo Level1.dat esta formado por 64 datos. El segundo archivo llamado Level2.dat tiene la forma de una dona y esta formado por 216 datos ubicados afuera del centro de la ciudad pero dentro de los límites de la ciudad de Cuernavaca. Este archivo excluye los datos de Level1.dat y los de Level3.dat.

El archivo Level3.dat está compuesto por 268 datos ubicados a las afueras de la ciudad. Es decir que de los 693 datos excluimos los datos de Level1.dat y Level2.dat. A continuación mostramos la partición elegida por el diseño 2.



Partición del Área total formada por tres archivos:
Level1.dat, Level2.dat y Level3.dat.
Figura 6-16.

6.4.1 ANALISIS GEOESTADISTICO DEL ARCHIVO LEVELL.DAT

Este archivo está compuesto de 64 muestras tomadas en el centro de la ciudad. En la gráfica 6-16, se muestra al área donde se obtuvieron las observaciones de Level1.dat correspondientes a la parte superior del cerro etiquetado con la letra A de la figura 6-15.

En la gráfica 6-17, donde se muestran los patrones de distribución de los datos del archivo level1.dat, esta gráfica se obtuvo a través de correr la rutina Postplot.

Se puede apreciar si hay algún patrón de distribución muestral. Se nota que los datos más altos se desplazan en la dirección suroeste - noreste. Estos puntos marcan una anisotropía corriente en dicha dirección.

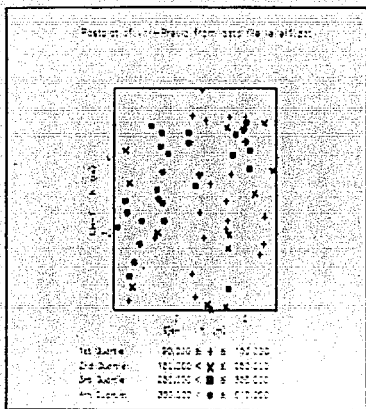
En la gráfica 6-18 los 2 últimos puntos marcados con el número 1 y 2 corresponden a terrenos muy transitados por peatones; en cambio los puntos etiquetados con los números 3, 4, 5, 6, y 7 que se encuentran cerca corresponden a terrenos baldíos.

Se nota que existen dos valores posiblemente aberrantes los cuales son los puntos 4 y 5. Se decidió investigar el porque de tal anisotropía. Una forma de hacerlo fue visitar estos lugares para poder llegar a una conclusión. Los dos terrenos están ubicados en el corazón del centro de Cuernavaca.

El punto 4 corresponde a un terreno ubicado en un calle donde hay una serie de vecindades de mal aspecto y por dicha calle no transita la gente. Es por eso que este terreno pierde su valor comercial.

El punto 5 corresponde a un lote que se encuentra ubicado entre una cantina y la barranca que pasa cerca del mercado central Adolfo López Matéos, esto ocasiona que el precio por metro cuadrado baja y por lo tanto forma otra anisotropía en esta dirección.

A continuación se analiza la gráfica de la distribución espacial de los datos del primer archivo llamado Level1.dat con objeto de detectar posibles patrones muestrales.



Gráfica Postplot del archivo Level1.dat.
Figura 6-17.

El variograma que se utilizó para este archivo fue un modelo esférico con una meseta de 25000, un rango de 2.5 y un efecto pepita de 7000. Una vez que se ajustó un variograma esférico al variograma experimental se resolvieron las ecuaciones Kriging. Posteriormente, se cuantificó la bondad de ajuste utilizando la rutina Xvalid.

La tabla 6-9 muestra los resultados de la rutina Xvalid del archivo Level1.dat, donde se aprecian el valor de la variable, la diferencia entre el observado menos el estimado, y la desviación estándar por cuartiles.

