

01174
11

UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO

FACULTAD DE INGENIERÍA

DIVISIÓN DE ESTUDIOS DE POSGRADO

SIMULACIÓN SECUENCIAL GAUSSIANA DE LA
POROSIDAD Y SU APLICACIÓN EN LA
CARACTERIZACIÓN GEOMECÁNICA DE
YACIMIENTOS

T E S I S

QUE PARA OBTENER EL GRADO DE:
MAESTRO EN INGENIERÍA

P R E S E N T A

OCTAVIO JESÚS SAAVEDRA ESTRADA

DIRECTORES DE TESIS: DR. DANIEL GARCÍA GAVITO

DR. VINICIO SURO PÉREZ



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 esta 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.

MI GRATITUD A LA DIVISION DE ESTUDIOS DE POSGRADO DE LA FACULTAD DE INGENIERIA Y A CADA UNO DE LOS PROFESORES QUE ME PERMITIERON CONTINUAR CON ESTA ETAPA DE MI PREPARACION.

A VIRIDIANA

INDICE

LISTA DE FIGURAS

RESUMEN

INTRODUCCIÓN

1. ANTECEDENTES

1.1. ANÁLISIS ESTADÍSTICO

1.1.1. ESTADÍSTICA UNIVARIADA

1.1.2. ESTADÍSTICA BIVARIADA

1.1.3. DESAGRUPAMIENTO

1.2. MODELOS PROBABILISTICOS

1.2.1. VARIABLES ALEATORIAS

1.2.2. ESTIMACIÓN DE LA DISTRIBUCIÓN DE PROBABILIDAD

1.2.3. MOMENTOS DE UNA VARIABLE ALEATORIA

1.2.4. MOMENTOS DE UNA DISTRIBUCIÓN BIVARIADA

1.2.5. VARIABLES REGIONALIZADAS

1.2.6. FUNCIONES ALEATORIAS

1.2.7. APLICACIÓN DE LAS FUNCIONES ALEATORIAS EN LA DETERMINACIÓN DE LA CONDICIÓN DE INSEGAMIENTO.

2. GEOESTADÍSTICA

2.1. FUNDAMENTOS PARA EL MODELADO DEL VARIOGRAMA

2.1.1. MODELOS TEÓRICOS PARA VARIOGRAMAS

2.1.2. ANISOTROPÍA

2.1.3. VARIOGRAMA Y COVARIANZA

2.2. KRIGING

2.2.1. KRIGING SIMPLE

2.2.2. KRIGING ORDINARIO

2.3. SIMULACIÓN ESTOCÁSTICA

2.3.1. PRINCIPIOS DE LA SIMULACIÓN ESTOCÁSTICA

2.3.2. SIMULACIÓN SECUENCIAL

2.3.3. SIMULACIÓN SECUENCIAL GAUSSIANA

2.3.4. TRANSFORMACIÓN NORMAL ESTÁNDAR

2.3.5. VERIFICACIÓN DE LA NORMALIDAD BIVARIADA

2.3.6. REESCALAMIENTO

2.4. SISTEMA DE COORDENADAS ESTRATIGRAFICAS

3. GEOMECÁNICA

3.1. PROPIEDADES ELASTICAS Y DE FALLA DE LA ROCA

3.2. CORRELACIONES

3.3. COEFICIENTE DE COMPACTACIÓN UNIAxIAL

4. METODOLOGÍA PARA LA CARACTERIZACIÓN GEOMECÁNICA

5. EJEMPLO DE APLICACIÓN Y ANÁLISIS DE RESULTADOS

CONCLUSIONES Y RECOMENDACIONES

BIBLIOGRAFÍA

NOMENCLATURA

APÉNDICE

LISTA DE FIGURAS

- Figura 1.1. Gráfico de dispersión de dos conjuntos de datos, U vs. V .
- Figura 1.2. a) Función de distribución de probabilidad y b) Función de distribución acumulativa del modelo gaussiano o normal.
- Figura 1.3. Función de distribución acumulativa definida por la ecuación (1.11) para una variable aleatoria V .
- Figura 1.4. Gráfico de un par (x_i, y_i) sobre un diagrama de dispersión
- Figura 2.1. Variograma que expresa la continuidad espacial de una variable regionalizada. Es la gráfica del promedio de la diferencia cuadrada entre todos los pares de datos de la variable de interés a una distancia dada de separación.
- Figura 2.2. Cálculo del variograma. (A). Cálculo de la distancia de separación $(Z_i - Z_j)^2$, donde Z es la variable de interés. (B) Gráfico de la distancia de separación vs. $(Z_i - Z_j)^2$. (C) Promedio de todos los datos en intervalos. (D) Ajuste de un modelo al variograma.
- Figura 2.3. Ilustración de las tolerancias sobre h para la selección de pares de datos en el cálculo del variograma.
- Figura 2.4. Elementos del variograma: Efecto pepita, rango y umbral.
- Figura 2.5. Modelos teóricos para variogramas con el mismo rango práctico (gráfica superior), y modelos de potencias para diferentes valores del parámetro ω , gráfica inferior.
- Figura 2.6. Anisotropía geométrica.
- Figura 2.7. Ejemplo de anisotropía zonal en la dirección del acimut ϕ (varianza mayor).
- Figura 2.8. Ejemplos idealizados del gráfico del variograma y la covarianza
- Figura 2.9. Kriging encontrando el valor del punto a ser estimado de manera tal que éste se ajuste al modelo del variograma.
- Figura 2.10. Simulación secuencial de porosidades en una capa.
- Figura 2.11. Procedimiento gráfico para la transformación de los valores de una distribución no-gaussiana a una distribución normal estándar.
- Figura 3.1. Rangos del coeficiente de compactación uniaxial para carbonatos.
- Figura 5.1. Distribución de los 19 pozos dato en el área de estudio.
- Figura 5.2. Perfiles de los pozos dato mostrando la longitud cubierta por la información disponible.
- Figura 5.3. Descripción estadística de los valores de porosidad disponibles para el estudio.
- Figura 5.4. Descripción estadística de los valores desagrupados de porosidad.
- Figura 5.5. Gráficas de las funciones de distribución acumulada tanto para los datos originales de porosidad como los desagrupados.
- Figura 5.6. Distribución normal estándar (normal score).
- Figura 5.7. Variograma modelado en la dirección NS.

TESIS CON
FALLA DE ORIGEN

Figura 5.8. Variograma modelado en la dirección EW.

Figura 5.9. Variograma modelado en la dirección NE-SW 45°.

Figura 5.10. Variograma modelado en la dirección NW-SE 45°.

Figura 5.11. Elipse de anisotropía mostrando los rangos máximos y mínimos.

Figura 5.12. Variograma en la dirección vertical.

Figura 5.13. Variograma omnidireccional.

Figura 5.14. Imagen de una simulación de la porosidad (mallado fino).

Figura 5.15. Imagen de los resultados de un reescalamiento de la porosidad.

Figura 5.16. Imágenes de la distribución de la porosidad a la profundidad de 2500 m.

Figura 5.17. Determinación del módulo M_s a partir del registro de densidad y del tiempo de tránsito.

Figura 5.18. Imágenes de C_o a la profundidad de 2500 m.

Figura 5.19. Imágenes de K a la profundidad de 2,600 m.

Figura 5.20. Imágenes de G a la profundidad de 2700 m.

Figura 5.21. Imágenes de E a la profundidad de 2800 m.

Figura 5.22. Imágenes de v a la profundidad de 2900 m.

Figura 5.23. Imágenes de C_m a la profundidad de 2500 m.

Figura 5.24. Imagen de la compactación del yacimiento bajo estudio.

RESUMEN

Este trabajo presenta una metodología para la estimación de la resistencia a la compresión uniaxial así como de las propiedades elasto-dinámicas: módulo volumétrico, módulo de corte, módulo de Young y relación de Poisson de las rocas; todas ellas como función de la porosidad empleando la correlación de Gutiérrez-García, basada en la teoría de autoconsistencia, y utilizando simulación secuencial gaussiana para el modelado de su distribución en el yacimiento.

Además, se ejemplifica la metodología propuesta con la determinación de la distribución del coeficiente de compactación uniaxial como función del módulo de Young y la relación de Poisson, y su aplicación en la determinación de la compactación en un yacimiento real.

INTRODUCCIÓN

El conocimiento del comportamiento mecánico de las rocas en los campos petroleros, es fundamental para el diseño e implementación de los programas de perforación y la caracterización de los yacimientos. Tradicionalmente, las propiedades mecánicas de las rocas son obtenidas a partir de mediciones directas en muestras de núcleos o de cálculos sobre las mediciones de los registros acústicos. La obtención de las propiedades geomecánicas mediante estos procedimientos es costoso, además de que sólo un número limitado de muestras es analizado.

Lo anterior ha conducido al desarrollo de nuevos métodos de predicción de las propiedades mecánicas de las rocas a partir de información más accesible; ejemplo de ello son las correlaciones con la porosidad. ⁽¹⁾

No obstante la valiosa información que se puede obtener en cada uno de los pozos, las variaciones laterales de la porosidad al igual que las otras propiedades de las rocas, no pueden ser delineadas a partir éstos debido a la considerable distancia que los separa y a la complejidad de las estructuras geológicas. La integración de los métodos geoestadísticos y el modelado estocástico son una alternativa para mejorar significativamente la descripción espacial de tales atributos. ⁽²⁾

Se puede considerar a un sistema como *determinístico* si éste produce la misma respuesta cuando se simula varias veces con la misma entrada. Entre los procedimientos determinísticos para la estimación de los atributos de interés se encuentran la regresión y Kriging. Por el contrario, se dice que un sistema es *aleatorio* cuando genera diferentes resultados (no relacionados entre sí) a partir de la misma entrada. Un sistema *estocástico* es aquel que es parte determinístico y parte aleatorio, es decir, un sistema híbrido. Estocástico implica que existe algo de impredecibilidad en la descripción de un conjunto de números o en una estadística. Entre los métodos estocásticos se pueden mencionar: Simulación Secuencial, Simulación Booleana, Descomposición Matricial, por nombrar solo algunos.

En los yacimientos se pueden presentar éstas tres categorías. Puede ser determinístico con elementos bien reconocibles y correlacionables a una escala de kilómetros entre pozos. Se puede considerar estocástico debido a una variabilidad aleatoria local y a su parte determinística y a mencionada. Y puede ser puramente aleatorio al no existir una

distribución de las propiedades geológicas fácilmente identificable en una unidad heterogénea.⁽³⁾

En general las predicciones determinísticas son "superiores" a las estocásticas. Después de todo, las primeras no tienen incertidumbre y usualmente contienen información sobre la naturaleza del sistema. Consecuentemente, si el sistema bajo consideración es perfectamente comprendido (si la naturaleza de la distribución entre pozos de la propiedad estudiada es perfectamente conocida), entonces los métodos determinísticos deben ser los métodos de predicción preferidos. Pero en la realidad, debemos reconocer que raramente tenemos un sistema perfectamente determinístico, las propiedades parecen ser aleatorias aún cuando su naturaleza física sea comprendida.⁽³⁾

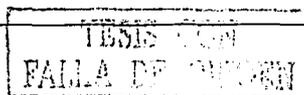
Los atributos de interés geomecánico al igual que los geológicos muestran patrones diferentes que son el resultado de procesos sedimentarios o estructurales. Una diferencia fundamental entre la estadística y la geoestadística es la extracción de información relacionada a la *localización* de cada observación asociada con tales procesos. Los atributos que tienen una localización específica asociada con cada observación, es decir, que tienen una distribución a través del espacio, tal como la porosidad o la permeabilidad por ejemplo, reciben el nombre de *variables regionalizadas*. La geoestadística se enfoca al estudio de estas variables.

No obstante que existen en la literatura diversas correlaciones para la determinación de las variables de interés geomecánico a partir de la porosidad, éstas han sido determinadas a partir de información proveniente de formaciones geológicas con características muy diferentes a las existentes en los yacimientos mexicanos. Por esta razón se seleccionó la correlación propuesta por Gutiérrez-García⁽⁴⁾ la cual fue determinada para las condiciones específicas de nuestros yacimientos.

El principal objetivo que se persigue con el desarrollo de este trabajo de tesis consiste en presentar una metodología para la caracterización geomecánica de los yacimientos petroleros, a partir de las imágenes estocásticas de la porosidad obtenidas mediante su simulación secuencial gaussiana.

En forma general, la metodología para llevar a cabo este trabajo se puede dividir en dos etapas:

- 1) Determinar la distribución espacial de la porosidad en el yacimiento empleando simulación secuencial gaussiana a partir de los valores obtenidos en cada pozo.



- 2) Mediante el empleo de correlaciones, transformar las imágenes de porosidad obtenidas en cada una de las variables que permiten determinar el comportamiento mecánico de las rocas del yacimiento.

TESIS CON
FALLA DE ORIGEN

1.- ANTECEDENTES

1.1.- ANÁLISIS ESTADÍSTICO

En términos generales se puede decir que la Probabilidad y Estadística es el estudio de los fenómenos aleatorios. Cuando se lleva a cabo la observación de un fenómeno y no es posible determinar con certeza cuál será el resultado antes de su medición, éste recibe el nombre de *experimento aleatorio*. El conjunto de todos los posibles resultados de un experimento aleatorio se denomina *espacio muestral*.

La Estadística trata dos grupos de datos diferentes: la colección de toda la posible información que caracteriza a un fenómeno de interés, llamada la *población*, y subconjuntos representativos de la población, denominados *muestras*.

El análisis estadístico permite describir a estas muestras a través de una serie de medidas numéricas que resumen su comportamiento.

1.1.1- ESTADÍSTICA UNIVARIADA ^{(5), (6)}

Una de las presentaciones de conjuntos de datos más comunes y útiles es la *tabla de frecuencias* y su correspondiente gráfica, el *histograma*. Una tabla de frecuencias registra la regularidad con la que los valores observados se ubican dentro de ciertos intervalos o *clases*.

Adicionalmente a las técnicas gráficas para describir los patrones de distribución ocultos en un conjunto de datos, existen medidas numéricas que se emplean para describir a éstos conjuntos, por ejemplo, las características importantes de los histogramas pueden ser capturadas en el *resumen estadístico*.

El resumen estadístico se puede dividir en 3 categorías:

- a) Medidas de tendencia central.
- b) Medidas de dispersión.
- c) Medidas de asimetría.

La estadística en el primer grupo busca un valor que pueda representar a toda la muestra por encontrarse en el centro de ella. Las más importantes medidas de tendencia central

son: la media, la mediana y la moda; éstas nos dan una idea de dónde se encuentra el centro de la distribución.

La *media* es el promedio aritmético de un conjunto de observaciones, sin embargo dado que cualquier observación en el conjunto se emplea para su cálculo, el valor de la media puede afectarse en forma desproporcionada por la existencia de algunos valores extremos.

La *mediana* es el valor que corresponde a la mitad de los datos ordenados de una muestra. La mediana puede ser una medida de tendencia central más representativa que la media cuando existan algunos valores extremos ya que éstos no afectan su valor.

La *moda* de un conjunto de observaciones es el valor de las observaciones que ocurre con mayor frecuencia. La moda muestra hacia qué valor tienden los datos a agruparse.

Una medida de tendencia central proporciona información acerca de un conjunto de datos pero no proporciona ninguna idea de la variabilidad de las observaciones en dicho conjunto. Las medidas de dispersión reflejan la separación de los elementos de una muestra.

La *varianza*, s^2 , de un conjunto de observaciones es en esencia, el promedio del cuadrado de las distancias entre cada observación y la media del conjunto de observaciones.

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.1)$$

donde:

n : es el número total de datos de la muestra.

x_i : es el valor del i ésimo dato.

\bar{x} : es la media del conjunto de datos.

La varianza es una medida razonablemente correcta de la variabilidad debido a que si muchas de las diferencias son grandes (o pequeñas), entonces el valor de la varianza será grande (o pequeño). Sin embargo, el valor de la varianza puede sufrir un cambio muy considerable, aún más que la media, por la existencia de algunos valores extremos en el conjunto de datos.

A la raíz cuadrada de la varianza se le denomina *desviación estándar*, s .

El *coeficiente de variación* es una medida estandarizada de la variación con respecto a la media, especialmente útil para comparar dos distribuciones de frecuencias cuando la escala de medición difiere de manera apreciable entre éstas:

$$\text{Coeficiente de variación} = \frac{s}{\bar{x}} \quad (1.2)$$

Una gráfica útil de un conjunto de datos adicional al histograma es la *distribución de frecuencias acumulativas*, la cual se obtiene graficando, en el eje vertical, la frecuencia relativa acumulativa de una clase contra el límite inferior de la siguiente sobre el eje horizontal y uniendo con segmentos todos los puntos consecutivos.

Con respecto a una distribución de frecuencias acumulativas se define un *cuartil-p* como el valor bajo el cual se encuentra una determinada proporción de los valores de la distribución. Por ejemplo el cuartil $q_{0.25}$ (percentil 25avo), es el valor bajo el cual se encuentra el 25% de los valores de la distribución.

Una característica del histograma que las medidas estadísticas anteriores no muestran es su asimetría. El coeficiente de asimetría permite determinarla a través del signo resultante al ser evaluado:

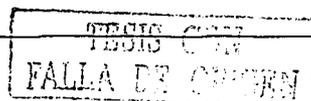
$$\text{Coeficiente de asimetría} = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^3}{s^3} \quad (1.3)$$

Un histograma con asimetría negativa tendrá un extremo con los valores más altos cargados hacia la derecha. Si el valor del coeficiente de asimetría es cercano a cero, el histograma será aproximadamente simétrico y la mediana tendrá un valor muy cercano a la media. Cuando se desea ajustar un modelo probabilístico a un fenómeno particular es conveniente comparar la forma del histograma de la muestra de datos con la de la función de probabilidad del modelo teórico.

Tomadas en conjunto las estadísticas permiten obtener un resumen de la información contenida en el histograma.

1.1. 2.- ESTADÍSTICA BIVARIADA ⁽⁵⁾

Las herramientas estadísticas vistas anteriormente pueden ser usadas para describir las muestras de las variables analizadas en forma independiente, sin embargo una de las



características más importantes de la información procedente de las ciencias de la tierra es la relación y dependencia entre sus diferentes atributos, por ello es necesario revisar algunos de los conceptos necesarios para describir la relación entre dos variables.

GRAFICO DE DISPERSIÓN

El desplegado más común de datos bivariados es el *grafico de dispersión*, éste consiste en graficar los valores de una de las variables en el eje de las abscisas vs. los valores de la otra variable en el eje de las ordenadas procedentes del mismo punto de muestreo (Figura 1.1). No obstante que puede existir una dispersión en la nube de puntos, los valores grandes de los dos conjuntos tienden a estar asociados entre sí, de igual forma que los valores pequeños.

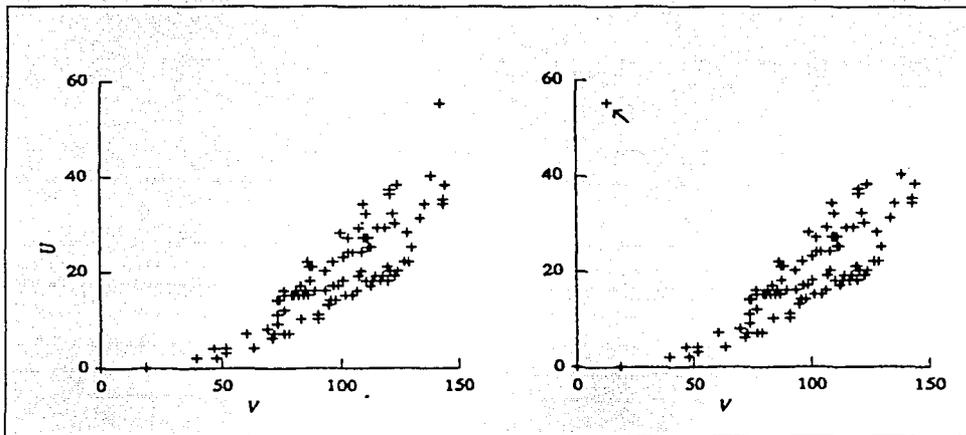


Figura 1.1. Gráfico de dispersión de dos conjuntos de datos, U vs. V . El valor indicado por la flecha en la gráfica derecha representa un punto que ha sido "accidentalmente" mal dimensionado.⁽⁵⁾

Adicionalmente a que este gráfico permite tener una idea cualitativa de la forma en que dos variables están relacionadas, un gráfico de dispersión también es útil para distinguir los datos fuera del comportamiento normal. Figura 1.1. Esto es importante en las primeras etapas del estudio donde los datos son revisados y depurados.

En forma muy general se puede decir que existen tres comportamientos que se pueden observar en un gráfico de dispersión: las variables pueden estar positivamente correlacionadas, negativamente correlacionadas o no-correlacionadas.

Dos atributos están positivamente correlacionados si los valores grandes de uno de ellos tienden a estar asociados con los valores grandes del otro, y similarmente con los valores pequeños. Para el caso de las rocas por ejemplo, la porosidad y la permeabilidad normalmente están correlacionadas positivamente.

Dos variables están negativamente correlacionadas si los valores grandes de una de las variables tienden a estar asociados con los valores pequeños de la otra variable. La tercera posibilidad es que las dos variables no estén relacionadas, es decir, un incremento en una de las variables no tiene efecto aparente sobre la otra.

CORRELACIÓN

Una medida de la variabilidad conjunta de dos variables es la *covarianza*. Esta es una medida de asociación entre sus correspondientes valores, así como de sus respectivas dispersiones. Cuando se tiene una alta probabilidad de que los valores grandes de una de las variables se encuentren asociados con los valores pequeños de la otra, la covarianza será negativa. Si la relación entre las variables es directamente proporcional, la covarianza será positiva. Por otra parte si las variables son estadísticamente independientes, la covarianza será cero.

La covarianza puede ser calculada para dos variables X e Y mediante la siguiente ecuación:

$$C(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) \quad (1.4)$$

donde \bar{x} e \bar{y} , son las medias de los valores de las variables X e Y respectivamente.

Si la covarianza de X e Y se divide por el producto de sus desviaciones estándar, el resultado es una cantidad sin dimensiones que recibe el nombre de *coeficiente de correlación*, ρ , y puede ser calculado con la siguiente expresión:

$$\rho = \frac{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{s_x \cdot s_y} \quad (1.5)$$

El coeficiente de correlación es una medida normalizada de la asociación lineal que existe entre dos variables en relación con sus dispersiones. Un valor de $\rho = 0$ indica la ausencia de cualquier asociación lineal, sin embargo no se puede asegurar que dos variables que tengan un coeficiente de correlación muy pequeño sean independientes ya que pueden tener otro tipo de relación que no sea lineal. Los valores -1 y 1 indican relaciones lineales perfectas negativa y positiva respectivamente; esto significa que se puede expresar una de ellas como una función lineal de la otra.

La covarianza entre dos variables depende fuertemente de la magnitud de los datos, por ejemplo, si se multiplican los valores de cada variable por 10, la covarianza será 100 veces más grande. Al dividir la covarianza por las desviaciones estándar de las dos variables se garantiza que el coeficiente de correlación siempre tome valores en el intervalo $-1 \leq \rho \leq 1$, por lo tanto es un índice independiente de la magnitud de los datos.

Tanto el coeficiente de correlación como la covarianza pueden verse afectados por pares de datos aberrantes. Una buena alineación de unos cuantos pares extremos puede mejorar considerablemente al coeficiente de correlación; y por el contrario, una buena correlación puede verse arruinada por un inadecuado alineamiento de los pares extremos.

1.1. 3.- DESAGRUPAMIENTO ^{(5), (7)}

Cuando se calcula la estadística de una muestra puede resultar que el valor de su media por ejemplo, sea una estimación "muy pobre" de la media de la población. Esto es debido a una mala estrategia de muestreo, o bien a que la información disponible se concentra en ciertas zonas de interés. Por lo tanto, para obtener un histograma que sea representativo de la zona de interés es necesario ponderar cada una de las muestras individuales de tal manera que las muestras agrupadas no tengan influencia en el conjunto total.

Dos de los métodos más comunes para desagrupar, aplicables a cualquier conjunto de datos son: el desagrupamiento poligonal y el método de desagrupamiento por celdas. El primer método asigna un polígono de influencia a cada muestra, de esta manera, las áreas de estos polígonos son utilizadas para calcular los pesos de ponderación.

En el desagrupamiento por celdas, el área de estudio es dividida en regiones rectangulares denominadas "celdas". Cada celda recibe un peso inversamente proporcional al número de muestras que caen dentro de la misma celda. Todas las

muestras dentro de una celda particular reciben igual peso y el peso total es de uno. La estimación que se obtiene con este método depende del tamaño que se selecciona para las celdas. Si las celdas son muy pequeñas, cada muestra caerá en una celda por lo que todas las muestras recibirán un peso de uno. Si las celdas son muy grandes entonces una gran cantidad de muestras pertenecerán a esa celda y todas recibirán igual peso, incluyendo las que no están agrupadas.

En el caso de tres dimensiones las celdas se convierten en bloques rectangulares cuya altura, anchura y profundidad deberán seleccionarse.

1.2 MODELOS PROBABILÍSTICOS

Los conceptos vistos hasta ahora han sido puramente descriptivos. Estos se han enfocado a la descripción y depuración del conjunto de datos de una muestra, sin embargo, si se desea realizar una estimación de parámetros en áreas no muestreadas utilizando únicamente la información disponible, es necesario recurrir a modelos probabilísticos que representen el comportamiento del atributo en las localidades de interés. Por esta razón se presentan a continuación conceptos indispensables para la conceptualización de tales modelos.

1.2.1. VARIABLES ALEATORIAS ^{(3), (6), (9), (7)}

Una variable aleatoria proporciona un medio para relacionar cualquier resultado del estudio de un fenómeno o experimento con una medida cuantitativa. La variable aleatoria se puede ver como una variable, por ejemplo Z , que puede tomar una serie de resultados o realizaciones, z_i , $i = 1, 2, \dots, n$, con un conjunto dado de probabilidades de ocurrencia (p_i , $i = 1, 2, \dots, n$)

Se dice que es "aleatoria" porque involucra las características probabilísticas de los resultados asociados en el espacio muestral. Esto debe llevar a la idea de *incertidumbre*, no acerca de la regla de asociación sino del resultado del evento. A pesar de que una variable aleatoria puede tomar cualquier valor dentro de un rango dado, cada valor está asociado con su frecuencia de ocurrencia a través de una función de distribución

Una variable aleatoria depende de la información en el sentido de que su distribución de probabilidad cambia a medida que están disponibles mas datos de los valores no muestreados.

Por ejemplo, si consideramos el evento de medir las porosidades en tres muestras de rocas llamando a los valores medidos ϕ_1 , ϕ_2 y ϕ_3 , el espacio muestral serán todos los valores dentro del rango de 0 a 1. Cada resultado ϕ , tendrá un valor numérico correspondiente a la porosidad de la muestra de roca. Estas porosidades pertenecen a un tipo de variables denominadas "variables aleatorias continuas", sin embargo existen otro tipo de variables denominadas "variables aleatorias discretas", de acuerdo con las siguientes definiciones:

Una variable aleatoria es *discreta* si el número de valores que puede tomar es entero.

Una variable aleatoria es *continua* si sus valores consisten en uno o más intervalos del conjunto de los números reales, por lo que se trata de conjuntos infinitos.