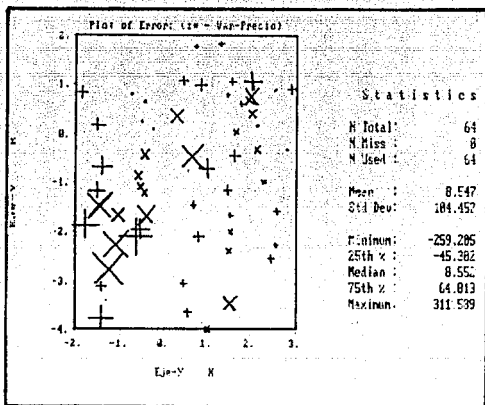
En el primer cuartil se nota que la banda de confianza para cada lado es de 102.559 y en el cuarto cuartil es de 170.734. Esto indica que para valores grandes la desviación estándar aumenta y que dichos valores altos están cerca de valores bajos.

En la práctica los valores altos de los precios por metro cuadrado de cualquier ciudad tienden a subir debido a la influencia de los valores altos vecinos. Si la demanda por terrenos no es muy buena, entonces el precio baja. En cambio si la demanda se incrementa el precio se eleva.

RESULTS

DATA FILE :	D:\GEO\LEVEL1.DAT			
X VARIABLE :	PJF-X	K		
Y VARIABLE :	LJI-Y	K		
KRIGING :	VAR-PRICE		# DATA USED :	64
TYPE :	ORDINARY		# MISSING DATA :	0
##### VARIANTE ESTIMATI DIFFERENCIA KRIGING STD #####				
MINIMUM	90.000	120.059	-259.205	102.559
25TH QUANTILE	150.000	157.252	-69.502	118.650
MEDIAN	250.000	258.505	-6.505	125.620
75TH QUANTILE	350.000	352.965	64.965	132.631
MAXIMUM	610.000	562.651	511.589	170.734
N	64	64	64	64
MEAN	279.688	288.157	18.507	126.004
STD. DEV.	162.372	112.525	100.052	12.000

Resultados de la Rutina Xvalid del archivo Level1.dat.
Tabla 6-9.



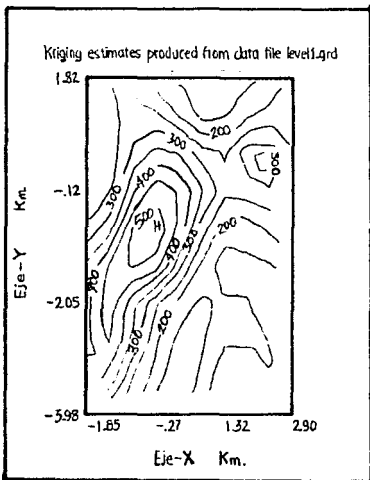
Gráfica de los errores de la diferencia del observado menos el estimado.
Figura 6-18.

Aquí se grafican las diferencias entre el estimado y el observado. Si el valor estimado es mayor que el observado se utiliza una cruz (x) y en caso contrario se usa una equis (+), y su tamaño corresponde al tamaño del error.

La demanda de compra de terrenos en el centro de la ciudad está vinculada con el hecho de que el centro es una zona altamente comercial.

Existen terrenos que adquieren un valor bajo respecto a terrenos vecinos aún estando en el centro. Como en el caso del terreno que está en un callejón su precio es relativamente bajo. Tal vez si se abriera este callejón, la gente empezaría a pasar y su precio se incrementaría al nivel de los precios de terrenos vecinos o de terrenos más transitados.

En la figura 6-19, se muestra la gráfica producida por la rutina Conrec, en ella se despliegan las curvas de nivel que muestran como se desplazan las anisotropías con su respectivo valor estimado. Este tipo de gráficas sirven para hacer predicciones.