El comportamiento de una variable aleatoria discreta Z puede estudiarse a través de su *función de probabilidad*, la cual relaciona los resultados de un experimento aleatorio con su probabilidad de ocurrencia. Esta función se denota como $P(Z = z)$, y se interpreta como la probabilidad de que la variable aleatoria Z tome el valor particular z .

La distribución de probabilidad de Z no solo implica la existencia de la función de probabilidad sino también la existencia de su *función de distribución acumulativa*, la cual representa la probabilidad de que Z sea menor o igual a un cierto valor específico x , esta función está dada por:

$$F(z) \equiv P(Z \leq z) = \sum_{z_i \leq z} P(Z = z_i) \quad (1.6)$$

En el estudio del comportamiento de una variable aleatoria continua no es posible relacionar cada valor de la variable Z con su probabilidad de ocurrencia, tal como sucede para el caso discreto. Una variable aleatoria continua puede tomar una infinidad de valores en un pequeño intervalo, en consecuencia, la probabilidad de que Z tome un valor particular es prácticamente cero. Sin embargo esto no implica que sea imposible, ya que el evento ($Z = z$) puede llegar a ocurrir.

La distribución de probabilidad de una variable aleatoria continua Z está caracterizada por una función $f(z)$, la cual recibe el nombre de *función de densidad de probabilidad*. Esta función no representa la probabilidad de que $Z = z$, mas bien, ésta representa un medio para determinar la probabilidad de que la variable Z se encuentre en el intervalo

$a \leq Z \leq b$. La densidad de probabilidad es un concepto auxiliar que indica cuales son los subintervalos con mayor o menor probabilidad:

$$\text{Prob}(a \leq Z \leq b) = \int_a^b f(z) dz \quad \forall a, b \quad (1.7)$$

La *función de distribución acumulativa* de una variable aleatoria continua Z , denotada por $F(z)$, es la probabilidad de que la variable Z tome un valor menor o igual a algún z específico, es decir

$$F(z) = \text{Prob}(Z \leq z) \quad (1.8)$$

FUNCIONES DE DISTRIBUCIÓN DEL MODELO GAUSSIANO (NORMAL)

Este es un modelo de distribución totalmente caracterizado por sus dos parámetros, la media y la varianza. Figura 1.2

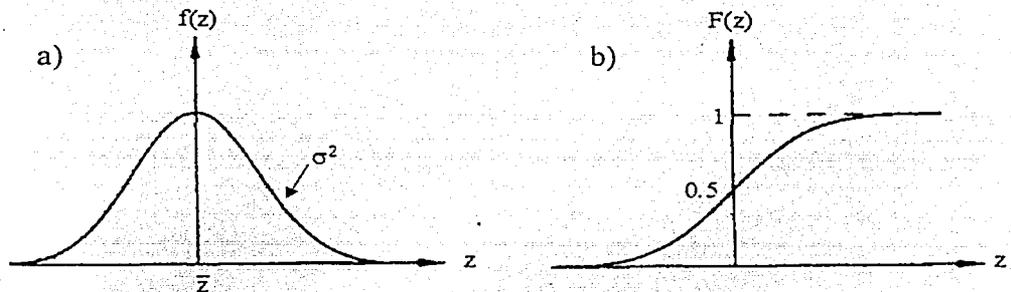


Figura 1.2. a) Función de distribución de probabilidad y b) Función de distribución acumulativa del modelo gaussiano o normal ⁽⁹⁾

Las funciones de distribución de las variables aleatorias permiten hacer un análisis probabilístico de los datos, sin embargo para ser desarrolladas directamente a partir de las observaciones del atributo estudiado, una opción es ajustar una curva de distribución teórica a su polígono de frecuencias.

TESIS CON
FALLA DE ORIGEN

1. 2. 2 ESTIMACIÓN DE UNA DISTRIBUCIÓN DE PROBABILIDAD ⁽⁵⁾

La estimación de una distribución se puede llevar a cabo determinando la proporción de valores que caen dentro determinadas clases particulares, o bien, estimando la proporción de valores que caen arriba o debajo de ciertos umbrales. En la primera aproximación se estima la distribución de la frecuencia; en la segunda, estamos estimando la distribución de la frecuencia acumulativa.

Si denotamos la función de distribución acumulativa, $F(z_c)$, como la proporción de valores por debajo de z_c , entonces la frecuencia acumulativa bajo el valor mínimo será cero, y la frecuencia acumulativa bajo el valor máximo será uno, es decir

$$F(z_{\min}) = 0, \quad F(z_{\max}) = 1$$

Existen dos aproximaciones generales para la estimación de la distribución acumulativa. La primera se conoce como aproximación no-paramétrica, consiste en calcular estimados de $F(z)$ en varias localidades de z : $F(z_1), \dots, F(z_n)$. La segunda, referida como aproximación paramétrica, consiste en determinar una función que describa completamente la distribución acumulativa para cualquier valor de z . En este trabajo nos basaremos en la aproximación no-paramétrica.

Para calcular la proporción de valores que se encuentran por debajo de un umbral z_c podemos transformar los valores z_1, \dots, z_n , a un conjunto correspondiente de variables indicadoras $i_1(z_c), \dots, i_n(z_c)$, de acuerdo a:

$$i_j(z_c) = \begin{cases} 1 & \text{si } z_j \leq z_c \\ 0 & \text{si } z_j > z_c \end{cases} \quad (1.9)$$

La proporción acumulativa de valores bajo cualquier umbral puede ser expresada como:

$$F(z_c) = \frac{1}{n} \sum_{j=1}^n i_j(z_c) \quad (1.10)$$

Para cualquier umbral z_c , podemos transformar los valores continuos de una variable Z en un indicador $I(z_c)$ usando la ecuación 1.10. En la práctica no se tiene acceso a todos los valores de la población sino solamente a un conjunto de muestras. No obstante, podemos aplicar la transformación indicadora a nuestras muestras disponibles y estimar la proporción real tomando un promedio ponderado de estas muestras indicadoras:

$$F(z_c) = \sum_{j=1}^n w_j \cdot i_j(z_c) \quad (1.11)$$

donde los n pesos w , son estandarizados para que sumen uno.

La Figura 1.3 muestra como se comporta la función de distribución acumulativa. Se puede observar que la altura de los saltos no es la misma en cada uno de los valores de la muestra. En cada valor de éstos, la función de distribución acumulativa se incrementa por w_j , el peso asignado a esa muestra particular.

Si las muestras disponibles cubren el área de interés totalmente sin que exista agrupamiento de las muestras en algunas zonas, entonces las muestras indicadoras pueden ser ponderadas con el mismo peso.

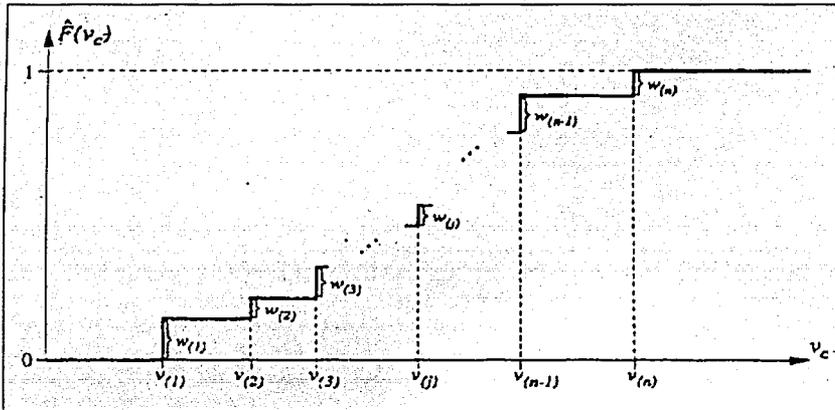


Figura 1.3. Función de distribución acumulativa definida por la ecuación (1.11) para una variable aleatoria V . Inicia en cero y se eleva hasta uno en una serie de saltos determinados por su peso correspondiente. ⁽⁵⁾

1. 2. 3 MOMENTOS DE UNA VARIABLE ALEATORIA ^{(9), (6)}

El uso de los momentos para caracterizar a la distribución de probabilidad de una variable aleatoria es muy útil, especialmente cuando es poco probable que se conozca la distribución de probabilidad.

TESIS CON
FALLA DE ORIGEN

MOMENTOS DE PRIMER ORDEN

El primer momento con respecto al origen es la *media* o *valor esperado* de una variable aleatoria. El valor esperado de la variable aleatoria Z es la suma ponderada de todas sus posibles ocurrencias, en donde la probabilidad de que Z tome el valor z_i es precisamente la ponderación. El valor esperado está dado por:

$$E[Z] = m = \sum_{i=1}^N p_i z_i \quad \text{si } Z \text{ es discreta} \quad (1.12)$$

$$E[Z] = m = \int_{-\infty}^{\infty} z f(z) dz \quad \text{si } Z \text{ es continua} \quad (1.13)$$

en donde $f(z)$ es la función densidad de probabilidad.

La media de una variable aleatoria se considera como una cantidad numérica alrededor de la cual los valores de la variable aleatoria tienden a agruparse. Por lo tanto, conociendo la media se tiene una idea del *valor central* de la distribución.

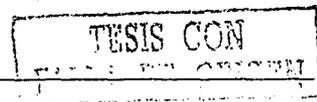
MOMENTOS DE SEGUNDO ORDEN

El segundo momento con respecto a la media recibe el nombre de *varianza*. Es una medida de la dispersión de la distribución de probabilidad de la variable aleatoria con respecto a su media. La varianza se define como:

$$Var[Z] = E[(Z - m)^2] \quad (1.14)$$

Como una medida de dispersión la varianza tiene el inconveniente de proporcionar valores con unidades que son el cuadrado de las que tiene la variable aleatoria. Otra medida de dispersión es la desviación estándar, S , definida como la raíz cuadrada de la varianza.

Si se desea comparar la varianza o la desviación estándar de varias variables con diferentes unidades entre sí, la comparación resultará injusta debido a que las distribuciones de las variables aleatorias con unidades pequeñas tendrán, aparentemente, menor variación. Para hacer este tipo de comparaciones se define otro indicador denominado *coeficiente de variación*, como el cociente de la desviación estándar entre la media, es decir



$$CV = \frac{S}{m} \quad (1.15)$$

El coeficiente de variación expresa la magnitud de la dispersión de una variable aleatoria con respecto a su valor esperado, es una medida estandarizada de la variación con respecto a la media.

1. 2. 4. MOMENTOS DE UNA DISTRIBUCIÓN BIVARIADA ⁽⁹⁾

Hasta ahora hemos considerado solamente una variable aleatoria a la vez. Sin embargo en ciencias de la tierra es muy importante la relación que puede existir entre dos variables, sobre todo si se desea estimar el valor de una de las variables a partir de la otra.

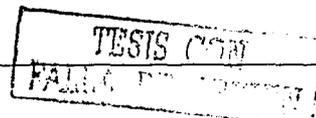
Así como la distribución de los resultados de una sola variable aleatoria Z , está caracterizada por su función de distribución acumulativa $F(z) = \text{Prob}(Z \leq z)$, la distribución conjunta de los resultados de un par de variables aleatorias Z e Y , está caracterizada por la función de distribución acumulada bivariada:

$$F(x, y) = \text{Prob}(X \leq x, Y \leq y) \quad (1.16)$$

El equivalente bivariado de un histograma es un *diagrama de dispersión*, donde cada par de datos (x_i, y_i) correspondientes a dos variables es graficado como un punto, Figura 1.4. El grado de dependencia entre las dos variables X e Y puede ser caracterizado por la dispersión de los puntos alrededor de la línea de 45° ; donde una perfecta dependencia lineal correspondería a todos los puntos graficados sobre esa línea. Figura 1.4.

De esta manera, el momento de inercia del gráfico de dispersión alrededor de la línea de 45° podría ser una característica de la carencia de dependencia. Este momento de inercia es llamado el *semivariograma* del conjunto de pares de datos (x_i, y_i) , $i = 1, \dots, N$.

$$\begin{aligned} \gamma(X, Y) &= \frac{1}{2N} \sum_{i=1}^N (x_i - y_i)^2 \\ &= \frac{1}{2} E\{[X - Y]^2\} \end{aligned} \quad (1.17)$$



El variograma, $2\gamma(X, Y)$, no es otra cosa que el promedio de la diferencia cuadrada entre dos componentes de cada par. Entre mayor sea el valor del variograma, mayor será la dispersión y menor relación habrá entre las variables.

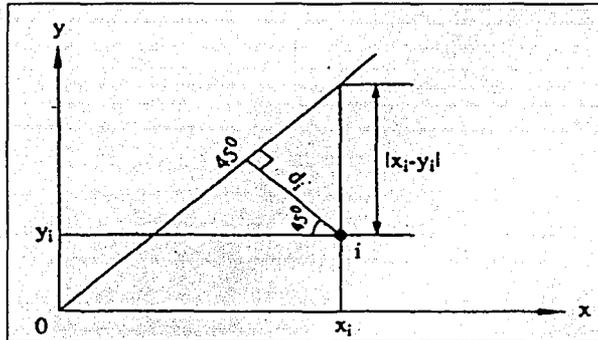


Figura 1.4. Par (x_i, y_i) sobre un diagrama de dispersión ⁽⁹⁾

Una alternativa al variograma es el momento bivariado definido como la covarianza:

$$C(X, Y) = E\{[X - m_x][Y - m_y]\} = E\{XY\} - m_x \cdot m_y \quad (1.18)$$

El variograma y la covarianza están relacionados mediante la ecuación:

$$\gamma(X, Y) = C(0) - C(X, Y) \quad (1.19)$$

1. 2. 5. VARIABLES REGIONALIZADAS ⁽⁹⁾

En la introducción se mencionó que los atributos que tienen una localización específica asociada con cada observación, es decir, que tienen una distribución a través del espacio, reciben el nombre de *variables regionalizadas*. Usualmente estas variables representan una característica de cierto fenómeno al que se le denomina *regionalización*. El espacio puede tener una o más dimensiones y puede incluir al tiempo. Por ejemplo, una característica geológica tal como el espesor de una formación puede ser representado como la distribución en dos dimensiones de la variable *espesor*. De hecho casi todas las

variables encontradas en ciencias de la tierra pueden ser consideradas como variables regionalizadas.

La definición de una variable regionalizada como una variable distribuida en el espacio es puramente descriptiva y no contempla alguna interpretación probabilística. Matemáticamente, una variable regionalizada se puede representar como una función $f(x)$ que toma un valor en cada punto x , de coordenadas (x_u, x_v, x_z) , para el caso del espacio en tres dimensiones. Sin embargo, la mayoría de las veces esta función varía tan irregularmente en el espacio que evita su estudio matemático directo.

Una variable regionalizada posee dos características aparentemente contradictorias:

- a) Un aspecto local, aleatorio; lo cual hace recordar la noción de variable aleatoria.
- b) Un aspecto general estructurado, el cual requiere una cierta representación funcional.

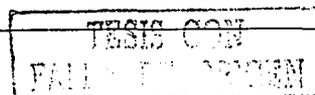
Una adecuada formulación debe tomar en cuenta este doble aspecto de aleatoriedad y estructura, de manera que proporcione una representación simple de la variabilidad espacial y conduzca a una aproximación consistente y operacional de la solución del problema. Una de estas formulaciones es la interpretación probabilística proporcionada por las funciones aleatorias.

1. 2. 6. FUNCIONES ALEATORIAS ^{(6), (7)}

Para introducir el concepto de función aleatoria consideremos el caso específico de la porosidad, por ejemplo si se tiene un valor de porosidad $z(x_1) = 10\%$ en un punto particular x_1 dentro de un yacimiento, esta porosidad puede ser considerada como una realización particular de una cierta variable aleatoria $Z(x_1)$, definida en el punto x_1 . De esta manera, el conjunto de porosidades para todos los puntos x dentro del yacimiento, es decir, la variable regionalizada $Z_R(x)$, puede ser considerado como una realización particular del conjunto de variables aleatorias $\{Z(x), x \in \text{yacimiento}\}$. A este conjunto de variables aleatorias se le llama *función aleatoria*, el cual se representa como $Z(x)$.

Esta definición de una función aleatoria expresa los aspectos aleatorio y estructurado de una variable regionalizada:

- a) Localmente, en un punto x_1 , $Z(x_1)$ es una variable aleatoria



- b) $Z(x)$ también es una función aleatoria en el sentido de que para cada par de puntos, (x_1) y $(x_1 + h)$, donde h representa una distancia de separación entre éstos, las correspondientes variables aleatorias $Z(x_1)$ y $Z(x_1 + h)$, no son en general, independientes, sino que están relacionadas mediante una correlación expresando la estructura espacial de la variable regionalizada inicial $Z_R(x)$.

La interpretación probabilística de una variable regionalizada $Z_R(x)$ como una realización particular de una cierta función aleatoria $Z(x)$, tiene un sentido operativo solamente cuando es posible inferir toda o parte de la ley de probabilidad que define a esta función aleatoria en su totalidad; sin embargo, no es posible inferir rigurosamente esta ley a partir de una sola realización $Z_R(x)$, la cual está, además, limitada a un número finito de puntos muestra z . Se requieren numerosas realizaciones $Z_1(x)$, $Z_2(x)$, ..., $Z_k(x)$ de la función aleatoria $Z(x)$, para inferir su ley de probabilidad. Como en la práctica estamos limitados a una sola realización $\{Z(x_i)\}$ de la función aleatoria en las posiciones x_i , se necesitan ciertas consideraciones. Estas consideraciones comprenden varios grados de homogeneidad espacial y se conocen como: *hipótesis de estacionaridad*.

En la práctica, aún si únicamente sobre una cierta región el fenómeno bajo estudio puede ser considerado como homogéneo, la variable regionalizada se repite así misma en el espacio. Esta homogeneidad o repetición proporciona el equivalente a numerosas realizaciones de la misma función aleatoria $Z(x)$, lo que permite una cierta *inferencia estadística*. De esta manera dos valores experimentales $z(x_0)$ y $z(x_0+h)$, en dos diferentes puntos x_0 y x_0+h , pueden ser considerados como dos realizaciones diferentes de la misma variable aleatoria $Z(x_0)$.

Así como una variable aleatoria está caracterizada por su función de distribución acumulada, una función aleatoria $Z(x)$ está caracterizada por el conjunto de todas sus funciones de distribución acumuladas K -variadas, para cualquier número K y cualquier selección de las K localizaciones x_k , $k = 1, \dots, K$:

$$F(z_1, \dots, z_k) = \text{Prob}\{Z(x_1) \leq z_1, \dots, Z(x_k) \leq z_k\} \quad (1.20)$$

De particular interés es la función de distribución acumulada bivariada ($K = 2$) de cualesquiera dos variables aleatorias.

ESTACIONARIDAD

Considérese una función aleatoria $Z(x)$ cuya ley de probabilidad univariada no depende de la localización x ; y similarmente, si la ley de probabilidad bivariada de $Z(x)$ y $Z(x+h)$ no depende de la posición x , sino solamente de la separación h , entonces a esta independencia de las leyes de probabilidad univariada y bivariada de la localización x se le denomina *estacionaridad*.

La estacionaridad es definida a través de los momentos de primer y segundo orden de la función aleatoria observada, y los grados de estacionaridad corresponden a los momentos particulares que permanecen invariantes a través del área de estudio.

La estacionaridad de segundo orden de una función aleatoria requiere que:

1. El valor esperado $E[Z(x)] = m$ no dependa de x , es decir, el valor esperado no varía a través del área de estudio.
2. La covarianza depende solamente de la distancia de separación h , es decir

$$C(h) = E[Z(x) \cdot Z(x+h)] - m^2 \text{ para toda } x \quad (1.21)$$

El vector h puede ser de una dimensión o mayor.

Si la covarianza $C(h)$ es estacionaria, la varianza y el variograma también son estacionarios:

$$C(0) = E\{[Z(x) - m]^2\} = VAR\{Z(x)\} \quad (1.22)$$

$$\gamma(h) = \frac{1}{2} E\{[Z(x+h) - Z(x)]^2\} \quad (1.23)$$

Una adecuada decisión de estacionaridad es crítica para la representatividad y fiabilidad de las herramientas geoestadísticas utilizadas. Por ejemplo, el conjuntar datos a través de diferentes facies geológicas puede enmascarar importantes diferencias estructurales; por otro lado, el dividir los datos en demasiadas subcategorías puede conducir a estadísticas poco representativas basadas en muy pocos datos por categoría. La regla en inferencia estadística es conjuntar la mayor cantidad de información *relevante* para formular predicciones correctas. ⁽⁷⁾

TESIS CON
FALLA DE ORIGEN

1. 2. 7. APLICACIÓN DE LAS FUNCIONES ALEATORIAS EN LA DETERMINACIÓN DE LA CONDICIÓN DE INSEGAMIENTO ⁽⁵⁾

Una función aleatoria es un modelo puramente conceptual que se selecciona cuando no se dispone de modelos determinísticos precisos que describan adecuadamente el fenómeno de interés. Por ejemplo, si se deseara estimar el comportamiento de un atributo geológico a partir de siete muestras disponibles, en cada punto no muestreado se podría estimar el valor real desconocido utilizando una combinación lineal ponderada de éstas muestras:

$$\hat{z} = \sum_{j=1}^7 w_j \cdot z_j \quad (1.24)$$

El problema sería cómo seleccionar los pesos para que el *promedio del error de estimación* sea cero.

Si se define el error r_i de cualquier valor particular estimado, como la diferencia entre el valor estimado y el valor real en esa misma localización

$$\text{error del } i\text{ésimo estimado} = r_i = \hat{z}_i - z_i, \quad (1.25)$$

entonces, el error promedio de un conjunto de n estimados sería

$$\text{error promedio} = m(r) = \frac{1}{n} \sum_{i=1}^n r_i = \frac{1}{n} \sum_{i=1}^n \hat{z}_i - z_i \quad (1.26)$$

Sin embargo aún no se ha resuelto el problema de cómo determinar los pesos para que el error promedio sea cero. La solución probabilística a este problema consiste en conceptualizar el perfil desconocido como el resultado de un proceso aleatorio y resolver el problema para nuestro modelo conceptual. Para cualquier punto en el que se desee estimar un valor desconocido, el modelo propuesto será una función aleatoria estacionaria formada por ocho variables aleatorias, una para el valor en cada una de las siete localizaciones muestreadas, $Z(x_1), \dots, Z(x_7)$, y una para el valor desconocido en el punto que se trata de estimar, $Z(x_0)$. Cada una de estas variables aleatorias tendrá la misma ley de probabilidad; en las ocho localizaciones, el valor esperado de la variable aleatoria será $E\{Z\}$. Cada par de variables aleatorias tendrá una distribución conjunta que dependerá solamente de la separación entre los dos puntos y no de sus localizaciones.

El estimado es así mismo una variable aleatoria ya que resulta de una combinación lineal ponderada de las variables aleatorias en las siete localizaciones muestreadas:

$$Z^*(x_0) = \sum_{i=1}^7 w_i \cdot Z(x_i) \quad (1.27)$$

Similarmente el error de estimación, es también una variable aleatoria:

$$R(x_0) = Z^*(x_0) - Z(x_0) \quad (1.28)$$

Sustituyendo (1.27) en (1.28) se puede expresar a $R(x_0)$ como:

$$R(x_0) = \sum_{i=1}^7 w_i \cdot Z(x_i) - Z(x_0) \quad (1.29)$$

Si se realiza la estimación en n localidades, entonces el promedio de los n errores es en si mismo una variable aleatoria:

$$\text{Error promedio} = A = \frac{1}{n} \sum_{i=1}^n R_i \quad (1.30)$$

En esta ecuación, las variables aleatorias R_1, \dots, R_n son los errores en las n localidades en las cuales se están calculando los estimados. El valor esperado de A estaría dado por:

$$E\{A\} = E\left\{\frac{1}{n} \sum_{i=1}^n R_i\right\} = \frac{1}{n} \sum_{i=1}^n E\{R_i\} \quad (1.31)$$

Una forma de garantizar que el valor esperado de A sea cero es insistiendo en que cada una de la variables aleatorias R_1, \dots, R_n tenga un valor esperado de cero. Se puede asegurar que esto pase aplicando el operador valor esperado a la ecuación (1.29):

$$E\{R(x_0)\} = E\left\{\sum_{i=1}^7 w_i \cdot Z(x_i) - Z(x_0)\right\} = \sum_{i=1}^7 w_i E\{Z(x_i)\} - E\{Z(x_0)\} \quad (1.32)$$

Como inicialmente se asumió que la función aleatoria es estacionaria, esto permite expresar la ecuación (1.32) como:

$$E\{R(x_0)\} = \sum_{i=1}^7 w_i E\{Z\} - E\{Z\} \quad (1.33)$$

El valor esperado del error en cualquier localización particular, $E\{R(x_0)\}$ es conocido como el *sesgo*. Igualando este valor esperado a cero para asegurar resultados insesgados se obtiene:

$$E\{R(x_0)\} = 0 = E\{V\} \sum_{i=1}^7 w_i - E\{V\}$$

$$E\{V\} \sum_{i=1}^7 w_i = E\{V\}$$

$$\therefore \sum_{i=1}^7 w_i = 1 \quad (1.34)$$

La conclusión es que si se desean estimados insesgados (el valor esperado del error de estimación igual a cero), se deben utilizar forzosamente combinaciones lineales ponderadas cuyos pesos sumen uno. A este resultado se le conoce como la *condición de insesgamiento*.

No obstante que el problema se derivó de la intención práctica de hacer el error promedio de un conjunto particular de estimaciones igual a cero, éste ha sido resuelto a través de un modelo de función aleatoria. El que funcione o no, depende en realidad de lo apropiado del modelo. Aún si la función aleatoria estacionaria es un modelo apropiado, el error promedio real puede no ser cero. La condición de insesgamiento ha estado basada en igualar un valor esperado a cero, sin embargo una variable aleatoria con un valor esperado de cero puede continuar generando un conjunto de resultados cuya media muestral no sea cero. Los parámetros de una variable aleatoria no necesitan ser idénticos a las correspondientes estadísticas calculadas a partir de un conjunto de resultados observados. ⁽⁵⁾

TESIS CON
FALLA DE ORIGEN

II- GEOESTADÍSTICA

La Geoestadística trata del estudio de las variables regionalizadas que fluctúan en el espacio y/o tiempo. Ofrece una colección de herramientas estadísticas enfocadas a comprender y modelar la variabilidad espacial. ⁽⁷⁾

La principal razón para modelar la variabilidad espacial es evaluar la incertidumbre en la estimación. Tal análisis y modelado no puede ser realizado sin una clara comprensión del origen de los datos, incluyendo su interpretación geológica. ⁽⁹⁾

El método geoestadístico es un procedimiento que se puede describir en forma general en cuatro etapas: ⁽¹⁰⁾

La primera etapa consiste en conocer los datos a través de un análisis estadístico simple (cálculo de medias, varianzas, cuartiles, valores mínimos y máximos, histogramas, etc.), y el modelado de su continuidad espacial a través de variogramas.

La segunda etapa se enfoca hacia la determinación y cuantificación de la relación entre los diferentes conjuntos de datos disponibles (si es posible). Por ejemplo la proveniente de registros geofísicos y los datos sísmicos.

En la tercera etapa se utiliza lo que ha sido aprendido y encontrado en los datos para determinar la distribución espacial de los puntos de control (el mallado para el subsiguiente contorno gráfico). Las herramientas usadas en esta etapa son kriging o cokriging.