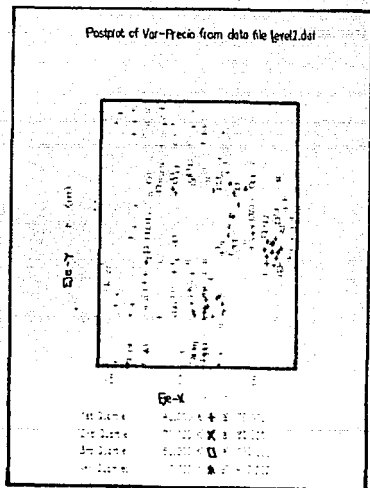


Resultados de las curvas de nivel producidas por la rutina Conrec.

Figura 6-19.

6.4.2 ANALISIS GEOESTADISTICO DEL ARCHIVO LEVEL2.DAT

El estudio consistió en considerar las 216 observaciones de este archivo. La gráfica producida por la rutina Postplot de la figura 6-20, muestra la distribución muestral de los datos. En la parte central se nota un hueco, el cual se formó al eliminar del área total, los datos de Level1.dat.



Gráfica Postplot del archivo Level2.dat
Figura 6-20.

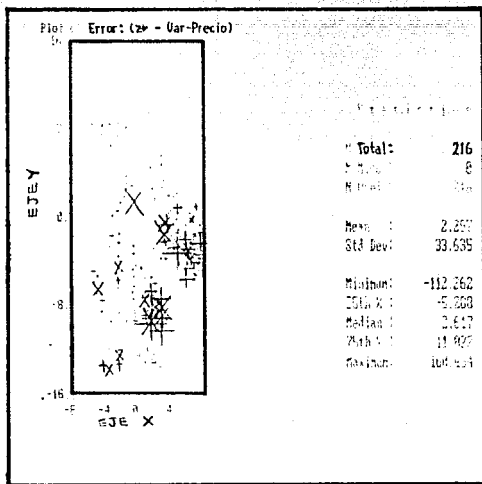
Se utilizó un modelo variograma esférico con un efecto pepita de 1000, una meseta de 5000 y un rango de 1.8. Después de resolver las ecuaciones Kriging con este modelo variograma, se utilizó la técnica de validación cruzada en donde se obtuvieron los siguientes resultados.

DATA FILE	:	D:\GEOM\LEVEL2.DAT			
VARIABLE	:	VAR-PRFICIO			
LOWER LIMIT	:	40.0000			
UPPER LIMIT	:	400.0000			
=====					
# OBSERVATIONS	:	216	SUM OF WEIGHTS	:	216.00000
# MISSING DATA	:	0			
# RETAINED	:	216			
MEAN	:	110.2689	MINIMUM VALUE	:	40.00000
VARIANCE	:	5277.7424	25TH PERCENTILE	:	70.00000
STD. DEVIATION	:	72.6481	MEDIAN	:	80.00000
%COEFF. VARIATION	:	65.8829	75TH PERCENTILE	:	110.00000
			MAXIMUM VALUE	:	400.00000
SKWENESS	:	2.2926			
KURTOSIS	:	18.0270			

Resultados Kvalid de los archivos Level2.dat.
Tabla G-10.

En la anterior tabla mostramos las estadísticas de este archivo de datos. El valor mínimo es 40, el valor máximo es 400 y la desviación estándar es 72.64 del total de 216 datos. Vemos que la desviación estándar es menos de la mitad de la desviación estándar del archivo Level1.dat.

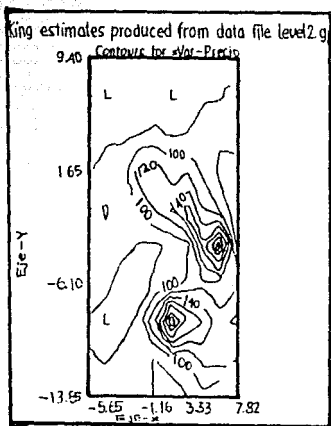
En la siguiente gráfica mostramos la variabilidad de los puntos estimados en comparación con los puntos observados, usamos la rutina Xvalid.



Gráfica de los errores de estimación utilizando la validación cruzada del archivo Level2.dat.