El último paso consiste en evaluar la precisión del mapa hecho en la etapa anterior, es decir, estimar la incertidumbre de los resultados utilizando simulación estocástica por ejemplo. La evaluación del riesgo es quizá la aportación más grande que la geoestadística provee en la solución de problemas de mapeo.

La ventaja del método geoestadístico es que propicia que se analicen detalladamente los datos antes de hacer un mapa. El método no solamente produce mapas con ciertas propiedades estadísticas deseables sino que también da una medida cuantitativa de la fiabilidad del mapa.

En los paquetes tradicionales de mapeo se empieza con una lista de datos x , y , z , y se produce un contorno del mapa. El proceso es bastante simple: entran datos, sale un contorno. Hay pocos parámetros para ajustar. La Geoestadística usualmente toma más

tiempo que los métodos tradicionales de mapeo, pero la confiabilidad de los resultados compensan el esfuerzo extra.

2.1.-FUNDAMENTOS PARA EL MODELADO DEL VARIOGRAMA ^{(10), (5), (11), (12)}

El variograma (o semivariograma) es una técnica matemática para cuantificar la correlación espacial o continuidad de una variable. El variograma se requiere en cualquier procedimiento geoestadístico para la predicción fuera de los puntos de control.

En el capítulo uno se introdujo la función semivariograma como el momento de una distribución bivariada. Dada una muestra de observaciones $z(x_i)$, $i = 1, 2, \dots, n$, y considerando que la media es constante, se puede obtener un estimador del semivariograma:

$$\hat{\gamma}(h) = \frac{1}{2n(h)} \sum_{i=1}^{n(h)} [z(x_i + h) - z(x_i)]^2 \quad (2.1)$$

El término $n(h)$ es el número de pares dentro de las observaciones que se encuentran a una distancia h de separación. El "semi" en el término semivariograma viene del hecho de que es una varianza dividida por dos.

Frecuentemente se tiende a pasar por alto la constancia de la media requerida para estimar un variograma. Una suave y sistemática variación en la media, tal como el aumento de la temperatura con la profundidad por ejemplo, recibe el nombre de *tendencia*. Las prácticas apropiadas de estimación del variograma requieren la eliminación de esta tendencia. Uno de los métodos que existen elimina esta tendencia substrayendo una función analítica (tal como una función polinomial) que modele la tendencia. ⁽¹²⁾

El variograma no es un solo número sino una función continua de una variable h llamada *espaciamiento*, la cual representa la distancia de separación entre los pares de datos. El espaciamiento en realidad es un vector, por lo que no sólo involucra la magnitud de la separación sino también al acimut de la línea a través de los pares de datos. En el caso más simple, el variograma es una función escalar de la distancia entre los puntos.

Al estimado del variograma real se le llama *variograma experimental*. La fórmula para calcular el variograma experimental involucra condiciones que dependen de las

localizaciones específicas. Al promediar, sin embargo, se supone que se cancela la dependencia con la localización y deja solamente una dependencia con la distancia h . Esta es una consideración más que un hecho. ⁽¹¹⁾

La gráfica del variograma muestra el promedio de la diferencias cuadradas entre los puntos dato contra la distancia entre éstos. Figuras 2.1 y 2.2. Los puntos de control que se encuentran cercanos entre sí, en promedio, no difieren en valor tanto como los puntos que se encuentran más alejados.

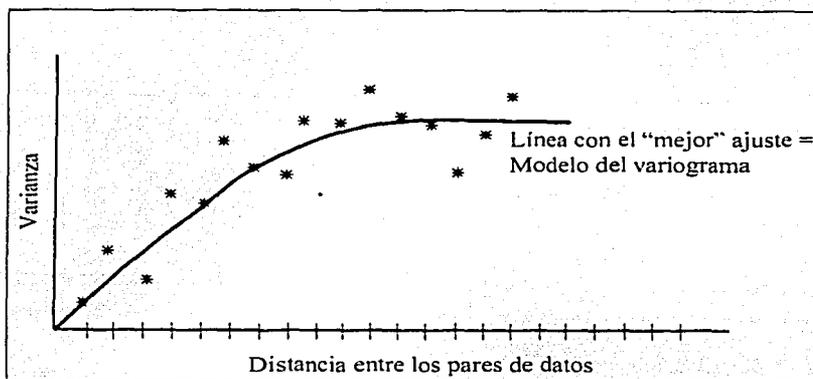


Figura 2.1. Variograma que expresa la continuidad espacial de una variable regionalizada. Es la gráfica del promedio de la diferencia cuadrada entre todos los pares de datos de la variable de interés a una distancia dada de separación. ⁽¹⁰⁾

Existe una regla que se aplica a cualquier estimación del variograma experimental: el número de pares tiende a disminuir con el incremento en el espaciamiento hasta un cierto límite, más allá del cual no hay más pares de datos. Debido a que la exactitud de las estimaciones es proporcional al número de pares, a mayor espaciamiento, son menos confiables las estimaciones. Una práctica es limitar la estimación a espaciamientos con un mínimo de 30 pares. ⁽¹¹⁾

Si se quiere estimar el variograma para un espaciamiento grande cercano en magnitud al diámetro del área muestreada, entonces sistemáticamente los únicos pares de observaciones que están lo suficientemente apartados son aquellos que se encuentran localizados en los extremos del área, lo que excluye los puntos centrales. Esta consideración es la justificación para una segunda regla práctica que aconseja limitar el

espaciamiento en el variograma experimental a la mitad de la distancia extrema en el dominio de las muestras para la dirección analizada.

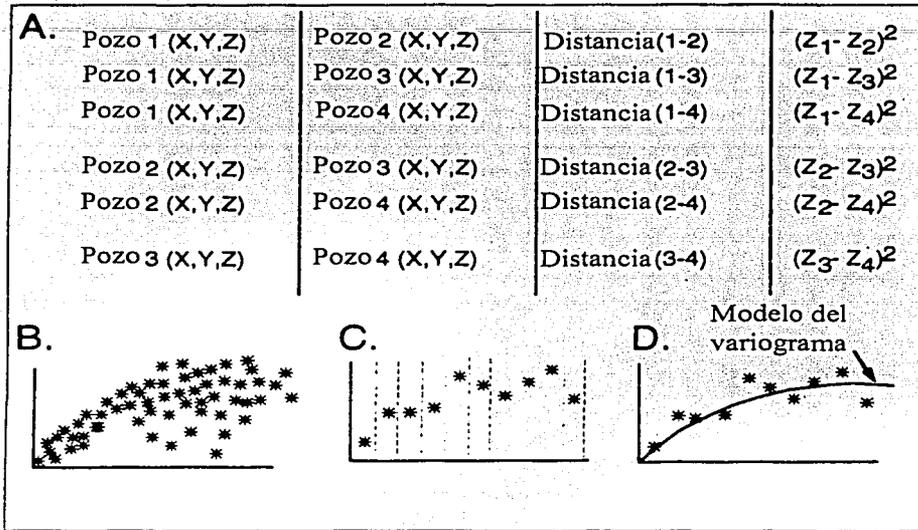


Figura 2.2. Cálculo del variograma. (A). Lista de todos los posibles pares de datos y cálculo de la distancia de separación $(Z_i - Z_j)^2$, donde Z es la variable de interés. (B) Grafico de la distancia de separación (eje de abscisas) vs. $(Z_i - Z_j)^2$. (C) Promedio de todos los datos en intervalos. (D) Ajuste de un modelo al variograma. ⁽¹⁰⁾

La situación más inconveniente para estimar el variograma es el más común: datos localizados irregularmente. El problema que se encuentra es que para cualquier espaciamiento h que se seleccione, existe la suficiente aleatoriedad en la localización de las muestras que muy pocos pares estarán separados "exactamente por h ". Por lo tanto, se tiene que tolerar esta inevitable aleatoriedad y aceptar cualquier par cuya separación sea cercana a h . En la práctica se especifican tolerancias, tanto en la distancia de separación como en la dirección. Un ejemplo típico se muestra en la Figura 2.3 donde se ha utilizado una tolerancia de ± 1 m sobre la distancia de separación, y $\pm 20^\circ$ en la dirección de h . Cualquier punto que quede dentro del área sombreada en este diagrama será aceptado como un punto localizado a 5 m de distancia en la dirección este.

TESIS CON
 FALLA DE CENSURA

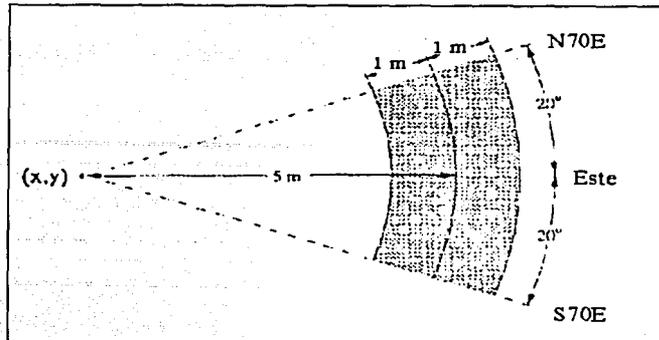


Figura 2.3. Ilustración de las tolerancias sobre h para la selección de pares de datos en el cálculo del variograma. ⁽⁵⁾

Además de las "reglas prácticas de dedo", no hay ninguna prueba estadística para evaluar cómo un variograma experimental se aproxima al verdadero variograma.

Una vez calculado el variograma experimental es necesario ajustarle un modelo adecuando ciertas funciones matemáticas a través de los datos. El modelado del variograma es una necesidad matemática que asegura que kriging trabajará.

Los parámetros clave que describen un modelo de variograma (Figura 2.4), son:

- Efecto pepita (nugget): es el valor del modelo a la distancia cero.
- Umbral (sill): representa la varianza de los datos.
- Rango: es la distancia que representa el punto de separación de la zona correlacionada a la zona no correlacionada del variograma.

EFFECTO PEPITA

El efecto pepita es una medida de la variabilidad a escala muy pequeña, implica cambios abruptos en la variable regionalizada a pequeñas distancias. Si éste fuera cero entonces los datos serían considerados exactos.

La presencia o ausencia del efecto pepita puede depender de la escala de muestreo relativa a la escala geográfica de variación del fenómeno. Tomando el ejemplo de un yacimiento fracturado, los pozos adyacentes pueden exhibir comportamientos muy diferentes que dependen de su proximidad relativa a las fracturas naturales. Si una roca

del yacimiento está muy fracturada en relación al espaciamiento normal de los pozos, la variabilidad debido al fracturamiento aparecería como un componente del efecto pepita.

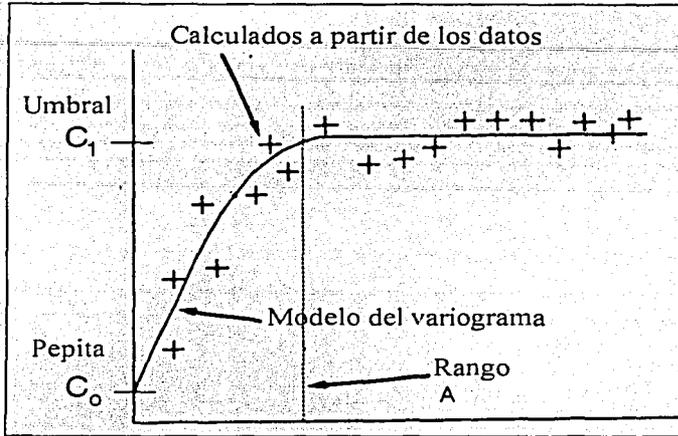


Figura 2.4. Elementos del variograma: El efecto pepita cuantifica la inconsistencia en las mediciones. El rango es el punto de separación entre los datos correlacionados y los no correlacionados.⁽¹⁰⁾

2.1. 1.- MODELOS TEÓRICOS PARA VARIOGRAMAS ^{(2), (7), (12)}

La necesidad de un modelo proviene del hecho de que podemos necesitar un valor del variograma para alguna distancia o dirección para la cual no tengamos un valor del variograma de las muestras. Se podría pensar en interpolar entre los valores conocidos de los variogramas de las muestras para resolver este problema; a pesar de que este método daría por resultado algún número, también daría un mayor problema. El uso directo del variograma experimental puede conducir a matrices singulares o soluciones múltiples en el sistema de ecuaciones del estimador kriging. La solución es reemplazar la tabulación obtenida para el variograma experimental por funciones admisibles que han sido probadas para asegurar una solución única a las ecuaciones resultantes. Es decir, se trata de funciones definidas positivas.

Sea V una función definida sobre algún dominio D que contiene el origen, entonces se dice que V es *definida positiva* sobre D si $V(0,0) = 0$ y $V(x, y) > 0$ para todos los demás puntos en D .

TESIS CON
FALLA DE ORIGEN

Los modelos más comunes son:

MODELO ESFÉRICO

$$\gamma(h) = \begin{cases} 1.5\left(\frac{h}{a}\right) - 0.5\left(\frac{h}{a}\right)^3 & \text{si } h \leq a \\ 1 & \text{si } h > a \end{cases} \quad (2.2)$$

donde a es el rango y h es el espaciamento vectorial.

Cerca del origen el variograma se parece a una línea. Se dice que el modelo es transitivo porque alcanza un umbral finito en un rango finito. (Figura 2.5.)

MODELO EXPONENCIAL CON RANGO PRÁCTICO a

$$\gamma(h) = 1 - \exp\left(\frac{-3h}{a}\right) \quad (2.3)$$

El modelo se aproxima al umbral (sill) asintóticamente. Se define un rango práctico a como la distancia a la cual el valor del modelo se encuentra a un 95% del umbral. (Figura 2.5)

MODELO GAUSSIANO

$$\gamma(h) = 1 - \exp\left(\frac{-3h^2}{a^2}\right) \quad (2.4)$$

Al igual que el modelo exponencial éste se aproxima al umbral asintóticamente. La gráfica del modelo tiene una forma parabólica cerca del origen. (Figura 2.5)

MODELO DE POTENCIAS

$$\gamma(h) = h^\omega \quad \text{con } 0 < \omega < 2 \quad (2.5)$$

En muchas situaciones dos o más modelos básicos, $\gamma_1(h)$ y $\gamma_2(h)$, deben ser combinados para ajustarse a la forma del variograma experimental (estructuras anidadas). Afortunadamente, una propiedad importante de los modelos positivos definidos de los variogramas establece que cualquier combinación lineal de éstos modelos con coeficientes positivos, es también un modelo positivo definido. Por lo tanto cualquier modelo de la forma

$$\gamma(h) = \sum_{i=1}^n w_i \cdot \gamma_i(h) \quad (2.6)$$

será también positivo definido.

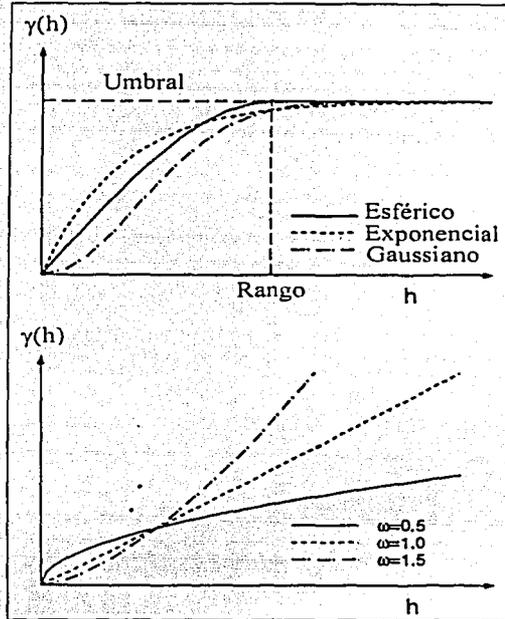


Figura 2.5. Modelos teóricos para variogramas con el mismo rango práctico (gráfica superior)⁽¹²⁾ y modelos de potencias para diferentes valores del parámetro ω , gráfica inferior.⁽¹²⁾

2. 1. 2.- ANISOTROPÍA ^{(5), (11)}

El análisis con variogramas puede ser usado para identificar y cuantificar el hecho de que la continuidad espacial puede ser mayor en una dirección que en otra definida por los puntos de control (datos). Un fenómeno se dice que es anisotrópico cuando su trayectoria de variabilidad espacial cambia con la dirección, por lo tanto, el modelado de la anisotropía requiere de funciones que dependan del vector \vec{h} , más que de la distancia h .

En este trabajo se presentarán los conceptos de la anisotropía en dos dimensiones. La extensión para el caso de tres dimensiones se puede encontrar en la referencia (5).

ANISOTROPÍA GEOMÉTRICA

Una anisotropía se dice que es geométrica cuando:

- 1) Los variogramas direccionales (variogramas calculados en una dirección específica) tienen la misma forma y umbral pero diferentes valores de rango, y
- 2) El diagrama de rangos, es decir, la gráfica de valores de rangos vs. el acimut θ de la dirección, es una elipse.

Si se considera por ejemplo, el diagrama de rangos elíptico de la Figura 2.6 (gráfica superior derecha), el eje mayor de la elipse correspondiente a la dirección de máxima continuidad, forma un ángulo θ con el eje de coordenadas y (dirección norte). Por convención, el ángulo del acimut θ se mide en el sentido de las manecillas del reloj a partir del eje y . La menor dirección de la anisotropía es perpendicular al eje mayor de la elipse y tiene un acimut $\phi = \theta + 90^\circ$.

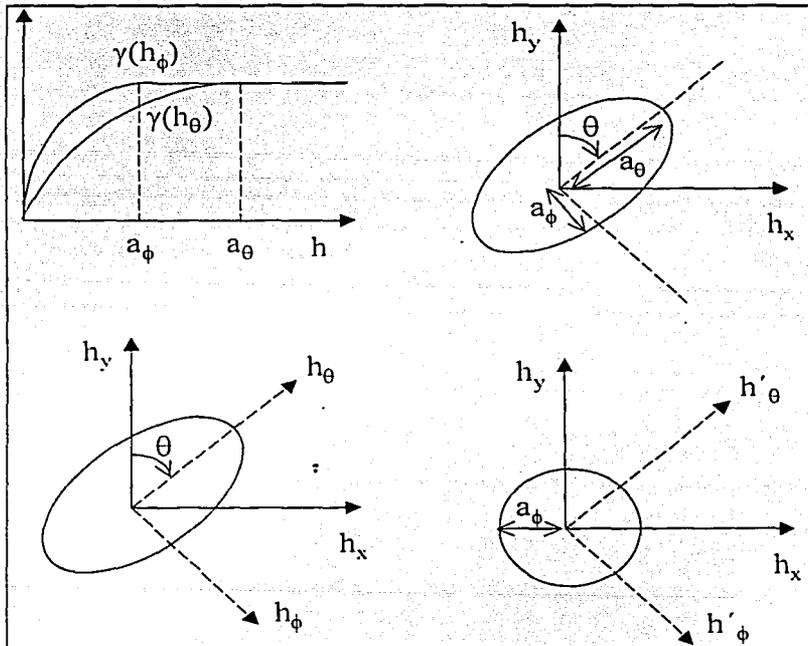


Figura 2.6. Anisotropía geométrica. ⁽⁵⁾

Los dos variogramas calculados en la dirección del acimut ϕ y θ son ajustados por modelos esféricos con el mismo valor de umbral pero diferentes valores de rangos, Figura 2.6 (extremo superior izquierdo). Los rangos mayor y menor de anisotropía, a_θ y a_ϕ , son graficados como el radio mayor y menor de la elipse. El factor de anisotropía, λ , se define como la relación entre el rango menor y el rango mayor. $\lambda = a_\theta/a_\phi < 1$

La corrección por anisotropía consiste en transformar el vector de coordenadas originales $h = (h_x, h_y)^T$, en un nuevo vector $h' = (h'_\phi, h'_\theta)^T$, de tal manera que el valor del modelo del variograma anisotrópico $\gamma(h)$ se identifique como un modelo isotrópico $\gamma'(|h'|)$ en el nuevo sistema de coordenadas:

$$\gamma(h) = \gamma'(|h'|) \text{ con } |h'| = \sqrt{h'^2_\phi + h'^2_\theta} \quad (2.7)$$

donde $\gamma'(\cdot)$ es un modelo isotrópico con un rango igual al rango menor a_ϕ de anisotropía.

La transformación se realiza en dos etapas:

- 1) El eje de coordenadas es rotado en el sentido de las manecillas del reloj de tal manera que identifique el eje principal de la elipse (Figura 2.6, gráfica inferior izquierda). El ángulo de rotación corresponde al acimut θ . El nuevo vector de coordenadas, denotado como $(h_\phi, h_\theta)^T$ es deducido a partir de

$$h_\theta = \begin{bmatrix} h_\phi \\ h_\theta \end{bmatrix} = \begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} h_x \\ h_y \end{bmatrix} \quad (2.8)$$

- 2) La elipse es reescalada a un círculo de radio igual al rango menor a_ϕ (Figura 2.6, gráfica inferior derecha). El reescalamiento de las nuevas coordenadas $(h_\phi, h_\theta)^T$ es escrito como

$$h_\theta = \begin{bmatrix} h'_\phi \\ h'_\theta \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \cdot \begin{bmatrix} h_\phi \\ h_\theta \end{bmatrix} \quad (2.9)$$

ANISOTROPÍA ZONAL

Una anisotropía que comprenda valores del umbral que varían con la dirección se dice que es zonal (Figura 2.7, gráfica derecha). El variograma en la dirección del acimut ϕ tiene un rango a_ϕ más grande y también un umbral más grande que en las otras direcciones.

Tal anisotropía puede ser modelada como la suma de un modelo isotrópico de transición $\gamma_1(|h|)$ y un modelo zonal $\gamma_2(h_\phi)$, el cual depende solamente de la distancia h_ϕ en la dirección de mayor varianza:

$$\gamma(h) = \gamma_1(|h|) + \gamma_2(h_\phi) \quad (2.10)$$

donde el modelo $\gamma_2(\)$ tiene un rango a_ϕ .

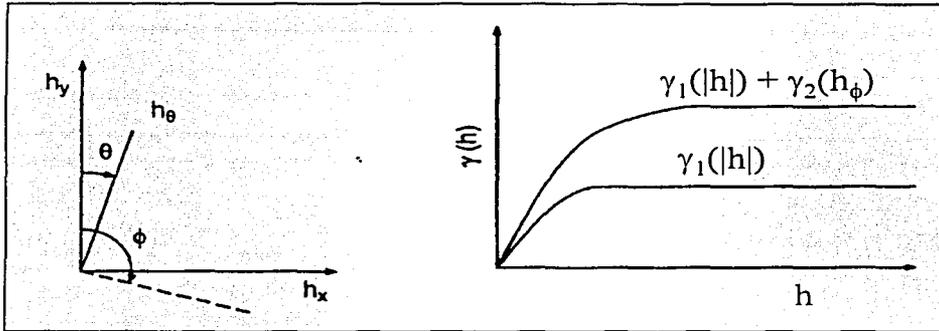


Figura 2.7. Ejemplo de anisotropía zonal en la dirección del acimut ϕ (varianza mayor). El modelo anisotrópico consiste de un modelo isotrópico $\gamma_1(|h|)$ y un componente zonal $\gamma_2(h_\phi)$, el cual depende solamente de la distancia h_ϕ .⁽⁵⁾

El componente $\gamma_2(h_\phi)$ se puede ver como un caso extremo del modelo de anisotropía geométrica; su modelado consiste de dos etapas:

1. El primer paso consiste en la rotación en el sentido de las manecillas del reloj de los ejes de coordenadas de tal manera que el eje y (norte), identifique la dirección de máxima continuidad definida como la dirección perpendicular a la de mayor varianza (umbral más alto). Para el ejemplo de la Figura 2.7 (gráfica izquierda), el ángulo de rotación es el ángulo acimut $\theta = \phi - 90^\circ$. Como para la anisotropía geométrica, el nuevo vector de coordenadas $(h_\phi, h_\theta)^T$ es calculado como:

$$h_\rho = \begin{bmatrix} h_\phi \\ h_\theta \end{bmatrix} = \begin{bmatrix} \cos \theta & -\text{sen} \theta \\ \text{sen} \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} h_x \\ h_y \end{bmatrix} \quad (2.11)$$

2. Los nuevos ejes son reescalados de tal manera que el modelo zonal no contribuye a la dirección de máxima continuidad (acimut θ). Este reescalamiento equivale a

fijar el rango a_θ en esa dirección al infinito, por lo que el factor anisotrópico $\lambda = a_\theta/a_\phi$, tiende a cero:

$$h_\theta = [h'_\theta] = [1 \quad 0] \cdot \begin{bmatrix} h_\phi \\ h_\theta \end{bmatrix} \quad (2.12)$$

Considerando por ejemplo, el componente zonal $\gamma_2(h_\theta)$ en la figura 2.8 (gráfica derecha), este modelo se puede ver como un modelo anisotrópico $\gamma_2(h)$ que identifica a un modelo esférico de rango a_θ en la dirección de mayor varianza pero que no contribuye en la dirección perpendicular del acimut θ . Con el nuevo sistema de coordenadas $(h'_\phi, h'_\theta)^T$, el modelo anisotrópico $\gamma_2(h_\theta)$ puede ser expresado como un modelo esférico isotrópico $\gamma'_2(|h'|)$ con rango a_ϕ :

$$\gamma_2(h) = \gamma'_2(|h'|) = Sph \left(\frac{|h'|}{a_\phi} \right) \quad (2.13)$$

con

$$Sph \left(\frac{|h'|}{a_\phi} \right) = \begin{cases} 1.5 \frac{|h'|}{a_\phi} - 0.5 \left(\frac{|h'|}{a_\phi} \right)^3 & \text{si } |h'| \leq a_\phi \\ 1 & \text{en otro caso} \end{cases}$$

donde

$$|h'| = \sqrt{h_\phi'^2 + h_\theta'^2} = h_c$$

2. 1. 3.- VARIOGRAMA Y COVARIANZA ^{(5), (12)}

El variograma y la covarianza están relacionadas mediante la ecuación:

$$\gamma(h) = C(0) - C(h) \quad (2.14)$$

No obstante que ambos momentos se refieren a la razón de la variación espacial de un atributo, conforme aumenta el variograma la covarianza disminuye. Se puede decir que la covarianza mide la similitud y el variograma la disimilitud. Tal como se muestra en la Figura 2.8, un variograma típico aumenta la disimilitud con la distancia.

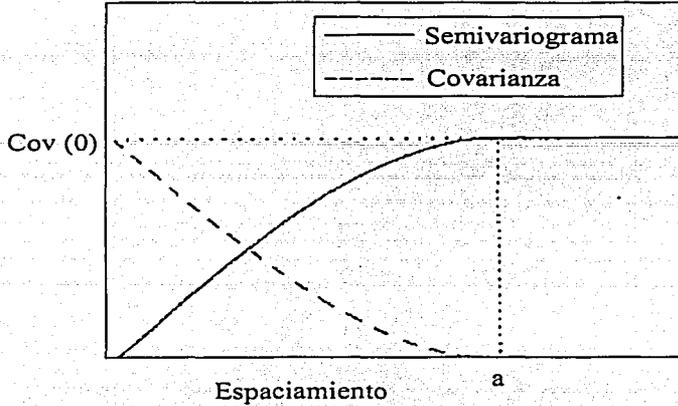


Figura 2.8. Ejemplos idealizados del variograma y la covarianza mostrando la localización del rango a .⁽¹²⁾

El variograma de una variable regionalizada que posee estacionaridad de segundo orden exhibe un umbral $Cov(0)$, que es igual a la varianza de las muestras exhaustivas de la variable. Dos valores observados a una distancia de separación mayor que a implica que son independientes.