Figura 6-21.

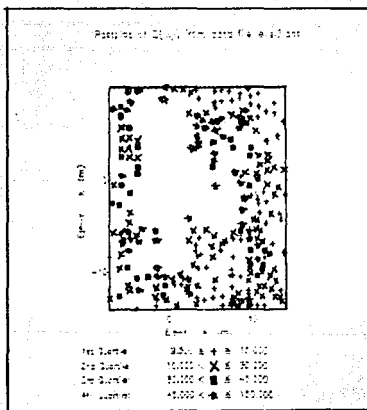
En la figura 6-22 Se muestra la gráfica producida por Conrec, en ella se despliegan las curvas de nivel las cuales muestran hacia dónde se desplazan las anisotropías.



Curvas de nivel del archivo Level2.dat.
Figure 6-22.

6.4.3 ANALISIS GEOESTADISTICO DEL ARCHIVO LEVEL3.DAT

En este archivo como en los dos anteriores se decidió tomar las muestras en un punto localizado en la esquina sur-este de la superficie de cada terreno. Las dimensiones de los terrenos en este archivo son las más grandes, ya que los tamaños fluctúan entre 10,000 y 100,000 metros cuadrados. Por supuesto que los terrenos que están ubicados en las afueras de la ciudad son más baratos. La gráfica de Postplot es la siguiente:



Postplot del archivo Level3.dat.
Figura 6-23.

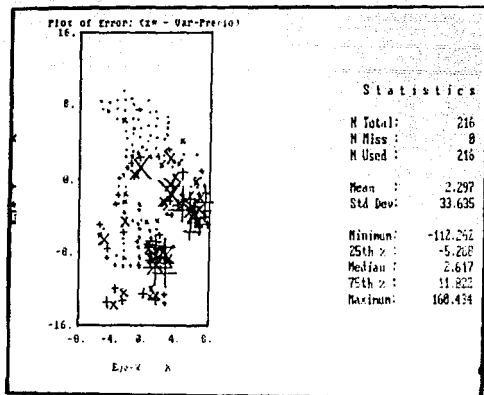
El modelo variograma que se utilizó fue un esférico con un efecto pepita de 80, una meseta de 350 y un rango de 7. Después de calcular las ecuaciones Kriging se procedió a usar la rutina Xvalid. Los resultados de validación cruzada son los siguientes:

Data File	: di: geo: level3.dat			
X Variable	: Eje-X	: K		
Y Variable	: Eje-Y	: K		
Kriging	: Z(X,Y)	# Data used	:	268
Type	: Ordinary	# Missing data	:	0
	Variable	Estimate	Difference	Kriging Std
Minimum	9.000	9.801	-71.955	10.830
25th Qtile	15.000	20.436	-5.539	12.896
Median	30.000	30.296	1.589	12.717
75th Qtile	45.000	41.977	6.547	13.512
Maximum	100.000	113.655	10.305	17.661
	268	268	268	268
Mean	33.274	33.781	.510	12.854
Std. Dev.	23.089	17.105	11.523	1.254

Resultados de Xvalid del archivo Level3.dat
Tabla 6-11.

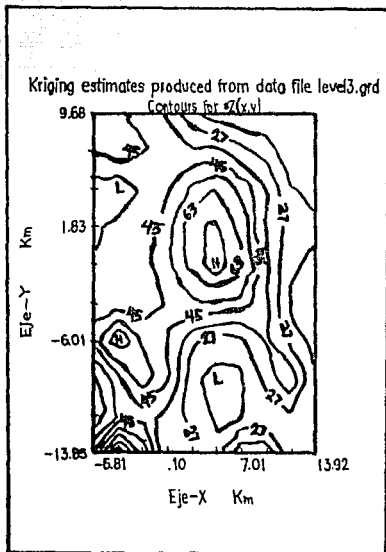
Es claro ver que la variabilidad en este caso es menor que en la variabilidad de los otros archivos, y que el valor esperado fluctúa alrededor del cero. Esto quiere decir que la varianza del error es mínima.