La estimación del variograma tiene una ventaja sobre la estimación de la covarianza, y es que el variograma no requiere el conocimiento de la media de la variable regionalizada.

2. 2.- KRIGING

Kriging es un algoritmo para la estimación de una variable regionalizada en cada nodo, sobre un área o dentro de un volumen, de manera tal que el cuadrado de la diferencia entre el valor del nodo de la malla y los puntos de control de alrededor, es consistente (en el sentido de mínimos cuadrados) con el modelo del variograma⁽¹⁰⁾. (Figura 2.9).

TESIS C. I.
FALLA DE ORIGEN

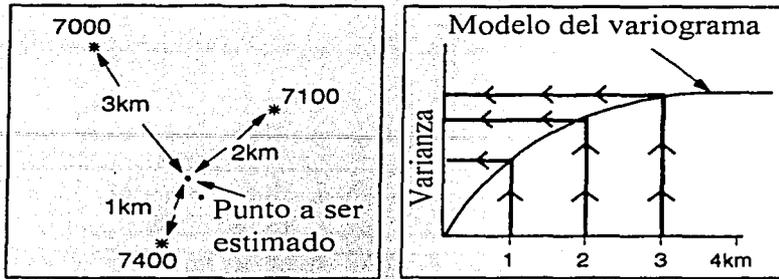


Figura 2.9. Kriging encuentra el valor del punto a ser estimado de tal manera que éste se ajuste de la mejor manera al modelo del variograma. ⁽¹⁰⁾

Kriging se obtiene matemáticamente mediante el cálculo del valor estimado como un promedio ponderado de los puntos de control alrededor de éste. Kriging toma en cuenta la distancia entre los puntos de control y el nodo de la malla a ser calculado y qué tan cerca se encuentran entre sí los puntos de control, manteniendo la relación espacial dada por el modelo del variograma. Los pesos son asignados de tal manera que *minimicen la varianza* del error de estimación en el sentido de mínimos cuadrados, de esta manera se elimina sistemáticamente el error por una sobrestimación o una baja estimación.

Una vez realizadas las estimaciones en cada nodo de la malla, la elaboración de un mapa de requiere dibujar los contornos conectados y suavizar las curvas (en caso de ser necesario) para hacer el mapa más real.

2. 2. 1. KRIGING SIMPLE ⁽⁹⁾

La idea básica de la regresión lineal es la estimación de un valor desconocido $z(x_0)$, mediante una combinación lineal de n valores conocidos $z(x_\alpha)$, $\alpha = 1, \dots, n$. Estos n valores conocidos pueden corresponder al mismo atributo que $z(x_0)$, en cuyo caso se habla de kriging, o a diferentes atributos en cuyo caso se habla de cokriging.

Un valor desconocido $z(x_0)$ es interpretado como una realización de una variable aleatoria $Z(x_0)$. Los n valores dato $z(x_\alpha)$, $\alpha = 1, \dots, n$ son asimismo interpretados como n realizaciones de n variables aleatorias $Z(x_\alpha)$, $\alpha = 1, \dots, n$.

Las $(n+1)$ variables aleatorias $Z(x_0), Z(x_1), \dots, Z(x_n)$ están caracterizadas por:

- Sus medias, por ahora se asume que son conocidas.

$$E\{Z(x_\alpha)\} = m(x_\alpha), \quad \alpha = 0, \dots, n$$

- Sus covarianzas:

$$\text{Cov}\{Z(x_\alpha), Z(x_\beta)\} = E\{Z(x_\alpha), Z(x_\beta)\} - m(x_\alpha) m(x_\beta) = C_{\alpha\beta} \quad \forall \alpha, \beta = 0, \dots, n$$

Las $(n+1)$ varianzas no son otra cosa que las autocovarianzas:

$$\text{Var}\{Z(x_\alpha)\} = E\{Z^2(x_\alpha)\} - m^2(x_\alpha) = C_{\alpha\alpha} \quad \forall \alpha = 0, \dots, n$$

ESTIMADOR LINEAL

Un valor desconocido z_0 es estimado mediante una combinación lineal de n datos mas un parámetro λ_0 :

$$z^*(x_0) = \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha z(x_\alpha) \quad (2.15)$$

$z^*(x_0)$ se dice que es el "estimado" del atributo analizado. En forma correspondiente, la combinación lineal de las n variables aleatorias $Z(x_\alpha)$ es asimismo una variable aleatoria llamada el "estimador":

$$Z^*(x_0) = \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha Z(x_\alpha) \quad (2.16)$$

Las ponderaciones λ_α son calculadas de acuerdo a los siguientes criterios:

- 1) El estimador es insesgado
- 2) La varianza de estimación es mínima.

El error real $z(x_0) - z^*(x_0)$, generalmente es desconocido, sin embargo algunos momentos de la variable aleatoria error, $Z(x_0) - Z^*(x_0)$, pueden ser calculados.

El valor esperado de la variable aleatoria error es:

$$E\{Z(x_0) - Z^*(x_0)\} = E\{Z(x_0)\} - E\{Z^*(x_0)\} \quad (2.17)$$

$$= E\{Z(x_0)\} - \lambda_0 - \sum_{\alpha=1}^n \lambda_\alpha E\{Z(x_\alpha)\} = m(x_0) - \lambda_0 - \sum_{\alpha=1}^n \lambda_\alpha m(x_\alpha) \quad (2.18)$$

Lo que deseamos es que el error esperado sea cero para asegurar insesgamiento, por lo que el parámetro λ_0 se toma como:

$$\lambda_0 = m(x_0) - \sum_{\alpha=1}^n \lambda_{\alpha} m(x_{\alpha}) \quad (2.19)$$

Sustituyendo (2.19) en (2.16) se obtiene:

$$Z^*(x_0) = m(x_0) + \sum_{\alpha=1}^n \lambda_{\alpha} [Z(x_{\alpha}) - m(x_{\alpha})] \quad (2.20)$$

El estimador insesgado $Z^*(x_0)$ aparece como el resultado de una estimación lineal del residual $[Z(x_0) - m(x_0)]$ a partir de la variable aleatoria residual de los datos $[Z(x_{\alpha}) - m(x_{\alpha})]$:

$$[Z^*(x_0) - m(x_0)] = [Z(x_0) - m(x_0)] = \sum_{\alpha=1}^n \lambda_{\alpha} [Z(x_{\alpha}) - m(x_{\alpha})] \quad (2.21)$$

Lo que resta es determinar los n pesos λ_{α} , $\alpha = 1, \dots, n$. Esto se realizará a través de la varianza (minimizada) del error.

Visualizando la variable aleatoria error $[Z(x_0) - Z^*(x_0)]$ como una combinación lineal Y de las $(n+1)$ variables aleatorias

$$Y = Z(x_0) - Z^*(x_0) = \sum_{\alpha=0}^n a_{\alpha} [Z(x_{\alpha}) - m(x_{\alpha})] \quad (2.22)$$

$$\text{con: } \begin{cases} a_0 = 1 \\ a_{\alpha} = -\lambda_{\alpha}, \alpha = 1, \dots, n \end{cases}$$

Y expresando la varianza del error como una doble sumatoria de las covarianzas $C_{\alpha\beta}$ (el desarrollo completo se puede encontrar en la referencia (9)), se obtiene:

$$\text{Var } Y = \sum_{\alpha=0}^n \sum_{\beta=0}^n a_{\alpha} a_{\beta} C_{\alpha\beta} \quad (2.23)$$

Lo que deseamos es que la varianza del error sea lo más pequeña posible, por lo tanto, los n pesos $a_{\alpha} = -\lambda_{\alpha}$, $\alpha = 1, \dots, n$ son seleccionados de tal manera que minimicen la $\text{Var } Y$. Esto se logra igualando a cero las n derivadas parciales de $\text{Var } Y$ con respecto a cada uno de los n parámetros a_{α} , $\alpha = 1, \dots, n$:



$$\frac{1}{2} \frac{\partial \text{Var} Y}{\partial a_\alpha} = \sum_{\beta=0}^n a_\beta C_{\alpha\beta} = a_0 C_{\alpha 0} + \sum_{\beta=1}^n a_\beta C_{\alpha\beta} \quad (2.24)$$

$$= C_{\alpha 0} - \sum_{\beta=1}^n \lambda_\beta C_{\alpha\beta} = 0, \alpha = 1, \dots, n \quad (2.25)$$

Lo que se ha establecido es lo que se conoce como "sistema normal" de ecuaciones, también conocido como ecuaciones de "regresión lineal", y renombradas como sistema "kriging simple":

$$\sum_{\beta=1}^n \lambda_\beta C_{\alpha\beta} = C_{\alpha 0}, \alpha=1, \dots, n \quad (2.26)$$

El sistema de ecuaciones es más fácil de visualizar en notación matricial:

$$K\lambda = k \Leftrightarrow \lambda = K^{-1}k \quad (2.27)$$

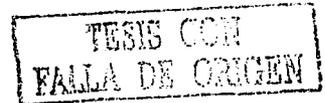
donde:

$$[\lambda] = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix}$$

$$[K] = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{bmatrix}$$

y

$$[k] = \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \end{bmatrix}$$



La correspondiente varianza minimizada del error, llamada también varianza de kriging simple es:

$$\sigma_{kS}^2 = \text{Var}\{Z(x_0) - Z^*(x_0)\} = C_{00} - \sum_{\alpha=1}^n \lambda_\alpha C_{\alpha 0} \geq 0 \quad (2.28)$$

Su deducción se puede consultar en la referencia (9).

Su forma matricial está dada por:

$$\sigma_{SK}^2 = C_{00} - \lambda^T \cdot k \quad (2.29)$$

El sistema (2.26) conduce a una y sólo una solución siempre que la matriz de covarianzas K sea positiva definida.

INTERPOLACIÓN ESPACIAL

La interpolación espacial corresponde al caso donde los n datos $Z(x_\alpha)$ se relacionan con el mismo atributo $Z(x)$ pero en diferentes ubicaciones, $x_\alpha \neq x_0$.

Por ejemplo, la porosidad $Z(x_0)$ en una localización no muestreada x_0 , será estimada a partir de n datos de porosidad $Z(x_\alpha)$ tomados en n localizaciones x_α alrededor de ésta:

$$Z^*(x_0) - m = \sum_{\alpha=1}^n \lambda_\alpha [Z(x_\alpha) - m] \quad (2.30)$$

donde m es el valor esperado común de las $(n + 1)$ variables aleatorias $Z(x_0)$, $Z(x_\alpha)$, $\alpha = 1, \dots, n$. En la práctica, m puede ser la media de todos los datos de porosidad disponibles dentro de la misma capa o formación homogénea.

Definiendo la función covarianza $C(h)$ como la covarianza entre cualesquiera dos variables aleatorias de porosidad $Z(x)$ y $Z(x+h)$, distantes por el vector h pero dentro de la misma formación:

$$C(h) = Cov\{Z(x), Z(x+h)\} \quad (2.31)$$

El sistema kriging simple será:

$$\sum_{\beta=1}^n \lambda_\beta C(x_\alpha - x_\beta) = C(x_\alpha - x_0), \quad \alpha = 1, \dots, n \quad (2.32)$$

El estimador de kriging simple (2.30) requiere el conocimiento previo o la estimación de la media común m . Ese paso previo puede ser concentrado con el proceso de la minimización de la varianza del error en un solo algoritmo, denominado "kriging ordinario".

TESIS CON
FALLA DE ORIGEN

2. 2. 2. KRIGING ORDINARIO ⁽⁹⁾

En todo el desarrollo de kriging simple se supuso que las medias eran conocidas. Veremos ahora un algoritmo de estimación que no requiera de este conocimiento previo de la media.

Considérese el estimador lineal:

$$Z^*(x_0) = \lambda_0 + \sum_{\alpha=1}^n \lambda_\alpha Z(x_\alpha) \quad (2.33)$$

La media del error será:

$$E\{Z(x_0) - Z^*(x_0)\} = m - \lambda_0 - \sum_{\alpha=1}^n \lambda_\alpha m = -\lambda_0 + m \left(1 - \sum_{\alpha} \lambda_\alpha\right) \quad (2.34)$$

y deberá ser igualada a cero, cualquiera que sea el valor desconocido de m. Esto sólo puede suceder si:

$$\begin{cases} \lambda_0 = 0 \\ \sum_{\alpha=1}^n \lambda_\alpha = 1 \end{cases} \quad (2.35)$$

De esta manera, se puede escribir un estimador lineal insesgado de $Z(x_0)$ como:

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_\alpha Z(x_\alpha), \quad \text{con: } \sum_{\alpha=1}^n \lambda_\alpha = 1 \quad (2.36)$$

La varianza del error es definida como:

$$\sigma_k^2 = \text{Var}\{Z(x_0) - Z^*(x_0)\} = \sum_{\alpha=0}^n \sum_{\beta=0}^n a_\alpha a_\beta C(x_\alpha, x_\beta) \quad (2.37)$$

con:

$$\begin{cases} a_0 = 1 \\ a_\alpha = -\lambda_\alpha, \alpha = 1, \dots, n \end{cases}$$

TESIS CON
FALLA DE ORIGEN

y:

$$C(x_\alpha, x_\beta) = \text{Cov}\{Z(x_\alpha), Z(x_\beta)\}$$

Lo que se desea es minimizar la varianza del error asegurando el insesgamiento del estimador, es decir, se debe cumplir que: $\sum_{\alpha=1}^n \lambda_\alpha = 1$. Estas restricciones dan por resultado un sistema de $(n+1)$ ecuaciones lineales conocido como "sistema kriging

ordinario". (El desarrollo completo de la deducción se puede encontrar en la referencia (9)).

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(x_{\alpha}, x_{\beta}) + \mu = C(x_{\alpha}, x_0), \quad \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \end{cases} \quad (2.38)$$

donde μ es un parámetro de Lagrange.

La correspondiente varianza minimizada del error, también llamada varianza de kriging ordinario, es:

$$\sigma_{k_0}^2 = E\{[Z(x_0) - Z^*(x_0)]^2\} = C(0) - \sum_{\alpha=1}^n \lambda_{\alpha} C(x_{\alpha}, x_0) - \mu \geq 0 \quad (2.39)$$

El sistema de ecuaciones de kriging ordinario también es más fácil de visualizar en notación matricial:

$$K\lambda = k \Leftrightarrow \lambda = K^{-1}k \quad (2.40)$$

donde:

$$[\lambda] = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix}$$

$$[K] = \begin{bmatrix} C_{11} & C_{12} & \dots & C_{1n} & 1 \\ C_{21} & C_{22} & \dots & C_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ C_{n1} & C_{n2} & \dots & C_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix}$$

y

$$[k] = \begin{bmatrix} C_{10} \\ C_{20} \\ \vdots \\ C_{n0} \\ 1 \end{bmatrix}$$

TESIS CON
FALLA DE ORIGEN

Al igual que el sistema de kriging simple, el sistema de kriging ordinario presenta una y sólo una solución, ya que la matriz de covarianzas de los datos, $K = [C(x_\alpha, x_\beta)]$, es positiva definida.

A pesar de que los sistemas de ecuaciones tanto para kriging simple como para kriging ordinario se pueden crear utilizando funciones de covarianzas o variogramas, la primera es preferida por razones computacionales. Escribiendo el sistema de ecuaciones en la forma de covarianzas se elimina el problema de ceros a lo largo de la diagonal, lo cual imposibilita a muchos algoritmos que tienen que dividir por estos términos. De esta manera el procedimiento más común es modelar la continuidad espacial de la variable regionalizada en términos de variogramas y después, transformar éstos a covarianzas a través de la relación (2.14).

Tradicionalmente kriging (simple u ordinario) ha sido utilizado para generar el "mejor" estimador lineal insesgado de los valores no muestreados $z(x)$, con la varianza kriging utilizada para definir intervalos de confianza de tipo gaussiano:

$$\text{Prob}\{Z(x) \in [z_k(x) \pm 2\sigma_k(x)]\} \cong 0.95 \quad (2.41)$$

Desafortunadamente las varianzas kriging del tipo (2.39) al ser independientes de los datos, solamente dan una comparación de alternativas configuraciones geométricas de los datos por lo que no miden la precisión de las estimaciones locales⁽⁷⁾.

2. 3.- SIMULACIÓN ESTOCÁSTICA

TESIS CON
FALLA DE ORIGEN

A finales de 1988 muchos de los métodos estocásticos para la caracterización de yacimientos aún no habían sido llevados a la práctica, se encontraban aún como trabajos teóricos por lo que existía la controversia sobre cuáles métodos tenían relevancia para el modelado de los yacimientos petroleros. El avance en el reconocimiento sobre el valor práctico de los métodos estocásticos fue tal, que tres años más tarde, en el Congreso de la SPE sobre Caracterización de Yacimientos en Crested, Butte, Colorado, la pregunta ya no era si los métodos eran adecuados, sino cómo elegir el método más apropiado para las condiciones específicas de cada yacimiento y los objetivos a lograr. ⁽²⁾

En la introducción de este trabajo se mencionaron algunos de los métodos estocásticos existentes, sin embargo nos concentraremos en la simulación secuencial por ser la que aplicaremos para el desarrollo de la tesis.

2. 3. 1. PRINCIPIOS DE LA SIMULACIÓN ESTOCÁSTICA ⁽⁷⁾

Considérese la distribución sobre un campo A de uno o más atributos $z(u)$, $u \in A$. La simulación estocástica es el proceso de construir modelos alternativos, igualmente probables, de alta resolución, de la distribución espacial de $z(u)$; cada realización es denotada con el superíndice i : $\{z^{(i)}(u), u \in A\}$. La simulación se dice que es "condicional" si todas las realizaciones resultantes respetan los valores de los datos en sus localizaciones.

La variable $z(u)$ puede ser categórica, por ejemplo indicando la presencia o ausencia de un tipo particular de roca, o puede ser continua tal como la porosidad en un yacimiento.

La simulación difiere de kriging o cualquier algoritmo de interpolación, en dos principales aspectos:

- 1) En muchos algoritmos de interpolación, incluyendo kriging, el objetivo es obtener el "mejor" estimado local de la variable sin considerar las estadísticas espaciales resultantes al considerar en conjunto a los estimados. Kriging provee un conjunto de representaciones locales, $z'(u)$, $u \in A$, donde prevalece la precisión local. La simulación genera alternativas representaciones globales, $z^{(i)}(u)$, $u \in A$, donde prevalece la reproducción de las trayectorias de continuidad espacial.
- 2) Excepto si se asume un modelo gaussiano para los errores, kriging proporciona solamente una medición incompleta de la precisión local, y no la apreciación conjunta de ésta cuando se consideran varias localizaciones. Las simulaciones son diseñadas específicamente para proporcionar tales mediciones de precisión, tanto local como considerando varias localizaciones. Estas mediciones están dadas por las diferencias entre L valores alternativamente simulados en cualquier localización (precisión local) o los L campos simulados alternativamente (precisión global).

2. 3. 2. SIMULACIÓN SECUENCIAL ⁽²⁾

Toda la familia de procedimientos secuenciales hacen uso del mismo algoritmo básico mostrado en la Figura 2.12.

1. Seleccionar en forma aleatoria un nodo de la malla en el cual aún no se haya simulado un valor.
2. Estimar la distribución de probabilidad condicional local (DPCL) en esa localización.

3. Determinar en forma aleatoria un solo valor a partir de la DPCL.
4. Incluir el nuevo valor simulado en el conjunto de datos condicionantes.
5. Repetir los pasos 1 a 4 hasta que todos los nodos de la malla tengan un valor simulado.

La única diferencia significativa en términos prácticos, entre los diversos procedimientos secuenciales es la forma en que la DPCL es estimada. Cualquier técnica que pueda producir un estimado de la DPCL puede ser usada como la base para la simulación secuencial. Kriging multivariado gaussiano por ejemplo, produce un estimado de la DPCL asumiendo que ésta sigue la clásica distribución normal en forma de campana y estimando su media y desviación estándar; esta es la aproximación usada en el diagrama de la Figura 2.10. En este caso al algoritmo se le llama *simulación secuencial gaussiana*.

Kriging indicador es un ejemplo de otra técnica que puede ser usada para estimar la DPCL. Con este procedimiento no se hace ninguna suposición acerca de la forma de la distribución, la cual es obtenida mediante la estimación directa de la probabilidad de estar debajo de una serie de umbrales. Cuando se utiliza este método para la simulación secuencial, el algoritmo se denomina simulación secuencial indicador.

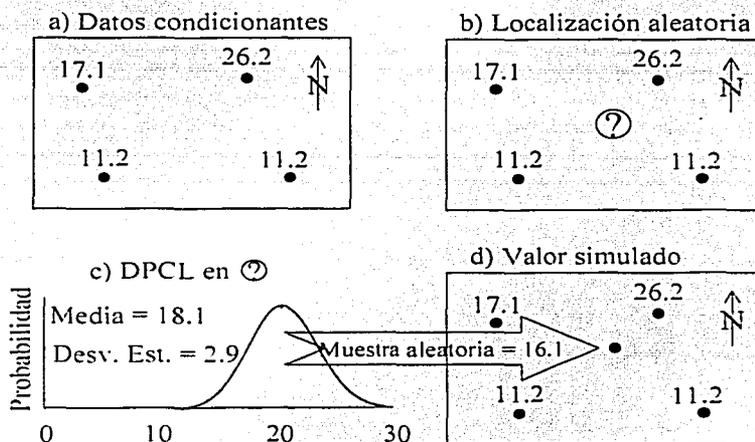


Figura 2.10. Simulación secuencial de porosidades en una capa. ⁽⁵⁾

2.3.3. SIMULACIÓN SECUENCIAL GAUSSIANA ⁽⁷⁾

Este algoritmo genera realizaciones de un campo gaussiano multivariado simulando secuencialmente cada variable de acuerdo a su función de distribución acumulada condicional, caracterizada totalmente a través de un sistema kriging simple u ordinario. Los datos condicionantes consisten de todos los datos originales y todos los valores previamente simulados que se encuentran dentro del vecindario de la localización que se esta simulando.

El procedimiento para la simulación condicional de una variable continua $z(u)$, modelada mediante una función aleatoria estacionaria gaussiana $Z(u)$ consiste en:

- 1) Determinar la función de distribución acumulada univariada $F(z)$ representativa de toda el área de estudio, y no solamente de las muestras de datos disponibles z . Se puede necesitar el desagrupamiento si los datos z están preferencialmente localizados.
- 2) Utilizando la función de distribución acumulativa $F(z)$, realizar la transformación de los datos z , a valores y , con una función de distribución acumulada normal estándar (normal score)
- 3) Verificar la normalidad bivariada de los valores y y transformados. Si el modelo gaussiano multivariado no puede ser retenido, entonces considerar modelos alternativos para la simulación.
- 4) Si se puede adoptar el modelo de función aleatoria multivariada gaussiana para la variable y , proceder con la simulación secuencial.
- 5) Definir una trayectoria aleatoria que visite cada nodo de la malla (no necesariamente regular) una sola vez. En cada nodo u , retener un número especificado de datos vecinos condicionantes incluyendo tanto los datos originales y , como los valores en los nodos previamente simulados.
- 6) Utilizar kriging simple u ordinario con el modelo del variograma de los valores de la distribución normal estándar (normal score) para determinar los parámetros (media y varianza) de la función de distribución acumulada de la función aleatoria $Y(u)$ en la localización u .
- 7) Generar un valor simulado $y^{(i)}(u)$ a partir de la función de distribución acumulada.
- 8) Adicionar el valor simulado $y^{(i)}(u)$ al conjunto de datos.

- 9) Proceder con el siguiente nodo de igual manera hasta que todos los nodos sean simulados.
- 10) Retransformar los valores normales simulados $\{y^{(l)}(u), u \in A\}$ en valores simulados de la variable original $z^{(l)}(u)$.

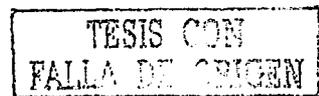
Si se desean múltiples realizaciones $\{z^{(l)}(u), u \in A\}, l = 1, \dots, L$, entonces el algoritmo previo es repetido L veces con cualquiera de estas opciones:

- a) La misma trayectoria aleatoria para visitar los nodos. En este caso la configuración de los datos y el sistema kriging son los mismos de una realización a otra, por lo que necesitan resolverse una sola vez. El tiempo de CPU es reducido considerablemente.
- b) Una trayectoria aleatoria diferente para cada realización. Por lo tanto la secuencia de configuraciones de datos es diferente; de esta manera deben de resolverse diferentes sistemas kriging. El tiempo de CPU es proporcional al número de realizaciones.

2. 3. 4. TRANSFORMACIÓN A UNA DISTRIBUCIÓN NORMAL ESTÁNDAR ^{(7), (5)}

La primera condición necesaria para que una función aleatoria sea multivariada gaussiana es que su función de distribución acumulada univariada sea normal. Desafortunadamente, muchos de los datos en ciencias de la tierra no presentan histogramas gaussianos simétricos. En tales casos se puede realizar una transformación normal estándar (una transformación no-lineal) sobre los datos originales. De esta manera los datos resultantes de la transformación adoptan un modelo gaussiano $Y(u)$, al menos en el caso univariado.

La Figura 2.11 muestra un procedimiento gráfico para la transformación de los valores de una distribución no-gaussiana a una distribución normal estándar



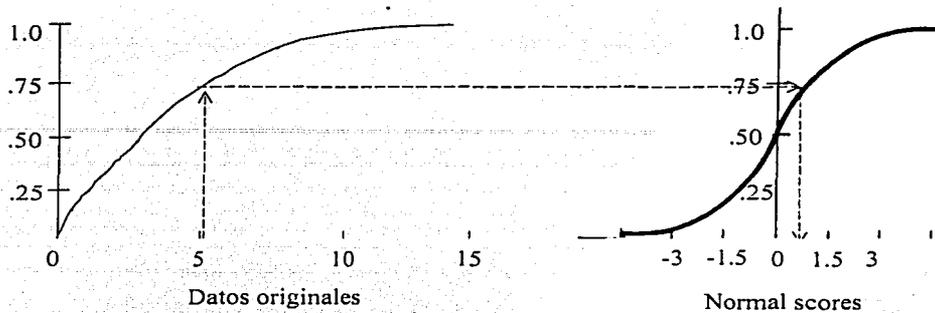


Figura 2.11. Procedimiento gráfico para la transformación de los valores de una distribución no-gaussiana a una distribución normal estándar. ⁽⁵⁾

2. 3. 5. VERIFICACIÓN DE LA NORMALIDAD BIVARIADA ⁽⁷⁾

La transformación a una distribución normal estándar vista en el punto anterior define una nueva variable Y , la cual es, por construcción, distribuida normalmente (univariada). Esta es una condición necesaria pero no suficiente para que los valores distribuidos espacialmente $Y(u)$, $u \in A$, sean multivariados gaussianos. La siguiente condición necesaria es que la función de distribución acumulada bivariada de cualquier par de valores $Y(u)$, $Y(u + h)$, $\forall u, \forall h$, sea normal. Los procedimientos para verificar esta condición pueden consultarse en la referencia (7).