A continuación presentamos los resultados la gráfica de los errores producida por Xvalid:



Gráfica de los errores (Z - Z) producida por la rutina Xvalid.
Figura 6-23.

A continuación se presentan los resultados de las curvas de nivel de la rutina Conrec construídas a partir del variograma experimental.



Gráfica Conrec del archivo Level3.dat.
Figura 6-25.

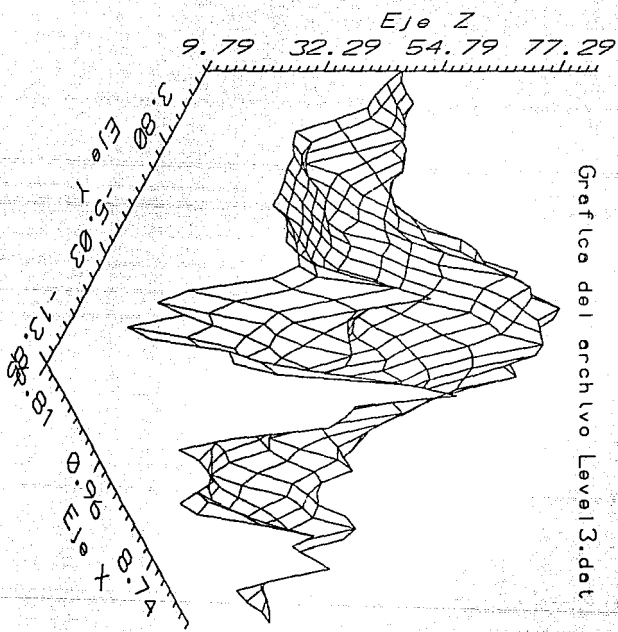


Grafico del archivo Level3.dat

Se nota que las dos áreas con alta variabilidad corresponden a las zonas etiquetadas con C (La Luna) y B (Tabachines).

Es fácil ver que las curvas de nivel de la gráfica con valores entre 27 y 36 se alargan en la dirección sur-este. En esta dirección pasa la autopista Cuernavaca-Cuautla, además en esta área se encuentra la zona industrial CIVAC.

La zona sur del centro marcada con una L en la figura 6-25, corresponden a terrenos de cultivo de caña de azúcar los cuales son terrenos ejidales. Si estos terrenos se regularizan, sus precios aumentarían considerablemente ya que están entre Palmira, Tabachines, San Gaspar, Sumiya y el boulevard Cuahnahuac. Las cuales son áreas residenciales por un lado y el boulevard forma parte de una comercial.

La varianza de la dispersión se redujo, debido a que la variabilidad del fenómeno en esta zona es menor. Esto se debe a que se redujo la media de la variabilidad y por lo tanto la varianza de la dispersión se redujo también. Los terrenos con menor valor por metro cuadrado son aquellos que están registrados a las afueras de la ciudad y además son en su gran mayoría de dimensiones grandes, podemos estandarizar un rango entre las 10 y 30 hectáreas.

CONCLUSIONES

La predicción espacial es un tema relativamente nuevo y poco se ha escrito en español para divulgar esta área de la Estadística.

El presente trabajo es una recopilación de varios artículos referentes a la predicción espacial y cuyo propósito es el de servir como libro de consulta así como el de proveer una ayuda para encontrar bibliografía.

El ejemplo desarrollado, nos da una idea de la utilidad de la teoría Geoestadística, ya que nos permite hacer predicciones sobre nuevas localidades y este hecho en la práctica se puede aplicar a diferentes áreas.

Fué muy importante ver cómo el método Kriging Ordinario se desarrolló históricamente, y cómo se ha ido extendiendo su aplicación a nuevas áreas.