La normalidad bivariada es necesaria pero no suficiente para la normalidad multivariada. En realidad, el procedimiento completo que muestra la suficiencia requiere que los datos sean numerosos y estén mallados, por lo tanto en la práctica, si no se puede a partir de las estadísticas de los datos mostrar que las propiedades gaussianas bivariadas son violadas, el modelo de función aleatoria gaussiano multivariado debe ser la primera opción para la simulación de variables continuas. ⁽⁷⁾

2. 3. 6. REESCALAMIENTO ⁽¹³⁾

A pesar que las técnicas geoestadísticas son capaces de generar descripciones del yacimiento a cualquier escala arbitrariamente pequeña, las simulaciones de flujo de fluidos a tales escalas pequeñas por ejemplo, requieren de computadoras tan poderosas no disponibles para toda la comunidad de ingenieros de yacimientos. De aquí que se

TESIS CON
FALLA DE ORIGEN

necesite una metodología donde la variabilidad geológica a pequeña escala representada por un número muy grande de bloques de mallado numérico puedan ser reescalados a niveles en los cuales la simulación del flujo sea práctica.

Parámetros aditivos, tales como la porosidad o la saturación de agua, no requieren más que el clásico promedio aritmético. La permeabilidad absoluta y relativa, sin embargo, representan problemas más complejos que requieren técnicas más refinadas.

2. 4.- SISTEMA DE COORDENADAS ESTRATIGRAFICAS ⁽¹⁴⁾

Una decisión importante en aplicaciones 3D se relaciona con el sistema de coordenadas. El yacimiento estudiado puede no tener un espesor uniforme, puede presentar plegamientos o fallas, de aquí que sus direcciones de anisotropía pueden variar a través del área con el echado local. Más aún, un modelo de variograma estacionario se requiere para kriging o los algoritmos de simulación.

Una solución consiste en dividir el área de estudio global en varias subáreas, cada una homogénea y suficientemente plana para ser modelada con variogramas estacionarios definidos a partir del tradicional sistema de coordenadas cartesiano.

Los problemas con esta solución son: a) La solución requiere de una gran cantidad de datos, suficiente para hacer la subdivisión y reevaluar localmente las direcciones de la anisotropía, o en algunos casos el modelo del variograma entero. b) Pueden ocurrir discontinuidades artificiales en las fronteras entre subáreas.

En casos donde la estratigrafía de la cima o el fondo de la capa pueda ser asumida y la variabilidad del espesor sea pequeño, se puede considerar la transformación de coordenadas propuesta por Journel y Gómez-Hernández:

$$x' = x, y' = y, z' = \frac{z(x, y) - \text{cima}(x, y)}{\text{espesor}(x, y)} \in [0, 1]$$

donde (x, y, z) son las coordenadas cartesianas originales, (x', y', z') son las coordenadas estratigráficas transformadas con la elevación z' definida en relación al espesor local de la capa.

Todos los variogramas, estimados y simulaciones son realizados en el sistema estratigráfico (x', y', z') , y los resultados son relocalizados en el sistema cartesiano original.

III. GEOMECÁNICA

Fairhurst define a la mecánica de rocas como el campo de estudio desarrollado para entender los procesos básicos de deformación de las rocas y su significado tecnológico.⁽¹⁵⁾

Las consecuencias de estos procesos en la ingeniería petrolera son considerables. Por ejemplo, el poder predecir el comportamiento mecánico de las formaciones es la clave para evitar la inestabilidad del agujero durante la perforación. Si la deformación de la roca resulta en una notable contracción del agujero debido al estado de esfuerzos inducido en la formación rocosa inmediatamente adyacente a la pared, el movimiento de la barrena puede ser restringido o la ubicación de la tubería de revestimiento, después de terminar la perforación, puede ser obstruida. Si la deformación resulta en una expansión del agujero, la roca puede fracturarse y resultar en una pérdida de circulación.⁽¹⁶⁾ La predicción del comportamiento mecánico de las rocas del yacimiento es esencial para un buen programa de terminación o estimulación. La compactación del yacimiento, que puede conducir a la subsidencia, es un factor crítico con respecto al comportamiento general del yacimiento.

3.1. PROPIEDADES ELÁSTICAS Y DE FALLA DE LA ROCA ^{(17), (18), (4)}

Los módulos elásticos son propiedades características de la roca que definen el comportamiento elástico entre los esfuerzos aplicados y las deformaciones generadas, estos módulos elásticos varían de acuerdo al tipo de roca.

DEFINICIÓN ESTÁTICA DE LOS MÓDULOS ELÁSTICOS

Módulo de Young (E).

Es una relación entre los esfuerzos y deformaciones longitudinales bajo la acción de una carga uniaxial. Se puede representar como:

$$E = \frac{\text{Esfuerzo Aplicado}}{\text{Deformación Normal}} = \frac{F/A}{\Delta L/L} \quad (3.1)$$

donde:

F/A es la carga por unidad de área

ΔL es el cambio en longitud

TESIS CON
FALLA DE ORIGEN

Relación de Poisson (ν).

Es la constante definida por la relación que existe entre la deformación lateral y la deformación longitudinal que sufre una roca cuando se le somete a un esfuerzo uniaxial.

$$\nu = \frac{\text{Deformación Lateral}}{\text{Deformación Longitudinal}} = \frac{\Delta D/D}{\Delta l/l} \quad (3.2)$$

donde ΔD es el cambio en diámetro.

Módulo de corte o rigidez (G).

Se define como la relación entre la fuerza tangencial sobre la cara de la roca y su deformación sin existir esfuerzos normales, se expresa como:

$$G = \frac{\text{Esfuerzo Tangencial}}{\text{Deformación de Corte}} = \frac{F/A}{\Delta l/l} \quad (3.3)$$

Módulo volumétrico (K).

Es la relación entre los esfuerzos compresionales (presión hidrostática) y la deformación volumétrica.

$$K = \frac{\text{Esfuerzo Compresivo}}{\text{Deformación Volumétrica}} = \frac{F/A}{\Delta V/V} \quad (3.4)$$

Los módulos elásticos están relacionados mediante las siguientes expresiones:

$$E = \frac{9KG}{3K + G} \quad (3.5)$$

$$\nu = \frac{3K - 2G}{2(3K + G)} \quad (3.6)$$

RESISTENCIA A LA COMPRESIÓN UNIAxIAL, σ_c .

La resistencia a la compresión uniaxial es una propiedad mecánica de falla de las rocas y numéricamente nos indica la dureza y la fuerza máxima que soportan las rocas antes de su fractura.

La resistencia a la compresión uniaxial presenta una relación inversa con la porosidad, ϕ , debido a que al incrementarse el esfuerzo de sobrecarga, la porosidad de la formación

decrece haciendo a la formación más compacta y por lo tanto más resistente a su fractura.

Las constantes elásticas pueden ser estáticas o dinámicas, según la forma que se emplee para determinarlas. Los métodos para hacerlo son variados y con diferente grado de precisión.

Las constantes estáticas se determinan en el laboratorio a través de la medición en núcleos mediante la aplicación gradual de esfuerzos, el tipo de esfuerzo dependerá del módulo que se quiera medir. Las pruebas de laboratorio más comunes que nos permiten determinar las propiedades de falla y las propiedades elásticas de la roca son las pruebas uniaxial y triaxial.

La gran ventaja de las pruebas de laboratorio es que podemos controlar las condiciones en que las realizamos. Es decir, podemos controlar los esfuerzos y deformaciones con gran precisión. Frecuentemente esto no es posible realizarlo en las pruebas de campo. La desventaja de las pruebas de laboratorio es que todas son llevadas a cabo con muestras que han sido modificadas de sus condiciones originales in situ.

El factor más importante a observar en la perturbación de las muestras es la calidad de la técnica de muestreo. Aún cuando una buena muestra es removida de un agujero sin sufrir daño considerable, durante su transporte al laboratorio puede sufrir algún deterioro, e incluso durante su misma preparación para la prueba.

Naturalmente ninguno de estos problemas se tiene en el campo. Las pruebas de campo manejan las muestras in situ, lo que da una gran ventaja. Los métodos de las pruebas en campo varían desde penetraciones simples de vibraciones, a pruebas con cargas más complejas y costosas. Es precisamente lo costoso de las pruebas una de las desventajas asociadas con las pruebas de campo. Las pruebas de campo menos costosas conducen a resultados con una cantidad importante de incertidumbre.

Los métodos de campo tienen otras desventajas adicionales a sus altos costos. Todas las pruebas de campo requieren someter un cierto volumen de la muestra in situ a una carga, y en muchas circunstancias este volumen no es precisamente conocido. También, se pueden generar presiones de poro por las cargas de las pruebas, y estas pueden distorsionar significativamente los resultados. Una tercera desventaja de las pruebas de campo es que ninguna prueba por sí sola, puede aportar valores con buena calidad para dos constantes elásticas. Algunas pruebas aportan una sola constante; por ejemplo la

prueba de "medición de presión" que permite conocer el módulo de corte. Otras pruebas proporcionan valores para una combinación algebraica de dos constantes, pero se necesita más información para encontrar sus valores individuales.

3. 2 CORRELACIONES

En la introducción se mencionó la utilidad de la porosidad como un índice geomecánico. ⁽¹⁾ Los trabajos antecedentes datan de 1956 y 1958 donde Wyllie y colaboradores introdujeron una relación semiempírica entre la porosidad y la velocidad acústica. Se considera a la porosidad como un factor importante de influencia en la elasticidad y el esfuerzo en las rocas ^{(19), (20)}. El esfuerzo de las areniscas ha sido reportado por Vernik (1993) y colaboradores, como decreciente en forma no lineal con el incremento de la porosidad ⁽²¹⁾. Similares relaciones han sido publicadas para carbonatos. Allison ⁽²²⁾ reportó en 1987 que el módulo elástico dinámico decrece exponencialmente con el incremento de la porosidad para rocas carbonatadas. Kamel y colaboradores encontraron en 1991 buenas correlaciones entre la porosidad determinada a partir de los registros de neutrón y densidad con el módulo elástico dinámico. ⁽²³⁾

En un trabajo presentado por Edlmann y colaboradores en 1998 sobre la predicción de las propiedades mecánicas de las rocas a partir de porosidades determinadas con registros, los autores determinaron ecuaciones matemáticas que correlacionan a la porosidad con las variables geomecánicas de la siguiente manera: el módulo de elasticidad (E) decrece con el incremento de la porosidad, la relación de Poisson (ν) se incrementa conforme incrementa la porosidad y la resistencia a la compresión uniaxial (C_0) decrece con el aumento de la porosidad. ⁽²⁴⁾

El trabajo de Edlmann y colaboradores reporta las propiedades mecánicas de las rocas de diferentes yacimientos con diferentes porosidades. Los autores desarrollan a partir de estos datos un nuevo método de predicción de las propiedades mecánicas de las rocas directamente de la porosidad. En el artículo se discute la medición de la porosidad derivada experimentalmente, el módulo elástico y los parámetros de esfuerzo de fractura, así como la interpretación de los resultados en correlaciones directas de las variables con la porosidad. Se enfatiza el uso práctico de estas propiedades mecánicas de la roca en la perforación, producción y simulación de yacimientos.

La porosidad, el módulo de elasticidad, la relación de Poisson, la resistencia a la compresión uniaxial y otras variables, fueron medidas en muestras de un amplio rango de yacimientos del Mar del Norte usando una máquina de prueba convencional triaxial. El artículo describe el procedimiento usado y presenta las correlaciones obtenidas a partir de las gráficas de cada una de las propiedades con la porosidad.

El trabajo de Gutiérrez-García ⁽⁴⁾ presenta el desarrollo de una metodología que permite determinar ecuaciones semi-empíricas para la predicción de las propiedades elasto-dinámicas como una función de la porosidad utilizando la teoría de autoconsistencia. Ellos determinaron también un modelo semi-empírico para mejorar la predicción de las propiedades de falla de la roca (resistencia a la compresión uniaxial) a partir de la impedancia acústica (módulo M), la cual se puede determinar a partir de los datos del registro sónico y de densidad. Al módulo M en el área de geomecánica en condiciones dinámicas se le conoce como módulo de onda P, en condiciones estáticas se le denomina módulo elástico restringido. En este trabajo sólo se le llamará módulo M.

Las ecuaciones propuestas por ellos de para un yacimiento real de México son:

$$M = 304.87^2 \frac{\rho}{\Delta t^2} \quad (3.7)$$

$$\frac{M}{M_s} = 0.98e^{-3.867\phi} \quad (3.8)$$

$$Co' = 2.54M_s - 21.17 \quad (3.9)$$

$$\frac{C}{Co'} = 0.1057e^{2.398\frac{M}{M_s}} \quad (3.10)$$

$$\frac{K}{K_s} = 0.991e^{(-5.92\phi)} \quad (3.11)$$

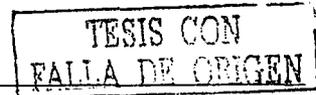
$$\frac{G}{G_s} = 0.992e^{(-8.29\phi)} \quad (3.12)$$

donde:

ϕ = porosidad

M = módulo M

M_s = módulo M del sólido de la roca



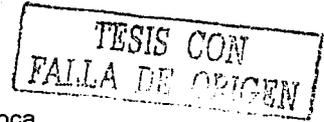
ρ = densidad de la roca

Δt = tiempo de tránsito

Co' = resistencia a la compresión uniaxial del sólido de la roca

K_s = módulo volumétrico del sólido de la roca

G_s = módulo de corte del sólido de la roca



3.3 COEFICIENTE DE COMPACTACIÓN UNIAxIAL, C_m .

La compactación de un yacimiento o la reducción de su volumen es fundamentalmente el resultado de una reducción en su espesor. Considerando que sus dimensiones laterales son grandes comparadas a su altura, el yacimiento se deforma predominantemente en el plano vertical. Por lo tanto, la compactación de la formación puede ser caracterizada convenientemente por la deformación vertical en el yacimiento, $\epsilon_z = \Delta z/z$, durante la producción, lo que expresa el cambio en altura (relativa a la altura inicial) causada por un incremento en el esfuerzo efectivo debido a una reducción en la presión de poro, bajo una sobrecarga constante.⁽²⁵⁾ Entonces se puede definir un *coeficiente de compactación uniaxial*, c_m , como la compactación de la formación por unidad de cambio en la reducción de la presión de poro:

$$c_m = \frac{1}{z} \frac{dz}{dp} \quad (3.13)$$

El coeficiente de compactación uniaxial depende de un número de factores tales como el tipo de roca, el grado de cementación, la porosidad, la profundidad. Desde un punto de vista mecánico, es el número de áreas de contacto entre las partículas individuales de la roca. En particular para las calizas, la forma y el esfuerzo de la estructura de la roca determina la deformidad de ésta. La porosidad es uno de los factores que es frecuentemente influenciado por estas condiciones de contacto mecánico, pero ciertamente no es el único. El esfuerzo efectivo también tiene influencia en la deformación de la estructura de la roca.⁽²⁵⁾

Si la presión de sobrecarga permanece constante, entonces se obtiene la fórmula para la compactación del yacimiento:⁽¹⁷⁾

$$\frac{\Delta h}{h} = \frac{1 - \nu - 2\nu^2}{(1 - \nu)E} \cdot \Delta p_f \quad (3.14)$$

donde h es el espesor o altura del yacimiento y Δp_f es la caída de presión.

De esta manera, el coeficiente de compactación uniaxial está definido como

$$C_m = \frac{1 - \nu - 2\nu^2}{(1 - \nu)E} \quad (3.15)$$

por lo tanto, la fórmula de la compactación de un yacimiento se puede definir como

$$\frac{\Delta h}{h} = C_m \cdot \Delta p_f \quad (3.16)$$

A partir de esta fórmula, la compactación Δh puede calcularse siempre que se conozcan las propiedades elásticas E y ν .

Geertsma (1973) ⁽²⁵⁾ presenta una serie de figuras que resumen una gran cantidad de datos experimentales relacionando al coeficiente de compactación uniaxial con la porosidad, tanto para areniscas como para calizas, Figura 3.1

TESIS CON
FALLA DE ORIGEN

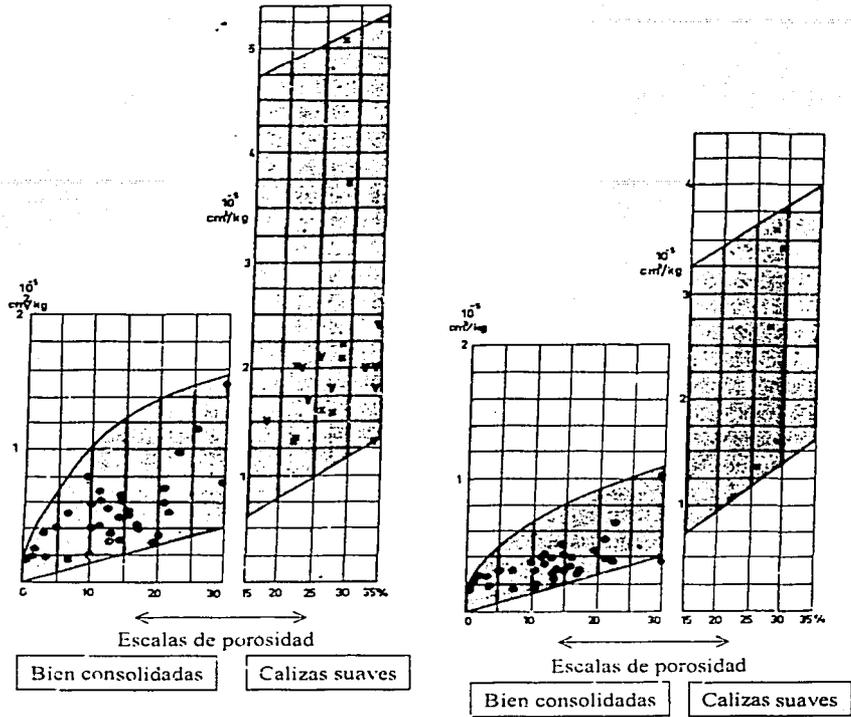


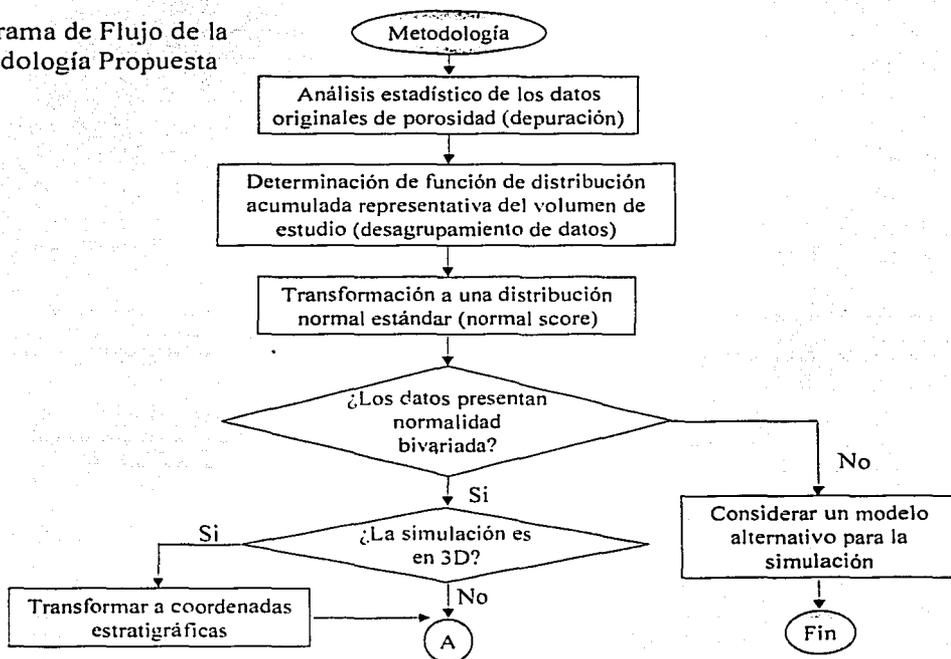
Figura 3.1. Rangos de C_m (eje vertical) para carbonatos. La grafica de la izquierda corresponde a un esfuerzo efectivo vertical, $\sigma_z = 100$ a 200 kg/cm², (profundidad = 1000 m). La gráfica de la derecha corresponde a $\sigma_z = 300$ a 600 kg/cm², (profundidad = 3000 m) ⁽²⁵⁾

TESIS CON
FALLA DE ORIGEN

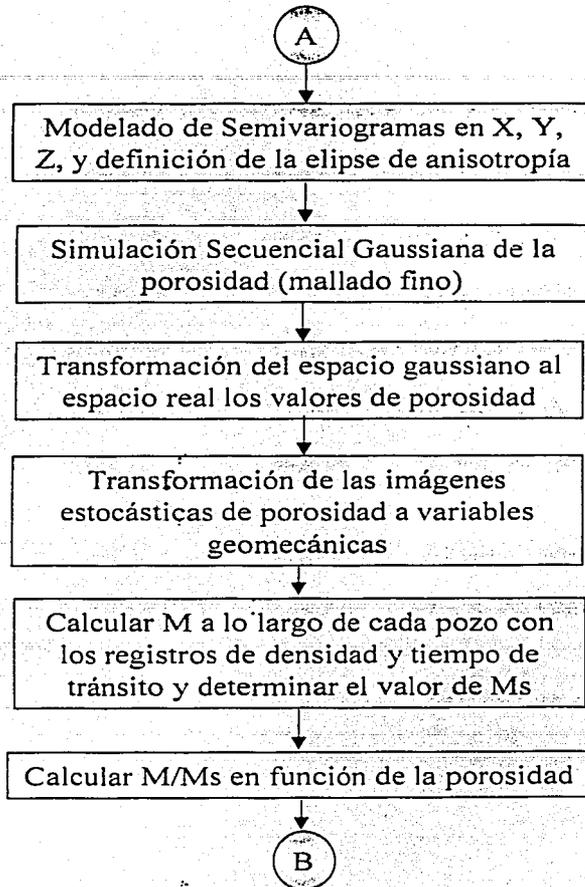
IV. METODOLOGÍA PARA LA CARACTERIZACIÓN GEOMECÁNICA.

Una vez revisado en los capítulos previos los conceptos teóricos necesarios para entender la metodología propuesta, se resumirá en forma de diagrama de flujo la serie de pasos a seguir para obtener las distribuciones espaciales de los módulos de elasticidad, de la resistencia a la compresión uniaxial y del coeficiente de compactación uniaxial en el yacimiento.

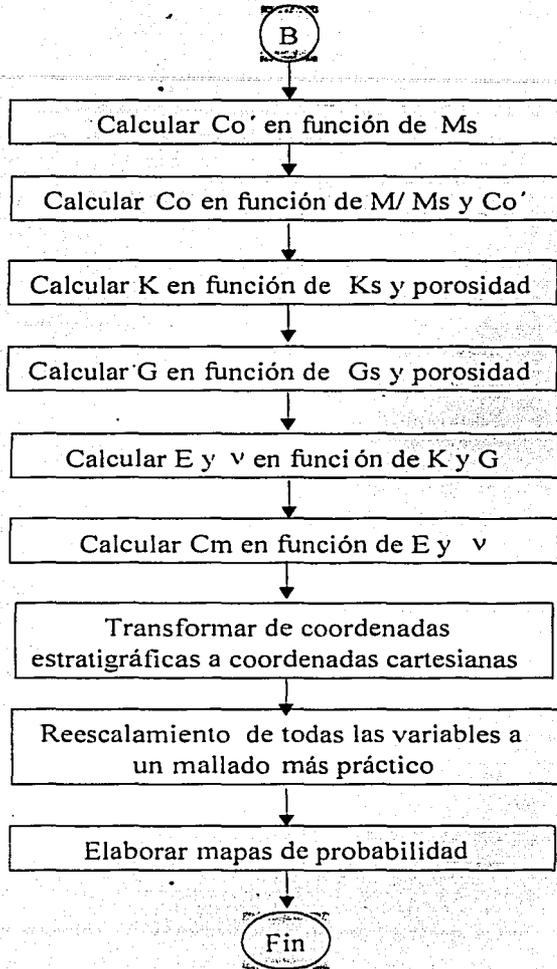
Diagrama de Flujo de la Metodología Propuesta



TESIS CON FALLA DE ORIGEN



TESIS CON
FALLA DE ORIGEN



TESIS C. S. S.
FALLA DE ORIGEN

V. EJEMPLO DE APLICACIÓN Y ANÁLISIS DE RESULTADOS

La metodología propuesta en el capítulo anterior se aplicó a un yacimiento real del cual se disponía de la información de la porosidad en 19 pozos obtenida a partir de registros geofísicos. El área seleccionada para la aplicación de la metodología fue de 50 km², con una longitud de 5,000 m en la dirección E-W y de 10,000 m en la dirección N-S. El espesor estudiado fue de 700 m, iniciando a la profundidad de 2,200m y finalizando a los 2900 m. La distribución de los pozos se muestra en la Figura 5.1.

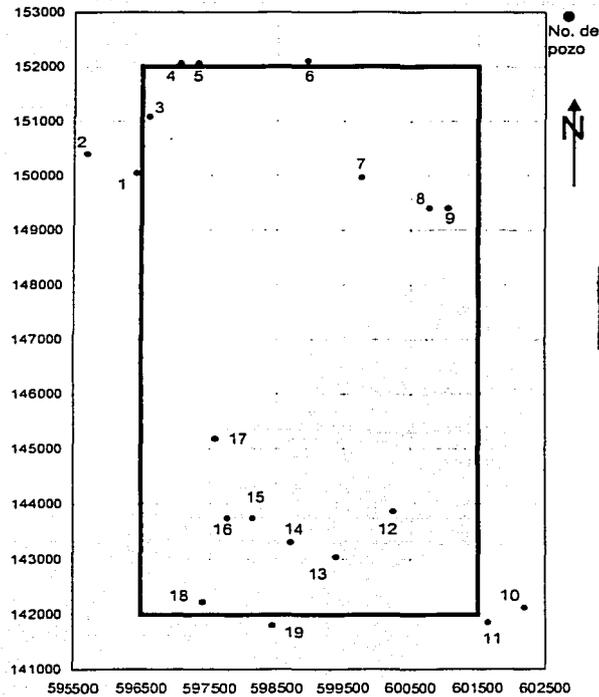


Figura 5.1. Distribución de los 19 pozos en el área de estudio.

Se puede observar que en toda la parte central del yacimiento no se cuenta con información de ningún pozo. Adicionalmente, es importante mencionar que solamente la información de un pozo cubre totalmente los 700 m del espesor analizado, la información de los restantes 18 pozos sólo cubre una porción menor de éste (Figura 5.2).

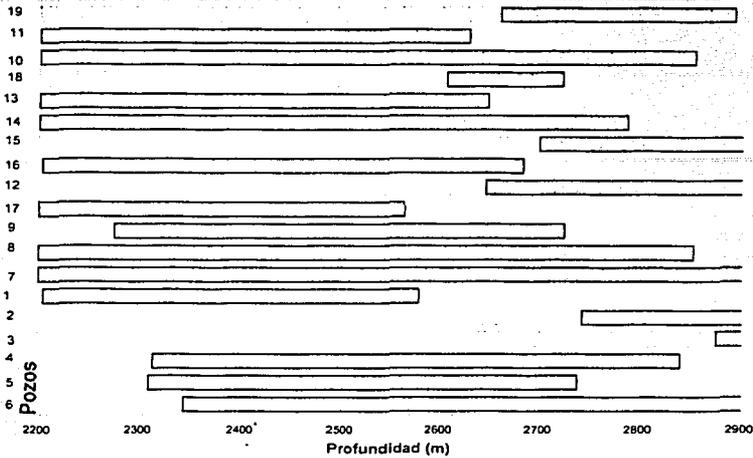


Figura 5.2. Perfiles de pozos mostrando la longitud cubierta por la información disponible.