En particular, creo que es importante destacar las diferentes etapas por las que pasó la teoría Geoestadística durante su desarrollo. Una de esas etapas ocurrió, cuando los estadísticos se dieron cuenta de que no estaban considerando las propiedades físicas de variabilidad de los depósitos minerales. Además, una vez que incluyeron la variabilidad en sus estudios, no la consideraron como un fenómeno local de ocurrencias aleatorias.

En resumen, si se conoce el variograma, entonces las ecuaciones kriging se pueden calcular interpolando toda la información disponible y la forma de hacerlo es asignando un peso a cada muestra dependiendo del valor de covarianza que la función variograma asigne.

El predictor Kriging es el mejor estimador lineal insesgado de una localidad conocida. Kriging trata de estimar valores de tal forma que la varianza del error asociado sea mínimo; pero cuando no es posible minimizar a la varianza. Entonces, el objetivo es detectar las áreas con la varianza de dispersión mayor, para saber en donde existen anisotropías. Por lo tanto estos indicadores ayudan a localizar áreas en donde las predicciones no son muy buenas.

Finalmente, se espera que este trabajo contribuya a dar a conocer el tema de predicción espacial sirviendo como un punto de apoyo para consulta y para estudios futuros.

BIBLIOGRAFIA

- A. W. Warrick y D.E. Myers, 1986, *Métodos Geostatistics Aplicados a la Ciencia de la Tierra*: American Society of Agronomy Soil Science Society of America; 677 South Segoe Road, Madison, WI 53711, USA Agronomy Monograph No. 9 (segunda edición).
- Amstrong (ed), 1989, *The loss of efficiency in Kriging Prediction*, Geostatistics, vol 1 pg 283-293.
- Amstrong and Delfiner (1980). *Towards a more Robust Variogram N-671*, Centre de Geostatistique, Fontainebleau.
- Amstrong, M. and Jabin, R., (1981) *Variogram models must be positive definite* Mathematical Geology, vol 13, No 5, 455-460
- Amstrong, M., (1984). *Commoner Problems seen in Variograms*,. Mathematical Geology, vol 16, No. 3, 305-313.
- Carroll, R.J. and rupper, D., (1982). *A comparison between maximum likelihood and Generalized least squares in a heterosedastic line model*: Journal of the American Statistic society, vol 77, pg. 878-882
- Cressie Noel (1990), *The Origins of Kriging*: Mathematical Geology, vol.22, No.3
- Cressie Noel (1989), *Geostatistics*: American Statistical Association, november, vol.43, no. 4
- Cressie Noel (1988), *Spatial Prediction and Ordinary Kriging*: Mathematical Geology, vol.20, no. 4
- Cressie (1986), *Kriging Nonstationary Data*: American Statistical Association, september, vol.81, No. 395, Applications
- Cressie Noel (1985), *Fitting Variogram Models by Weighted Least Squares*: Mathematical Geology, vol. 17, No. 5, p. 563-586.
- Cressie Noel (1985) *When are Relative Variograms useful in Geostatistics*: Mathematical Geology, vol.17, No.7
- Cressie, N. A. C. and Hawkins, D.M., 1980, *Robust Estimation of the Variogram*: Journal of the International Association of Mathematical Geology, vol. 12, p. 115-125.
- Cressie, Noel and Hawkins M. Douglas (1984). *Robust Kriging a Proposal*: Mathematical Geology, vol. 16, No. 1, pg. 3-17.
- Delfiner, P., (1976). *Linear estimation of non-stationary phenomena Advance Geostatistics for the mining industry*. Guaracio et al (eds), Reidel Publishing, Dordrecht, 49-68.

- Diamond, P. and Armstrong, M., 1983, Robustness of variograms and conditioning of kriging matrices: Internal note N-804, Centre de geoestatistique, Fontainebleau, France.
- Diamond, P. y Armstrong, M., (1984). robust of Variogram and Conditioning of Kriging matrices, *Mathematical Geology*, vol 16, no. 8, 809-822.
- Statistic Encyclopedia (1989) Volumen 4, pg 424, vol 6 pg 632, vol 8 pg 843. Variogram, vol 4 pg. 489; Kriging, vol 4 pg 411; Stochastic Processes, vol 8 pg 838, 848, 858; Stationary Processes, vol 6 pg 632.
- Harper, W.V., and Fur, J.M. (1986), "Geostatistical Analysis of Potentiometric data in a Wolfcamp Aquifer of the Palo Duro Basin Texas", technical report ONWI-587, Battelle Memorial Institute, Columbus, OH.
- Journel A. G., and Huijbregts, Ch., 1978, *Mining Geostatistics*: Academic Press, New York, 600 p.
- Myers Donald, 1989; *Vector Conditional Simulation*, Universidad de Tucson Arizona, Departamento de Matemáticas
- Myers Donald, E., (1987). *On variogram estimation*. Department of Mathematics, University of Arizona Tucson Arizona 85721, USA.
- Myers D. E., (1989). *To be or not to be ... Stationary ? That is the Question*. *Mathematical Geology*, vol 21, No. 3.
- Matheron (1963), *Principles of Geostatistics*, *Economic Geology*, vol. 58, pg 1246 - 1266. Vol I (1963a), vol II (1963b).
- Matheron (1965). *Les Variables regionalisees et leur estimation*. Masson, Paris. The mathematical statistics aspect of geoestatistics
- Matheron, G., (1971). *The theory of Regionalized Variables and its applications*. Centre of Geoestatistique, Fontainebleau, France, pg 212.
- Matheron (1962). *Tratado de la Geoestadística Aplicada*, Vol. 1, (1962), Vol 2 (1963). Technip, Paris.
- Mood Graybill (cuarta edición, 1978), *Introducción a la Teoría de la Estadística*. Colección de Ciencia y Tecnología
- Ord (1990), *Kriging*. *Enciclopedia Estadística*, recopilación de temas, pg 411.
- Rendu, J. M. (1978), *An introduction to Geoestatistical Methods of Mineral Evaluation*. The South African Institute of Mining and Metallurgy, in Johannesburg.

- Stein M. L., *The loss of efficiency in Kriging predicción cause by misspecificacions of the covarianza estructura.* Departamento de Matemáticas de la universidad de Chicago, 5734 S. University Ave., Chicago Illinois 60637.
- Warrich y Myers (1986), *Geostatistical methods applied to soil science*, Sociedad Americana de Agronomia, *Methods of soil analysis, part 1., Physical and mineralogical methods - agronomy monograph no. 9* (second edition).
- Yalom, A. M. (1957), *Some classes of Random Fields in n Dimensional Space, Related to Stationary Random Proceses: Theory of Probability and its Applicatios*, 2, pg 273-320.
- Zimmerman., D. L. and Zimmerman, M. B. (1989), "A monte Carlo Comparison of Spatial Variogram Esstimators, Kriging Predictors and Kriging Variances, manuscrito no publicado enviado a *Technometrics*.
- Procesos Estocásticos, Rodney Coleman (1987), editorial Limusa.
- Richard D. McArthur, 1987, *An evaluation of Sample Designs for Estimating a Locally Concentrated Pollutant.*, *Commun. Statist. Simula.*, 16(3), 735/759 (1987).
- Gandin, L. S., 1963, *Objective Analysis of meteorological fields: Gidrometeorologicheskoe Izdatel'stvo (GINIZ) Leningrad* (Translated b Israel program for Scientific Translations, Jerusalem, 1965. pg 238.
- Goldberg, A. S. 1962, *Best Linear Unbiased Prediction in the Generalized linear regression model.* *Int. Ame. Stat. Assoc.*, V 57 369-375.
- Edward H. Issaaks "Applied Geostatistics" Stanford University
R. Mohan Srivastava FSS International British Columbia New York
Oxford University Press 1989.