Como primer paso se procedió a hacer una depuración y el análisis estadístico de la información, obteniéndose los resultados presentados en el histograma de la Figura 5.3. para los 31,684 valores de porosidad totales.

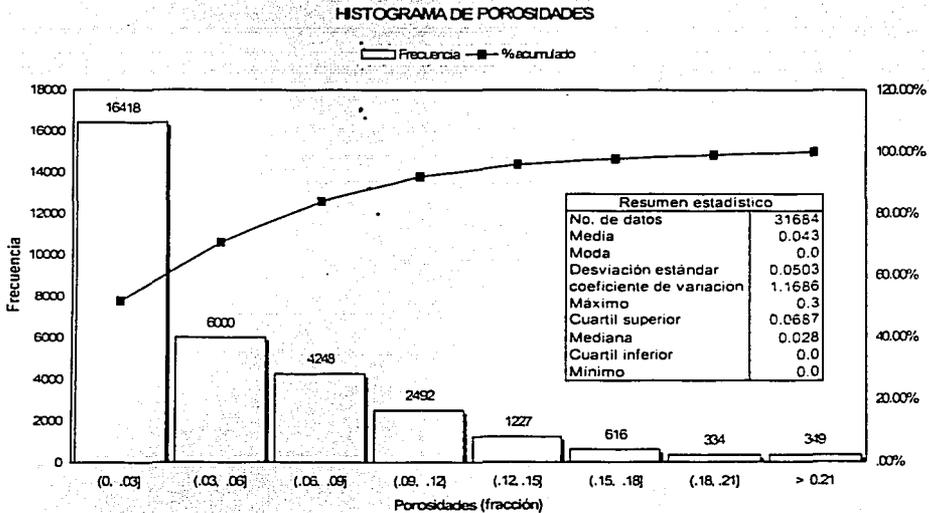


Figura 5.3. Descripción estadística de los valores de porosidad disponibles para el estudio.

Los valores de porosidad obtenidos a partir de los registros geofísicos no fueron calculados en este trabajo. Se consideró que la información disponible era correcta.

El análisis del histograma de la Figura 5.3 demuestra que existe una asimetría positiva y no una distribución de frecuencias aproximadamente simétrica, lo que describiría a una distribución uniforme.

El valor de la media es de 0.043 no obstante que poco más del 50% de los datos tienen un valor de porosidad menor o igual a 0.03. Por esta razón es más representativo de la distribución el valor de la mediana igual a 0.028.

El siguiente paso fue realizar el desagrupamiento de la información, para ello se utilizó el programa **declus** de **GSLIB** ⁽⁷⁾ obteniéndose el histograma de la Figura 5.4.

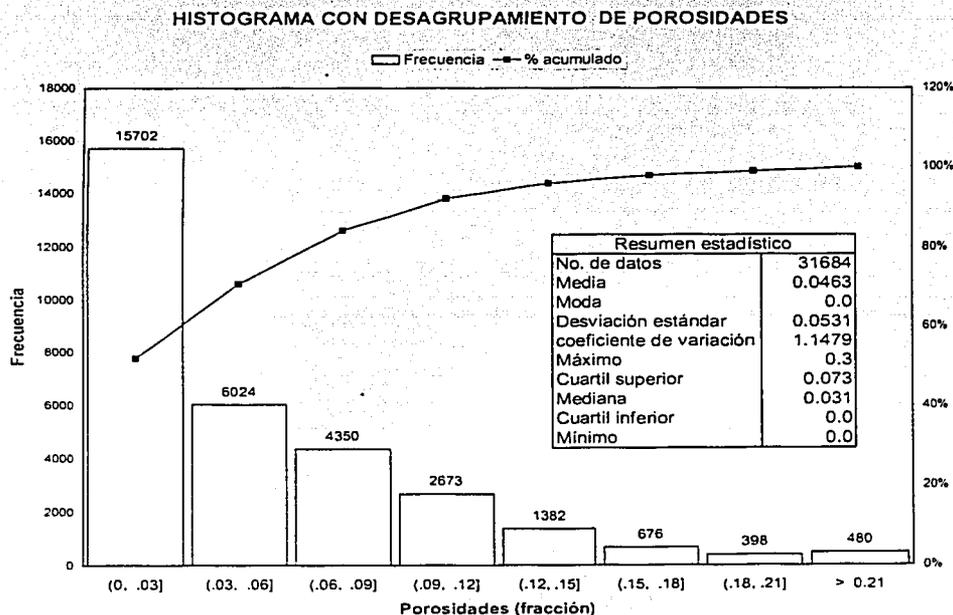


Figura 5.4. Descripción estadística de los valores desagrupados de porosidad.

Los resultados muestran que no se produjo un cambio sustancial en la información original. Esto es explicable debido a la poca cantidad de pozos disponibles. Lo anterior se

reafirma con las gráficas de distribución acumulada tanto para los datos originales como para los desagrupados. Figura 5.5.

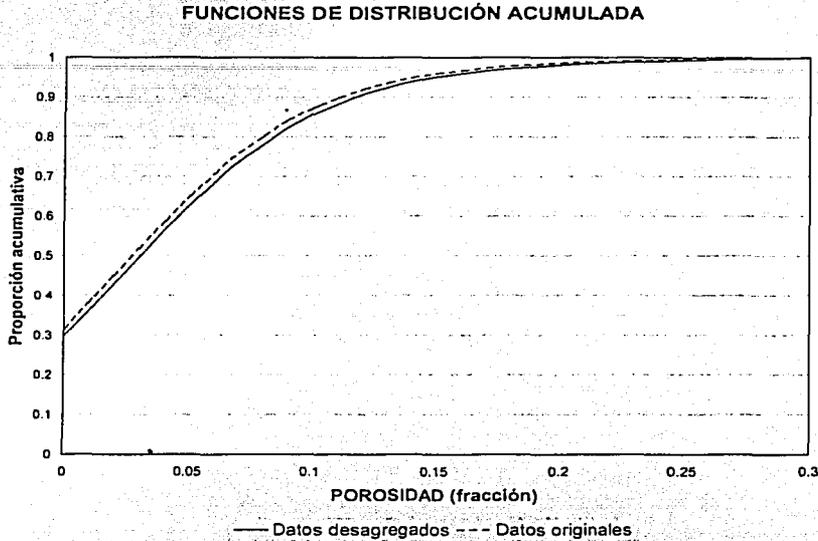


Figura 5.5. Gráficas de las funciones de distribución acumulada tanto para los datos originales de porosidad como los desagrupados.

Una vez realizado el desagrupamiento de los datos, se empleó el programa `nscore` de GSLIB ⁽⁷⁾ para llevar a cabo la transformación de la distribución original a una distribución normal estándar (normal score). El histograma resultante se muestra en la Figura 5.6. El valor de la media igual a cero, y el de la desviación estándar igual a uno, confirman la correcta transformación a una distribución normal estándar.

Los siguientes pasos serían la verificación de que los datos presenten normalidad bivariada para justificar la aplicación de la simulación secuencial gaussiana y, considerando que se trata de una simulación en 3D, se debería hacer la transformación de coordenadas cartesianas a coordenadas estratigráficas; sin embargo, en este trabajo no se realizó la verificación considerando que el comportamiento general de la porosidad en los yacimientos tiende a ser gaussiana, y que probablemente los datos de este ejemplo no lo muestran debido a su escasa distribución areal. Debido también a la falta de información

para establecer la cima y el fondo de cada capa estratigráfica del yacimiento, no se llevó a cabo esta transformación.

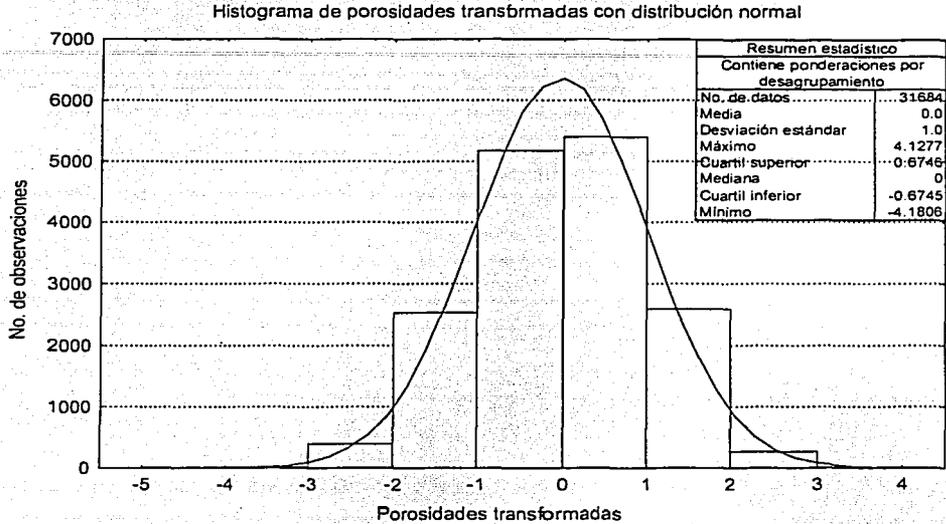


Figura 5.6. Distribución normal estándar (normal score).

El modelado de los variogramas se realizó tanto en las cuatro direcciones principales horizontales así como en la dirección vertical. Para ello se utilizó el programa **gamv** de **GSLIB** ⁽⁷⁾ para datos irregularmente espaciados. Así mismo, para el ajuste de los modelos teóricos a los variogramas experimentales se utilizó el programa **VARIOWIN 2.2**, creado por el Dr. Pannatier.

Las Figuras 5.7 a 5.10 muestran los variogramas en las cuatro direcciones principales: NS, EW, NE –SW 45° y NW – SE 45°, así como los correspondientes modelos ajustados. De su análisis se desprende que existe una anisotropía geométrica con un rango mayor de 1,600 m en la dirección NW – SE 45°, y un rango menor de 800 m en la dirección NE – SW 45°, tal como se aprecia en la elipse de anisotropía de la Figura 5.11.

TESIS CON
FALLA DE ORIGEN

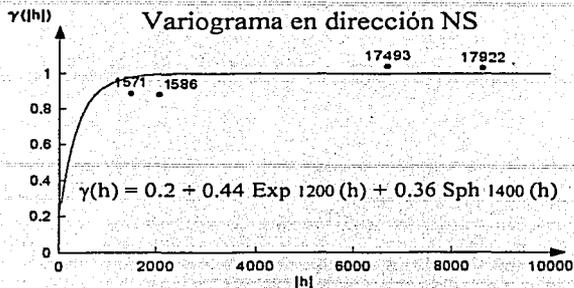


Figura 5.7. Variograma modelado en la dirección NS.

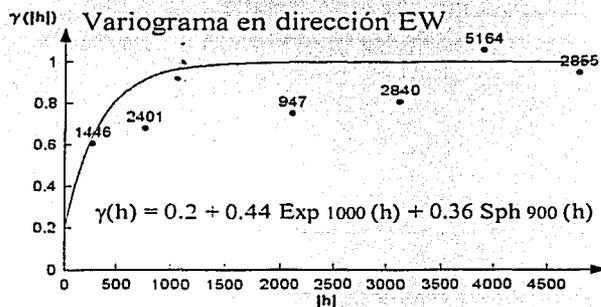


Figura 5.8. Variograma modelado en la dirección EW.

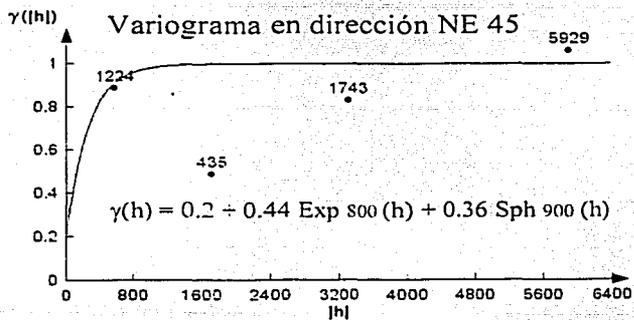


Figura 5.9. Variograma modelado en la dirección NE-SW 45°.

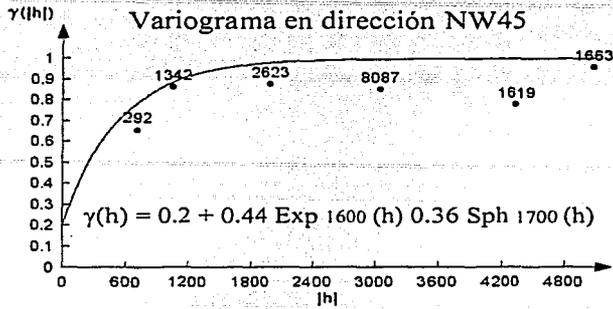


Figura 5.10. Variograma modelado en la dirección NW-SE 45°.

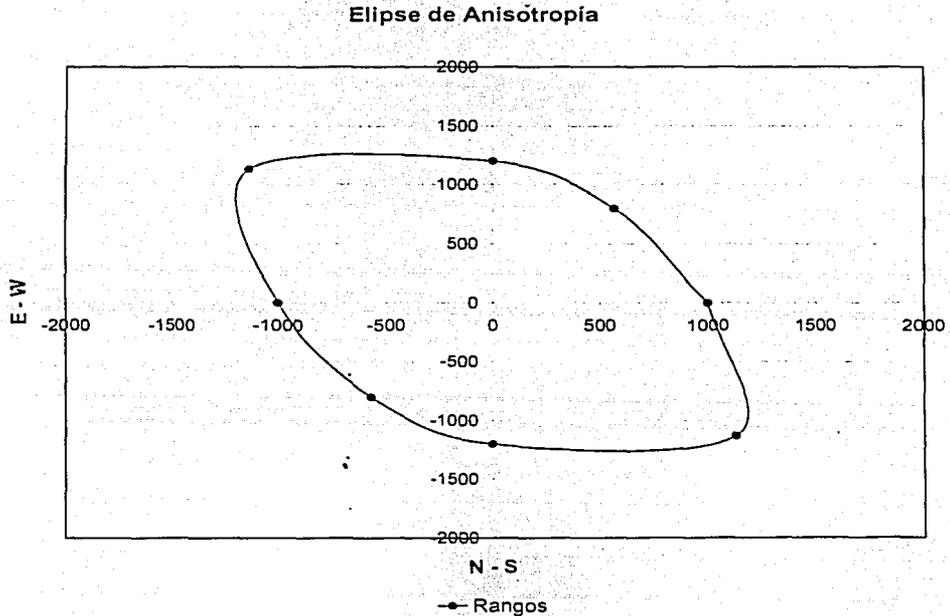


Figura 5.11. Elipse de anisotropía mostrando los rangos máximos y mínimos.

El modelo ajustado obtenido que representa el comportamiento general de la variabilidad en todas las direcciones del plano está dado por:

$$\gamma(h) = 0.2 + 0.44 \text{Exp}(h_1) + 0.36 \text{Sph}(h_2) \quad (5.1)$$

donde

$$h_1 = \begin{bmatrix} h_x \text{Cos}135 + h_y \text{Sen}135 \\ 1600 \\ -h_x \text{Sen}135 + h_y \text{Cos}135 \\ 800 \\ h_z \\ 30 \end{bmatrix}$$

$$h_2 = \begin{bmatrix} h_z \\ 30 \end{bmatrix}$$

El variograma en la dirección vertical así como el variograma omnidireccional, donde es último representa "un promedio" del comportamiento general, se presentan en las Figuras 5.12 y 5.13.

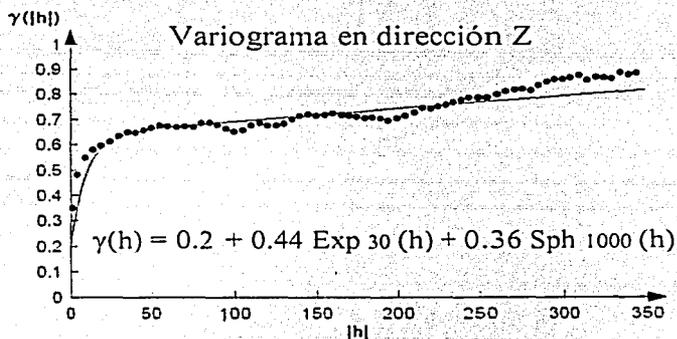


Figura 5.12. Variograma en la dirección vertical.

TESIS CON
FALLA DE ENGEN

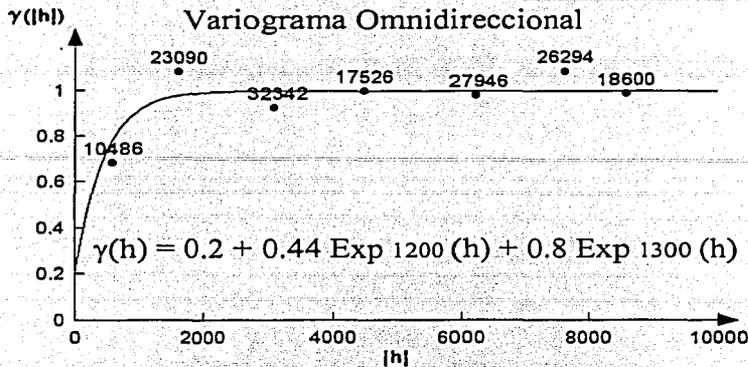


Figura 5.13. Variograma omnidireccional.

Una vez modelados los variogramas, el siguiente paso fue llevar a cabo la simulación secuencial gaussiana con un mallado "fino". El grado de fineza dependerá del volumen que representan las muestras disponibles así como del equipo de cómputo y del tiempo disponibles. Es importante mencionar que los valores de las muestras de porosidad disponibles provienen de registros geofísicos, y no obstante que se dispone de una gran cantidad de valores para cada pozo, en realidad están representando la variabilidad de la porosidad en volúmenes de un pie de largo por un pie de radio aproximadamente, por lo tanto, el tamaño de la malla inicial no debe ser tan grande que altere la variabilidad que representan las muestras. Una vez realizada la simulación, se puede llevar a cabo un reescalamiento a un tamaño de malla más adecuado.

La figura 5.14 muestra los resultados de la imagen de una simulación para una pequeña sección del yacimiento de 500 m por 500 m, con 5 m en la dirección vertical. Para ello se utilizaron los programas de GSLIB⁽⁷⁾ *sgsim*, para la simulación secuencial gaussiana, y *backtr*, para la transformación de los valores del espacio gaussiano al espacio real. La malla está generada por: $n_x = 100$, $dx = 5$, $n_y = 100$, $dy = 5$. Donde: n_x y n_y representan el número de nodos en sus respectivas direcciones, así como, dx y dy , representan el tamaño de los nodos en metros.

La Figura 5.15 muestra la imagen de la simulación de los valores de porosidad de la figura anterior después de aplicarse un reescalamiento promediando aritméticamente los valores. El tamaño la malla es de: $n_x = 20$, $dx = 25$ m, $n_y = 20$, $dy = 25$ m.

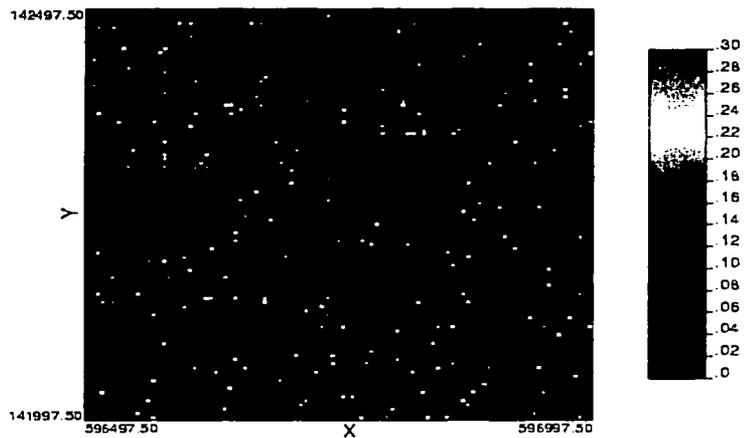


Figura 5.14. Imagen de una simulación de la porosidad con un mallado fino para una sección de 0.25 km^2 del yacimiento y un espesor de 5 m. La escala gráfica indica la porosidad en fracción.

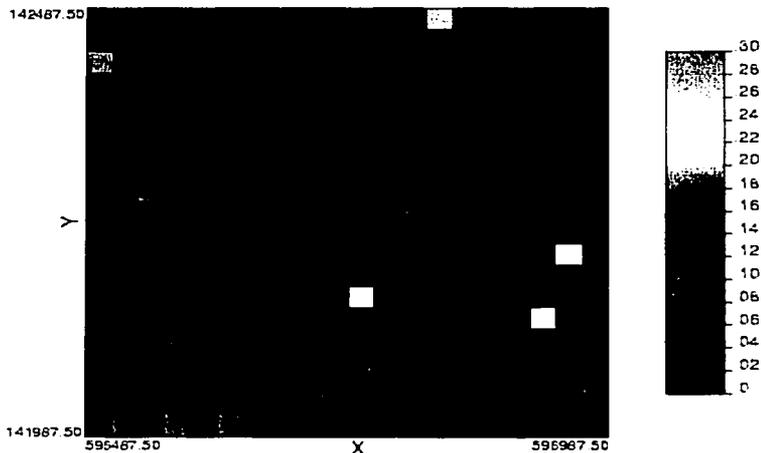


Figura 5.15. Imagen de los resultados del reescalamiento de los valores de la porosidad a un mallado más práctico después de aplicar un promedio aritmético.

Tanto la Figura 5.14 como la Figura 5.15 representan una sola imagen de la sección analizada. Para la realización de este trabajo de tesis se llevó a cabo la simulación de 100 imágenes para cada sección.

Para obtener las imágenes de la porosidad del yacimiento se realizaron 100 simulaciones con un mallado fino a cada 100 m de profundidad. Desde los 2200 m hasta los 2900 m. Debido a que no se realizó el cambio a coordenadas estratigráficas, los espesores considerados en cada sección fueron únicamente de 30 m con el objeto de evitar tomar dos capas estratigráficas diferentes como homogéneas.

Cada una de las imágenes resultantes fueron sometidas a un reescalamiento de tal manera que se formara una malla de: $n_x = 20$, $dx = 250$, $n_y = 40$, $dy = 250$, $n_z = 1$, $dz = 30$.

La presentación de los resultados es un factor que depende de los objetivos del trabajo. Una posibilidad consiste en que las imágenes resultantes muestren las probabilidades de que el atributo analizado rebase cierto valor límite o bien, mostrar las probabilidades de estar bajo cierto umbral. En esta trabajo se consideró que los valores obtenidos para el coeficiente de compactación uniaxial serían utilizados en un simulador, por lo tanto, se buscaban valores representativos que describieran cada zona del yacimiento. El valor más representativo de las 100 imágenes fue su mediana debido a la poca influencia de los valores extremos en su obtención.

Con base en lo anterior se presentan para cada profundidad analizada los valores de las medianas, así como las probabilidades de que las variables se encuentren dentro de un determinado rango a manera de ejemplo.

En este capítulo sólo se presenta una imagen de cada variable. El resto de las imágenes se presentan en el apéndice al final del trabajo.

POROSIDAD

La imagen izquierda de la Figura 5.16 representa los valores de la mediana de las 100 imágenes simuladas de porosidad a la profundidad de 2500 m en el área de estudio de 50 km² con un espesor de 30 m. La imagen a la derecha representa las probabilidades de que la porosidad en esta parte del yacimiento sea menor o igual a 0.06. Las cruces representan la ubicación de los pozos.

TESIS CON
FALLA DE ORIGEN

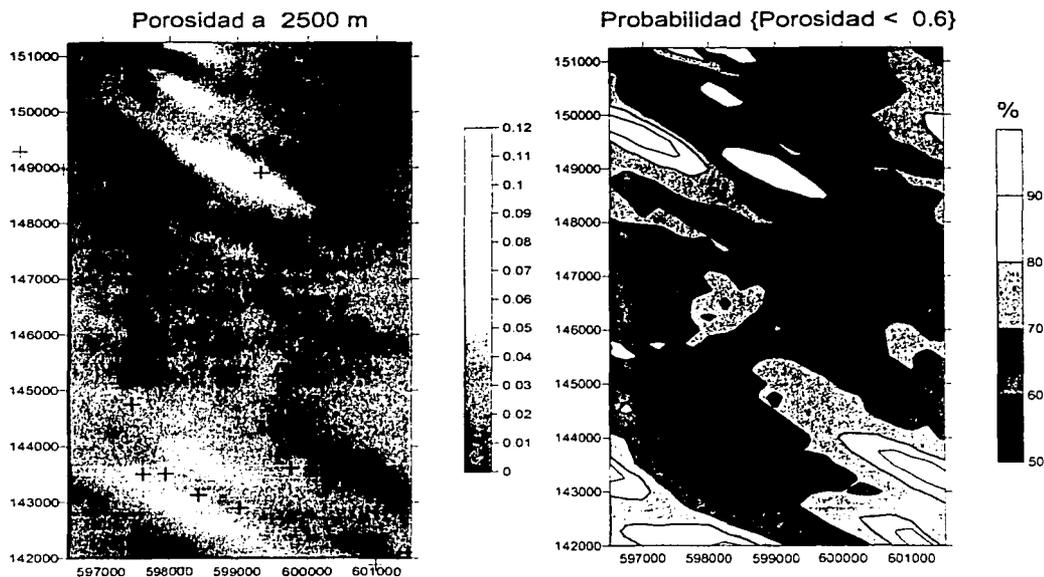


Figura 5.16. Imágenes de la porosidad a la profundidad de 2500 m. La imagen izquierda representa la mediana de 100 simulaciones realizadas. La imagen derecha muestra las probabilidades de que la porosidad se encuentre por debajo de 0.06.

Los resultados muestran que la porosidad en la mayor parte de esta sección del yacimiento (2500 m) se encuentra entre 0.02 y 0.06. Si se revisa el histograma de los datos en la Figura 5.4 se puede apreciar que poco más del 70% de los valores se encuentran por debajo de 0.06. Ahora bien, considerando que al obtener la mediana los valores extremos desaparecen, los resultados obtenidos respetan las características globales de los datos.

Otra característica importante que se ve refleja en los resultados obtenidos es la presencia de la anisotropía en la distribución de la porosidad. Durante el modelado de los variogramas se encontró que existía una mayor continuidad de la porosidad en la dirección NW-SE 45°. Esto es precisamente lo que se muestra en la figura 5.16. Este comportamiento global se ve reflejado por la implementación de la simulación estocástica. Si se hubiera realizado únicamente una estimación determinística, probablemente no se vería reflejada esta distribución.

TESIS CON
FALLA DE ORIGEN

Se puede observar también que en la parte central del área analizada no se presenta una gran variabilidad. Esto se puede deber a que no se cuenta con ningún pozo ubicado en esa zona por lo que los valores asignados sólo mantienen la estadística general.

Con respecto a las probabilidades se aprecia que a pesar de que los valores obtenidos de las medianas se encuentran en el rango de 0.02 a 0.06 para la mayor parte del área, su probabilidad de ocurrencia está entre 60 y 70 por ciento, no obstante que el rango analizado abarca los valores con mayor frecuencia en los datos. Esto se puede deber a la existencia de una gran variabilidad en los valores generados en los nodos que no tienen pozos cercanos para su estimación, los cuales son la gran mayoría.

Revisando el conjunto de imágenes de la porosidad desde la profundidad de 2,200 m a 2,900 m se pueden identificar siete zonas principales en la porción del yacimiento estudiado. No obstante que el rango de variación de los valores de la porosidad es pequeño, del orden de 0.01 a 0.1, éstos tienen una gran influencia en el comportamiento de las demás variables, las cuales, presentan variaciones fuertes para valores pequeños de ésta, tal como se apreciará en sus distribuciones obtenidas.

La zona noroeste presenta porosidades del orden de 0.02 a 0.03, con excepción de la profundidad de 2,300 m donde se muestra un incremento a porosidades del orden de 0.05. La zona norte muestra los valores de porosidades más elevadas de todo el yacimiento. Estos varían de 0.7 a 0.1, con incrementos y decrementos a lo largo del intervalo total. La zona noreste por su parte, está caracterizada por valores de porosidad en el rango de 0.04 a 0.06, sin embargo a la profundidad de 2,700 m se presenta una disminución de ésta a valores de 0.03 en promedio.

Con respecto a la parte central del yacimiento se mantiene un comportamiento homogéneo con poca variabilidad de la porosidad. Los primeros 200 metros presentan valores en el rango de 0.03 a 0.04, el intervalo restante indica porosidades del orden de 0.04 a 0.05. Esta zona no presenta cambios abruptos tales como los encontrados en las zonas restantes, esto es lógico ya que no hay información en los datos que la describa.

La zona suroeste está dividida en dos partes principales. Hasta la profundidad de 2,500 m se presentan valores de porosidad bajos, alrededor de 0.02, sin embargo a partir de esta profundidad los valores aumentan a un rango de entre 0.05 y 0.07. La zona sur por su parte muestra un comportamiento más irregular. Los primeros 200 metros tienen los valores más elevados del conjunto de resultados de la porosidad, de 0.11 a 0.12; los

siguientes 400 m presentan valores en el rango de 0.05 a 0.06, finalizando el intervalo con valores pequeños de entre 0.02 a 0.04.

Finalmente la zona sureste con porosidades del orden de 0.02 a 0.03 en la mitad superior, con excepción de la profundidad de 2,300 m, donde es de 0.06. La mitad inferior presenta porosidades que variación del 0.04 al 0.08.

El hecho de que no se encuentre una variación uniforme de la porosidad al aumentar la profundidad, indica el alto grado heterogeneidad de las formaciones del yacimiento.

RESISTENCIA A LA COMPRESIÓN UNIAxIAL, Co.

Para la determinación de Co, primeramente se requirió obtener el valor del módulo M a partir de los registros de tiempo de tránsito y de densidad en cada pozo. Su valor se graficó vs la porosidad, de esta manera se encontró el valor del módulo M del sólido, Ms, el cual es igual al valor de M cuando la porosidad es igual a cero, Figura 5.17. El valor promedio obtenido fue de 110 GPa.

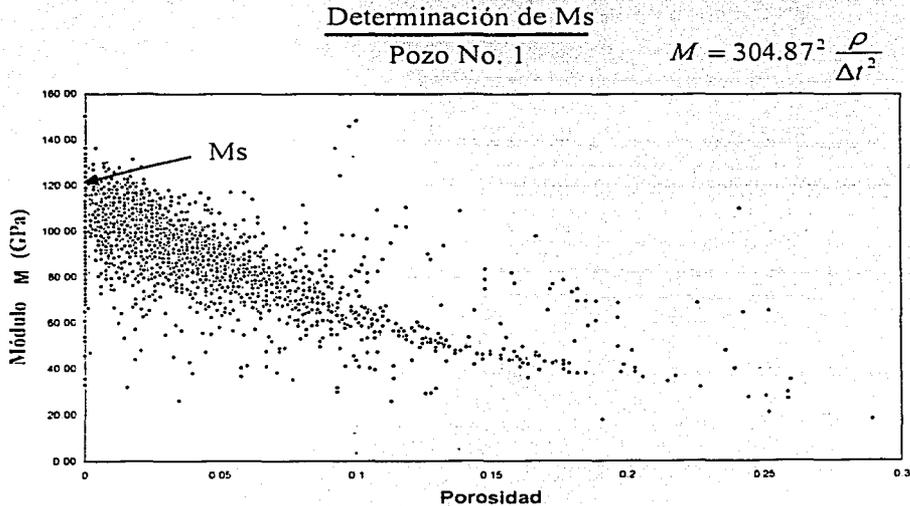
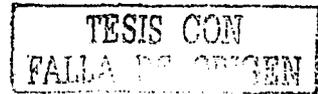


Figura 5.17. Determinación del módulo Ms a partir del registro de densidad y del tiempo de tránsito.



Mediante la aplicación de las ecuaciones 3.7 a 3.10 se obtuvieron los resultados de la distribución de C_o en el yacimiento. En la figura 5.18 se puede apreciar que el comportamiento de su distribución a lo largo del área es muy similar al de la porosidad a la misma profundidad, con la particularidad de que los valores grandes de porosidad corresponden a las zonas de menor resistencia a la compresión uniaxial. Esto se debe a que su valor se ve afectado únicamente por la porosidad en las ecuaciones utilizadas en este trabajo para su cálculo, ya que se utilizó como constante el valor de M_s obtenido anteriormente. Por lo tanto, los comentarios hechos anteriormente para la porosidad también son válidos para el coeficiente de compactación uniaxial.

No obstante que a esta profundidad la porosidad presenta una variación muy pequeña (de 0.02 a 0.08), la variación de la C_o es más considerable, de 170 a 290 MPa.

Los resultados muestran que la resistencia a la sobrecarga no aumenta necesariamente con la profundidad, como podría esperarse en una tendencia normal.

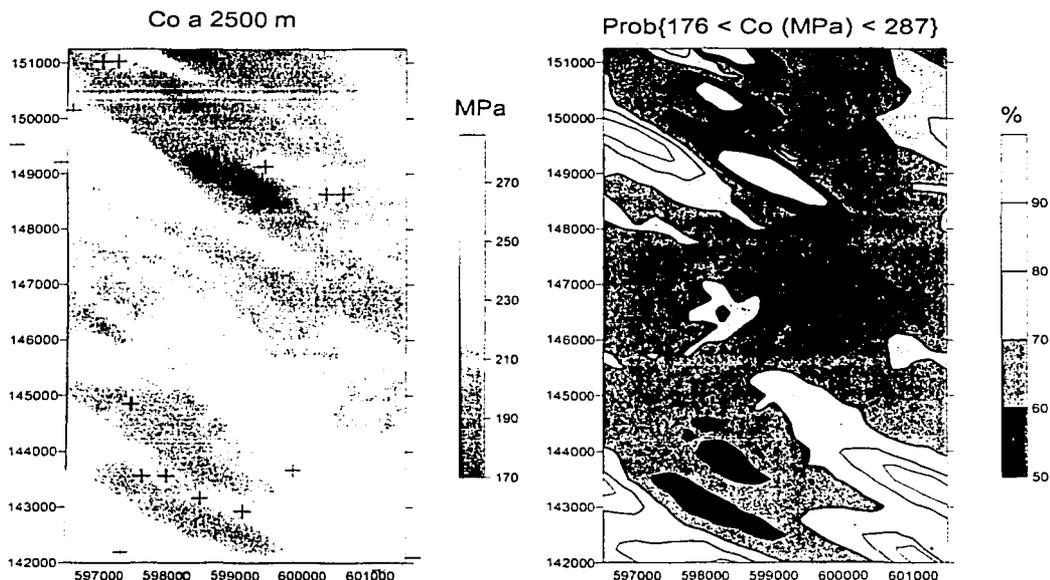


Figura 5.18. Imágenes de C_o a la profundidad de 2500 m. La imagen izquierda representa la mediana de 100 simulaciones realizadas. La imagen derecha muestra las probabilidades de que C_o sea mayor a 176 MPa y menor a 287 MPa.

VARIABLES ELASTO-DINÁMICAS

Los módulos K y G fueron calculados mediante las ecuaciones 3.11 y 3.12 en función de la porosidad. La distribución encontrada muestra una disminución en sus valores con el incremento de la porosidad (ver apéndice). Los valores de K varían en el rango de 47 a 62 GPa con un valor promedio de 54 GPa. Por su parte el módulo G se encuentra en el rango de 23 a 31 GPa con un promedio de 27 GPa. En las figuras 5.19 y 5.20 se presenta su distribución a la profundidad de 2,600 m y 2,700 m respectivamente.

No obstante que el rango de probabilidades estipulado para ambas variables es muy grande, la certidumbre de la estimación es del 70% para la mayor parte del área analizada, esto confirma que si se redujera la búsqueda a valores dentro de un intervalo pequeño, las probabilidades serían menores al 50%.

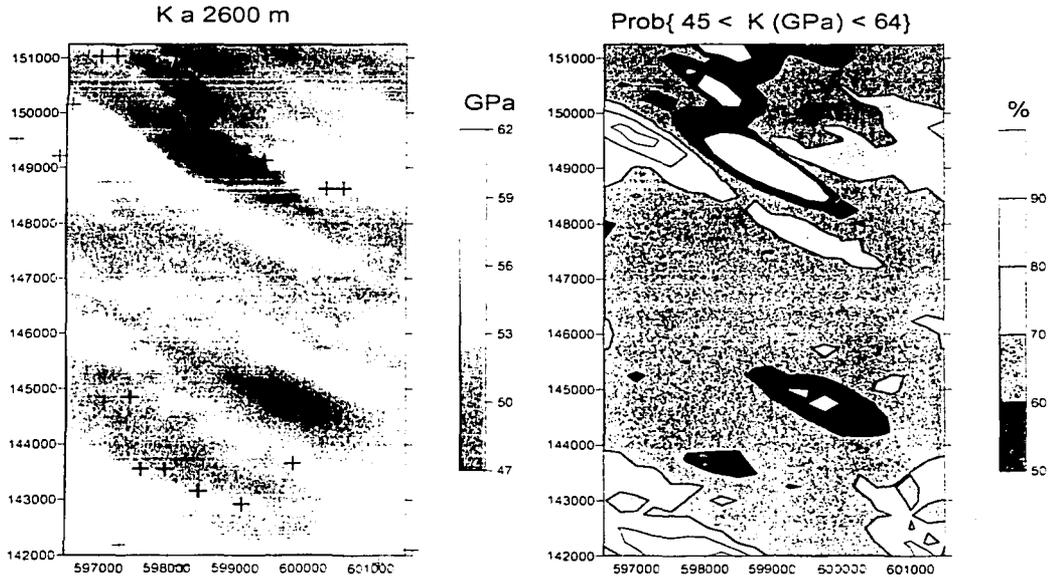


Figura 5.19. Imágenes de K a la profundidad de 2,600 m. La imagen izquierda representa la mediana de 100 simulaciones realizadas. La imagen derecha muestra las probabilidades de que K sea mayor a 45 GPa y menor a 64 GPa.

TESIS CON
FALLA DE ORIGEN

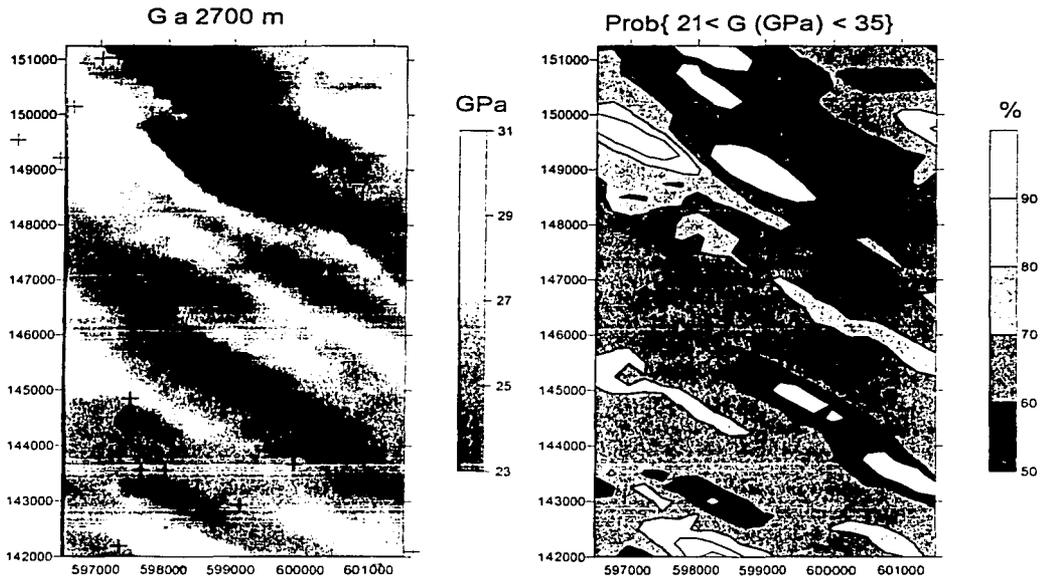


Figura 5.20. Imágenes de G a la profundidad de 2700 m. La imagen izquierda representa la mediana de 100 simulaciones realizadas. La imagen derecha muestra las probabilidades de que G sea mayor a 21 GPa y menor a 35 GPa.

Los resultados del módulo de Young, E , y la relación de Poisson, ν , no fueron obtenidos directamente de la porosidad sino mediante las ecuaciones 3.5 y 3.6, las cuales están en función únicamente de K y G .

Los valores del módulo de Young se encuentran en el intervalo de 60 a 80 GPa con un valor medio de 68 GPa, figura 5.21.

TESIS CON
FALLA DE ORIGEN

ESTA TESIS NO SALE
DE LA BIBLIOTECA

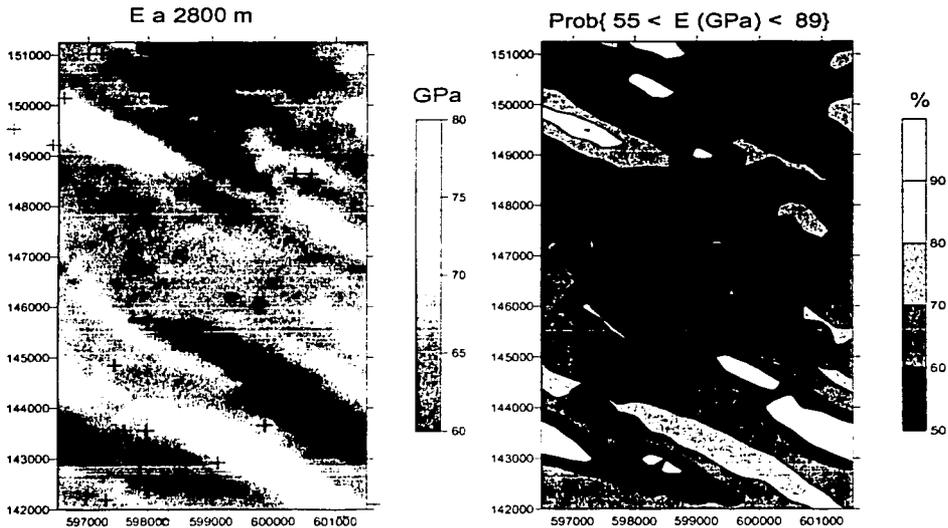


Figura 5.21. Imágenes de E a la profundidad de 2800 m. La imagen izquierda representa la mediana de 100 simulaciones realizadas. La imagen derecha muestra las probabilidades de que E sea mayor a 55 GPa y menor a 89 GPa.

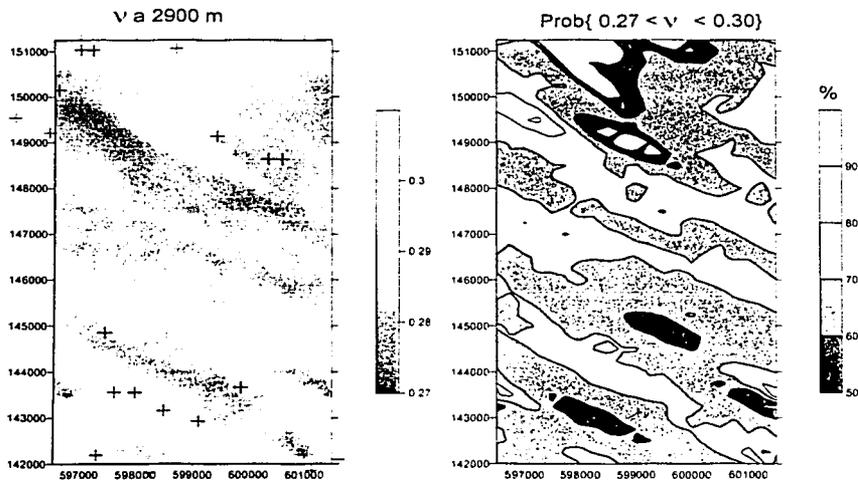


Figura 5.22. Imágenes de ν a la profundidad de 2900 m. La imagen izquierda representa la mediana de 100 simulaciones realizadas. La imagen derecha muestra las probabilidades de que ν sea mayor a 0.27 y menor o igual a 0.30.

La distribución del coeficiente de compactación uniaxial, C_m , a la profundidad de 2500 m obtenida mediante la ecuación 3.15 se presenta en la figura 5.23.

Al comparar los valores obtenidos para el coeficiente de compactación uniaxial con la gráfica propuesta por Geertsma obtenida mediante datos experimentales (figura 3.1 derecha), se observa que éstos están dentro del rango esperado para porosidades de 0 a 5 % y a una profundidad de 3000 m.

Normalmente para llevar a cabo la simulación de un yacimiento se utiliza un valor promedio de C_m en toda el área o volumen analizado. Los resultados obtenidos muestran que existe una gran variabilidad tanto horizontal como verticalmente (ver apéndice). Ello implica que el considerar que todas las rocas del yacimiento se compactan de igual manera puede dar origen a simulaciones poco representativas de la realidad.

Finalmente, considerando una caída de presión de 166 kg/cm^2 , el espesor comprendido entre las profundidades de 2200 a 2900 metros y los resultados del coeficiente de compactación uniaxial, se estimó la compactación del yacimiento con los resultados mostrados en la figura 5.24.

Este tipo de gráfica se seleccionó para apreciar más fácilmente los contrastes entre las zonas de mayor compactación con las menos compactadas.

TESIS CON
FALLA DE CALIFICACION

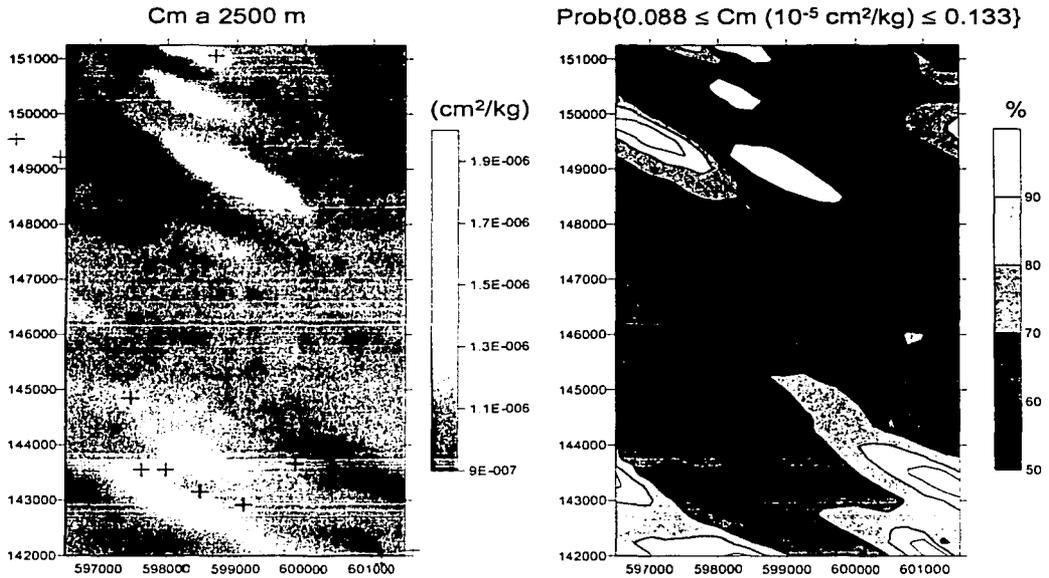


Figura 5.23. Imágenes de Cm a la profundidad de 2500 m. La imagen izquierda representa la mediana de 100 simulaciones realizadas. La imagen derecha muestra las probabilidades de que Cm sea mayor a 0.088 y menor a 0.133 ($10^{-5} \cdot \text{cm}^2/\text{kg}$)

TESIS CON
FALLA DE ORIGEN

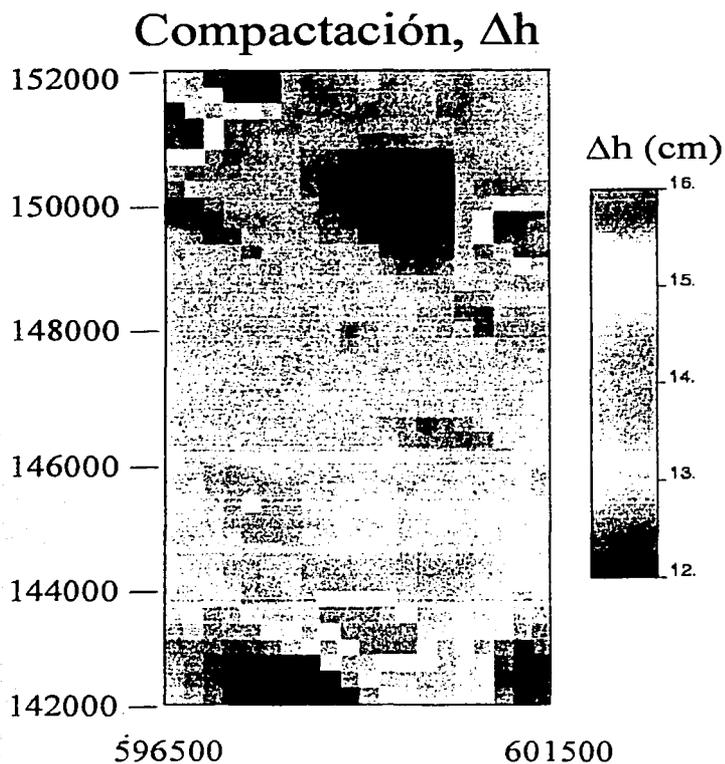
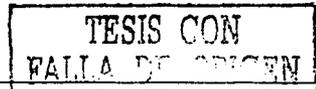


Figura 5.24. Imagen de la compactación del yacimiento para una caída de presión de 166 kg/cm^2 , un espesor comprendido entre las profundidades de 2200 a 2900 metros y los resultados del coeficiente de compactación uniaxial.

TESIS CON
FALLA DE ORIGEN

CONCLUSIONES Y RECOMENDACIONES

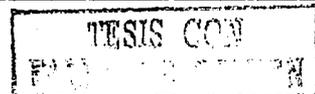
- Se integró un procedimiento para la caracterización geomecánica de las formaciones a partir de la simulación secuencial gaussiana de la porosidad y correlaciones empíricas de las propiedades poroelásticas: módulo volumétrico, módulo de corte, módulo de Young y relación de Poisson.
- El empleo de los métodos geoestadísticos en la distribución espacial de la porosidad para el estudio del campo analizado, concuerda con la dirección preferencial NW 45° SE de la porosidad obtenido en otros estudios para el mismo campo.
- La distribución espacial de propiedades poro-elásticas y de resistencia de las rocas está disponible, al detalle necesario, para realizar la simulación del comportamiento poro-elástico de las formaciones y así, el estudio y análisis asociado con este comportamiento, tal como la compactación y subsidencia, producción de arena e inestabilidad mecánica de formaciones.
- Se recomienda el empleo del método propuesto para estimar la distribución de compresibilidades de las rocas en forma espacial para su aplicación en los simuladores de flujo de fluidos en yacimientos.
- Es recomendable que el analista complemente su estudio con un conocimiento geológico-geofísico del área para desarrollar los variogramas requeridos en la simulación secuencial gaussiana.
- Los resultados de las probabilidades obtenidas en la simulación de las propiedades analizadas proporcionan el grado de confiabilidad del estudio, ésta es la aportación valiosa que brinda la Geoestadística en el modelado espacial.
- La incorporación de la Geoestadística en la caracterización de los yacimientos permite tener una visión más real de éste. Se visualizan más claramente las heterogeneidades dando al yacimiento su verdadera dimensión en complejidad.



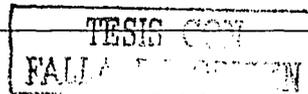
BIBLIOGRAFÍA

LIBROS:

- 1) Farquhar, R. A., Somerville, J. M., y Smart, B. G.: "Porosity as a Geomechanical Indicator: An Application of Core and Log Data and Rock Mechanics", SPE 28853, Octubre, 1994.
- 2) Srivastava, R. M. An Overview of Stochastic Methods for Reservoir Characterization. En Yarus, J. M. y Chambers, R. L., editores, Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies. AAPG, U.S.A., 1994.
- 3) Jensen, L. J., et al., Statistics for Petroleum Engineers and Geoscientists, Prentice Hall PTR, U.S.A., 1997.
- 4) Gutiérrez, E. Gustavo, Determinación de la resistencia a la compresión uniaxial a partir de registros. Tesis de Maestría, UNAM, Facultad de Ingeniería, División de Estudios de Posgrado, 2001.
- 5) Isaaks, Edward y Srivastava, Mohan, An Introduction to Applied Geostatistics, Oxford University Press, U.S.A., 1989.
- 6) Canavos, George, Probabilidad y Estadística. Aplicaciones y Métodos, McGraw-Hill, 1ª edición, México, 1988.
- 7) Deutsch, Clayton y Journel, André, GSLIB: Geostatistical Software Library and Users Guide, 2ª edición, Oxford University Press, Inc., U.S.A., 1992.
- 8) Journel, A. G., y Huijbregts, Ch. J., Mining Geostatistics, Academic Press, Gran Bretaña, 1991.
- 9) Journel, Andre, Fundamentals of Geostatistics in Five Lessons, Short Course Presented at the 28th International Geological Congress Washington, D.C., Vol. 8, American Geophysical Union.
- 10) Wolf, D. J., Withers, K. D. y Burnaman, M. D. Integration of Well and Seismic Data Using Geostatistics. En Yarus, J. M. y Chambers, R. L., editores, Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies. AAPG, U.S.A., 1994.
- 11) Hohn, M. Edward, Geostatistics and Petroleum Geology, Kluwer Academic Publishers, 2ª ed., Gran Brataña 1999.



- 12) Olea, Ricardo. Fundamentals of Semivariogram Estimation, Modeling, and Usage. En Yarus, J. M. y Chambers, R. L., editores, Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies. AAPG, U.S.A., 1994.
- 13) Mansoori, John. A Review of Basic Upscaling Procedures: Advantages and Disadvantages. En Yarus, J. M. y Chambers, R. L., editores, Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies. AAPG, U.S.A., 1994.
- 14) Chu, J, Xu, W y Journel, A. 3-D Implementation of Geostatistical Analyses – The Amoco Case Study. En Yarus, J. M. y Chambers, R. L., editores, Stochastic Modeling and Geostatistics: Principles, Methods, and Case Studies. AAPG, U.S.A., 1994.
- 15) Fairhurst, C., Rock Mechanics. Pergamon Press, Londres, 1963.
- 16) Gnirk, P. F. "The Mechanical behavior of uncased wellbores situated en elastic/plastic media under hydrostatic stress." Trans. AIME, Vol. 253, 1972.
- 17) Erling, F, et al. Petroleum Related Rock Mechanics, Elsevier Science Publishing Company, Países Bajos, 1992.
- 18) Tiab, Djebbar y Erle C. Donaldson. Petrophysics. Gulf Publishing Company. U.S.A., 1996.
- 19) Wyllie, M., Gregory, A. y Gardner, L. "Elastic Wave Velocities in Heterogeneous and Porous Media", Geophysics, 21, 1956.
- 20) Wyllie, M., Gregory, A. y Gardner, L. "An experimental Investigation of Factors Affecting Elastic Wave Velocities in Porous Media", Geophysics, 23, 1958.
- 21) Vernik, L., Bruno, M. y Bovberg, C., "Empirical Relations Between Compressive Strength and Porosity of Siliclastic Rocks", Int. J. Rock Mech. Min. Sci & Geomech. Abstr., 30 (7), 1993.
- 22) Allison, R. J., "Non-Destructive Determination of Young's Modulus and Its Relationship With Compressive Strength, Porosity and Density", en "Deformation of Sediments and Sedimentary Rocks", Jones, M. E. and Preston, R. (Eds.), Geological Society Special Publication No. 29, 1987.
- 23) Kamel, M., et al, "Porosity-Mechanical Properties Crossplots: Possible Indicators of Fracture Intensity", SPWLA 32ª Annual Logging Symposium, junio. 1991.



- 24) Edimann, K., et al: "Predicting Rock Mechanical Properties from Wireline Porosities", SPE/ISRM 47344, Julio, 1998.
- 25) Geertsma, J., "Land Subsidence Above Compacting Oil and Gas Reservoirs", JPT, junio, 1973.

TESIS CON
FALLA DE ORIGEN

NOMENCLATURA

a	Rango
$C(X, Y)$	Covarianza entre las variables X e Y
C_0	Efecto pepita
$C_1 = Cov(0)$	Umbral
C_m	Coefficiente de compactación uniaxial.
C_0	Resistencia a la compresión uniaxial
C_0'	Resistencia a la compresión uniaxial del sólido de la roca
CV	Coefficiente de variación
dx	Tamaño de los nodos en la dirección x
dy	Tamaño de los nodos en la dirección y
E	Módulo de Young
$E[Z] = m$	Valor esperado de una variable aleatoria Z
F/A	Carga por unidad de área
$F(z) = P(Z \leq z)$	Función de distribución acumulada una variable aleatoria Z .
$F(x, y) = \text{Prob}(X \leq x, Y \leq y)$	Función de distribución acumulada bivariada
$f(z)$	Función de densidad de probabilidad de una variable aleatoria continua Z .
G	Módulo de corte
G_s	Módulo de corte del sólido de la roca
h	Distancia de separación entre pares de datos
K	Módulo volumétrico
K_s	Módulo volumétrico del sólido de la roca
M	Módulo M
M_s	Módulo M del sólido de la roca
n	Número de datos en una muestra
n_x	Número de nodos en la malla en la dirección x
n_y	Número de nodos en la malla en la dirección y
$n(h)$	Número de pares de datos separados por una distancia h
$P(Z = z)$	Función de probabilidad de una variable aleatoria discreta Z .
s	Desviación estándar de una muestra de datos
s^2	Varianza de una muestra de datos
u	Vector de coordenadas de la posición de un atributo z .

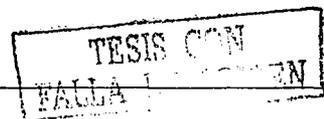
$VAR[Z]$	Varianza de una variable aleatoria Z
\bar{x}	Media de una muestra de datos
Z	Variable aleatoria. Pueden ser discreta o continua.
z	Representa un valor particular de una variable aleatoria Z
Z(x)	Función aleatoria
Z_R	Variable regionalizada
Z^*	Estimado del atributo Z
ϕ	Porosidad.
μ	Parámetro de Lagrange
ρ	Densidad de la roca
ρ	Coficiente de correlación
ν	Relación de Poisson
$\gamma(X, Y)$	Variograma de las variables aleatorias X e Y
ΔD	Cambio de diámetro de una muestra de roca
Δh	Compactación del yacimiento
ΔL	Cambio en longitud de una muestra de roca
Δp_i	Caída de presión en el yacimiento
Δt	Tiempo de tránsito

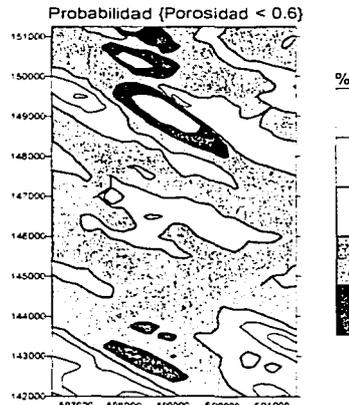
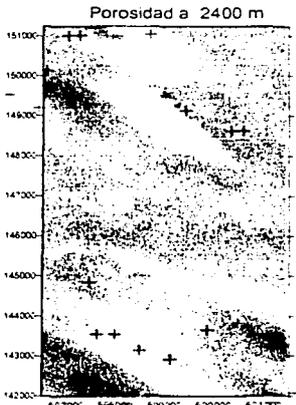
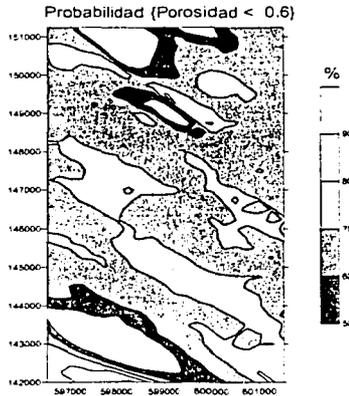
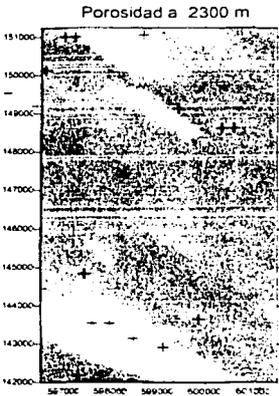
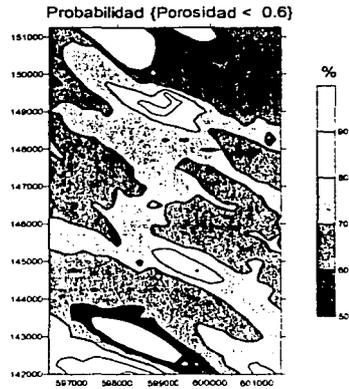
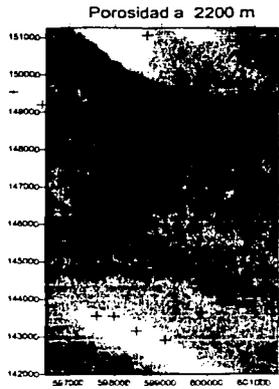
APÉNDICE

La serie de imágenes que se presentan a continuación son los resultados obtenidos mediante la simulación secuencial gaussiana. Las variables analizadas son: porosidad, resistencia a la compresión uniaxial, C_o ; módulo de Young, E ; módulo de corte, G ; módulo volumétrico, K ; relación de Poisson, ν y coeficiente de compactación uniaxial, C_m .

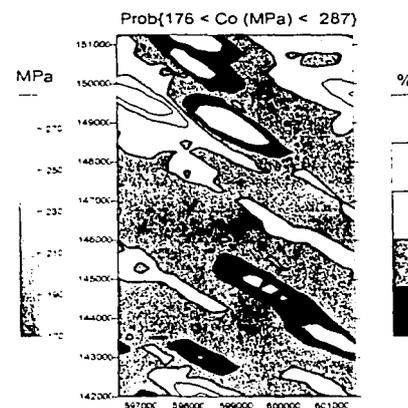
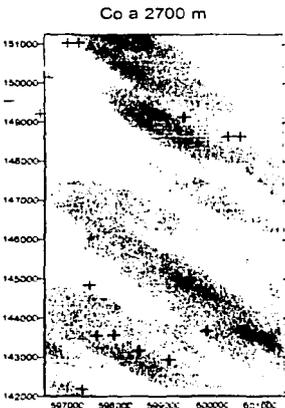
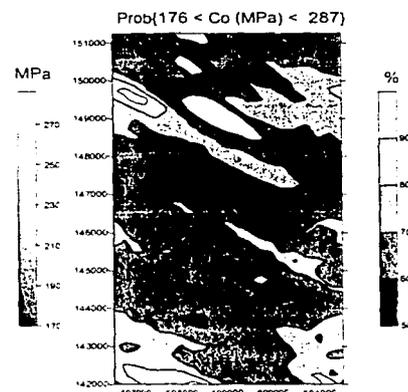
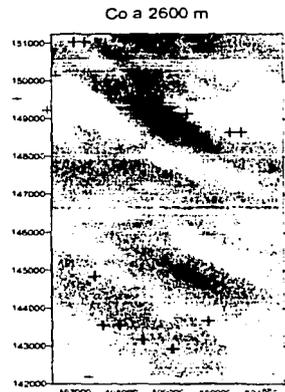
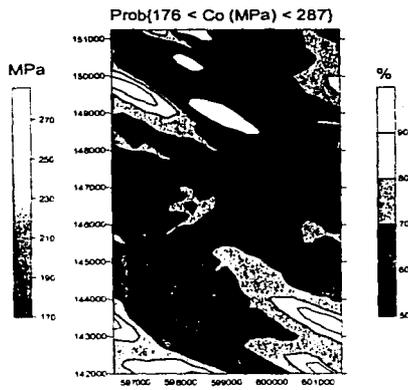
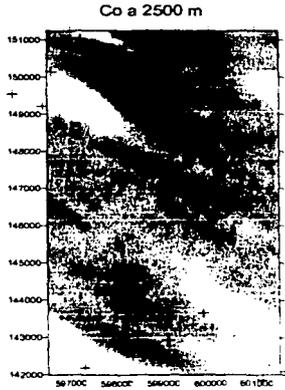
Las imágenes de la izquierda en cada par de figuras representan la distribución del valor de la mediana resultante de 100 simulaciones para cada variable. Las cruces negras representan la ubicación de los pozos disponibles como dato.

Las imágenes de la derecha en cada par de figuras representan las probabilidades de que la variable analizada se encuentre dentro del rango indicado en la parte superior de cada imagen, esto como resultado de las 100 simulaciones realizadas. Las zonas blancas representan probabilidades menores al 50%.

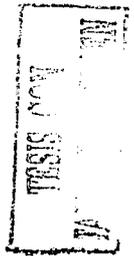
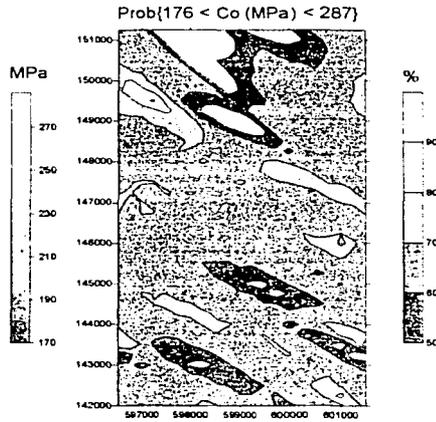
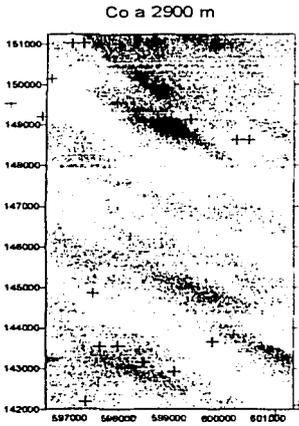
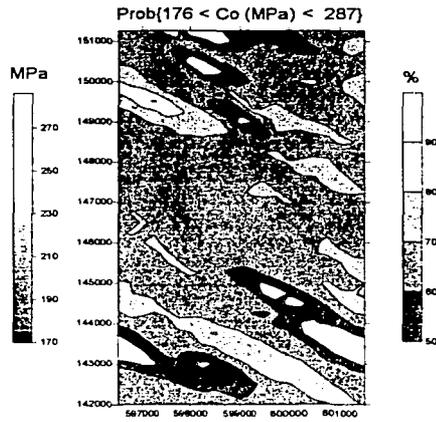
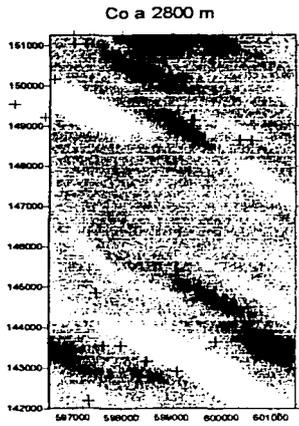


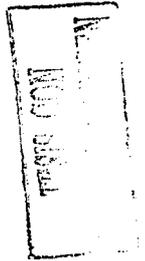
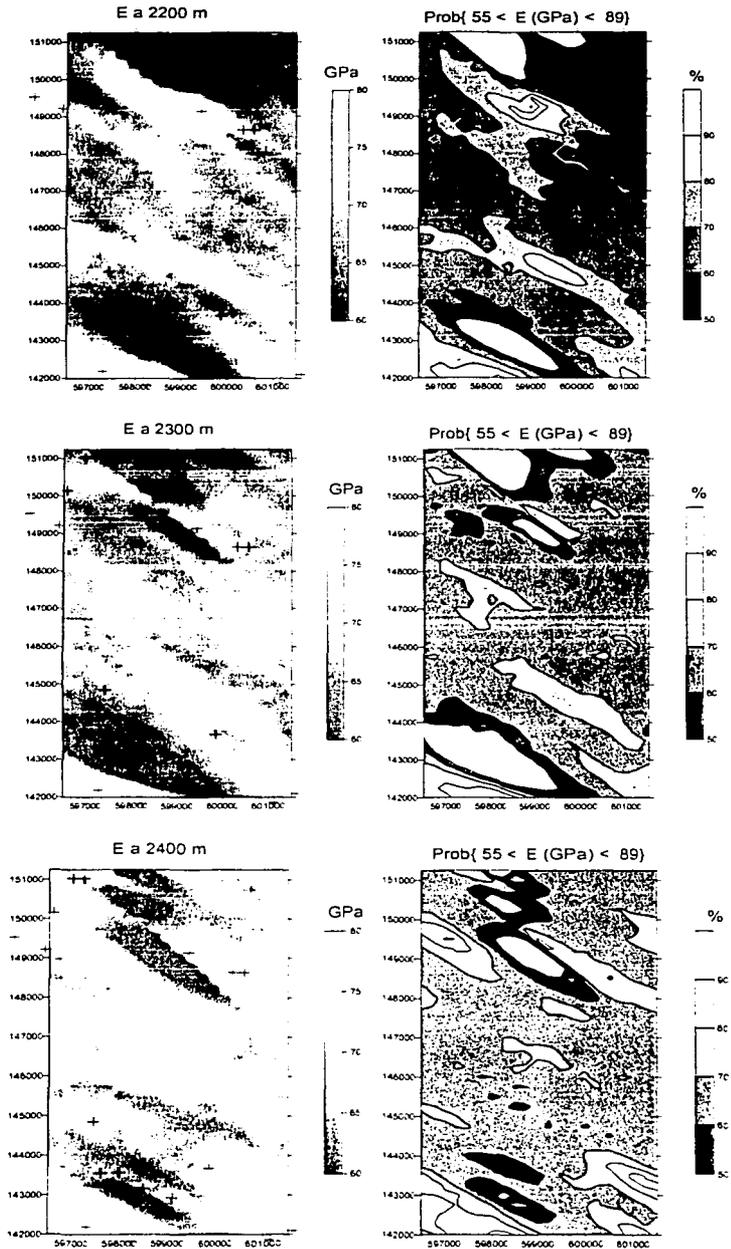


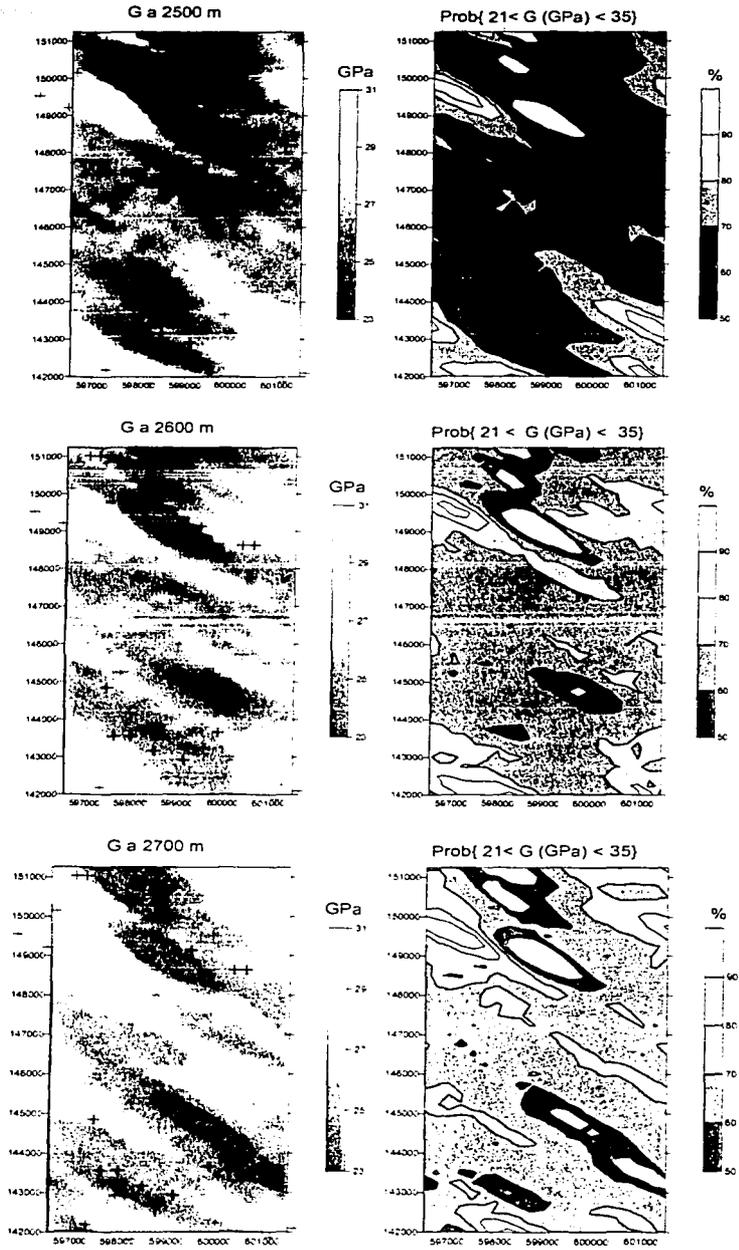
TESIS UCH
FALLA DE IMPRIMEN



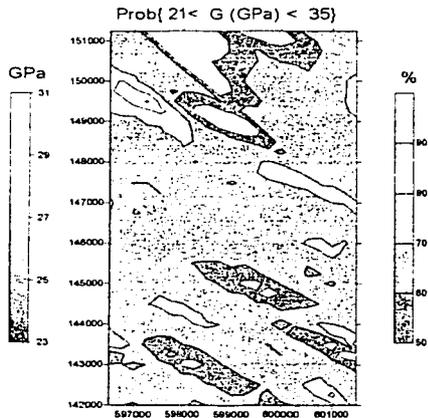
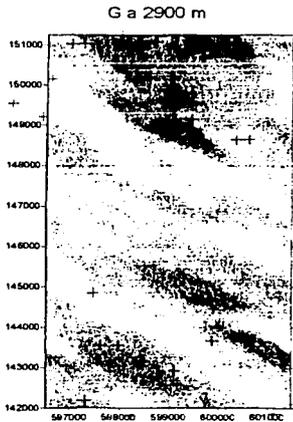
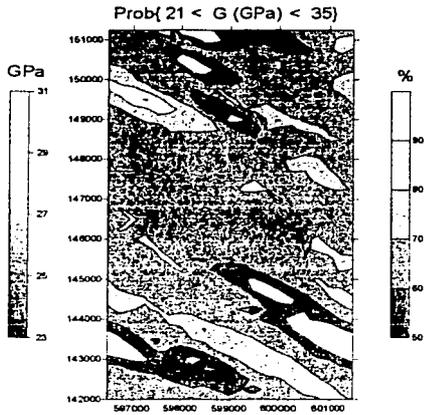
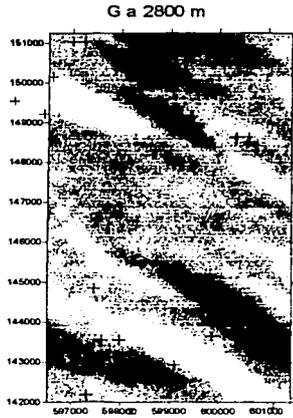
TESIS DE GRADUACION
MAY 2008



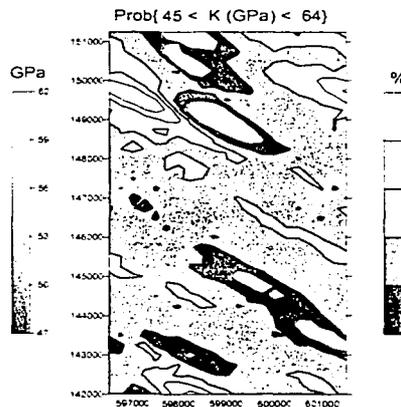
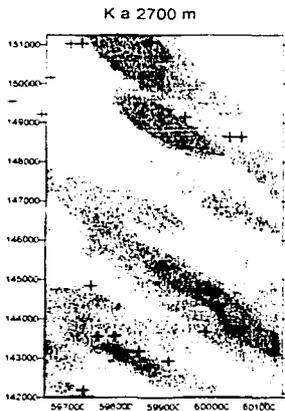
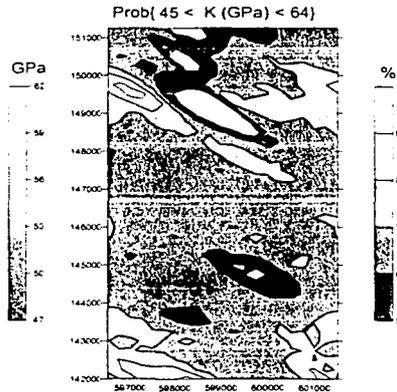
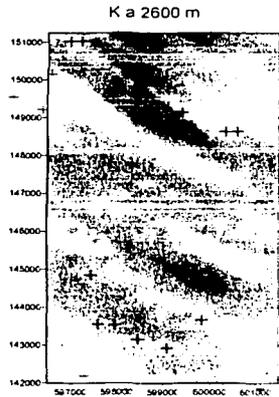
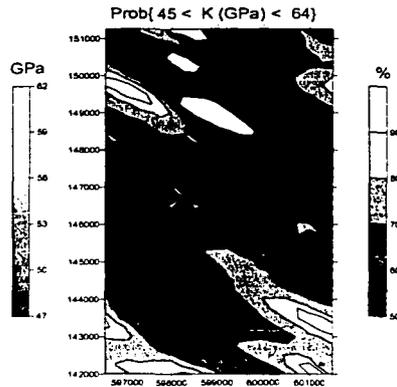
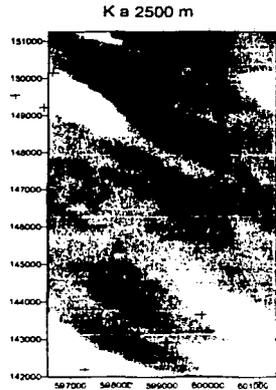




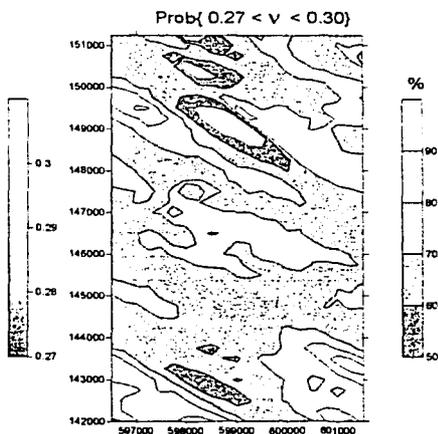
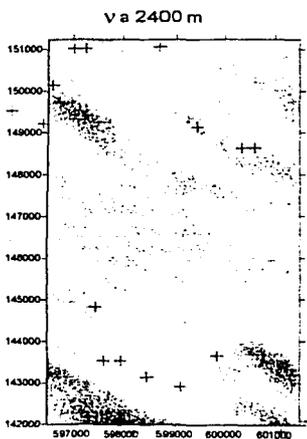
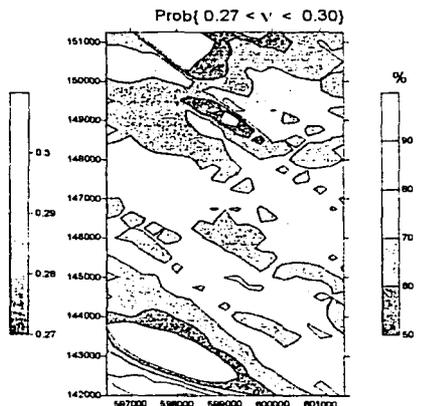
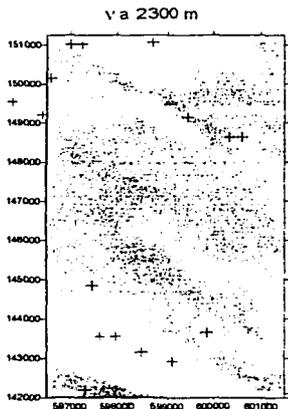
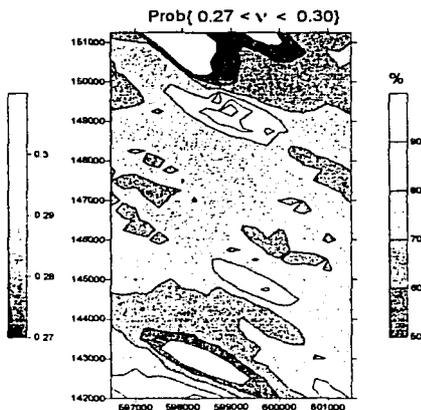
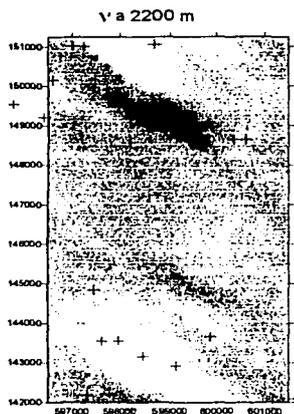
TESIS CON
FALLA DE CUBIETA



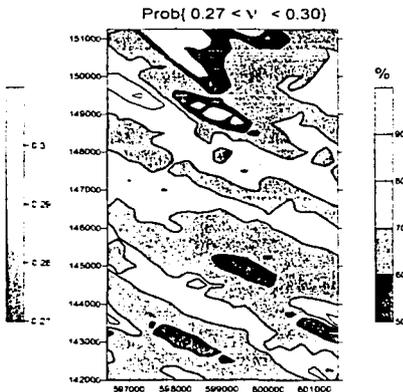
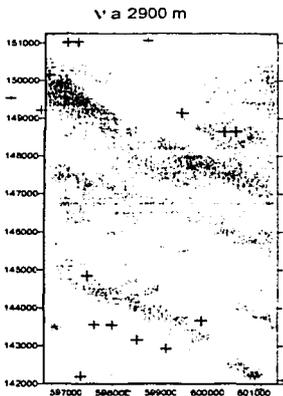
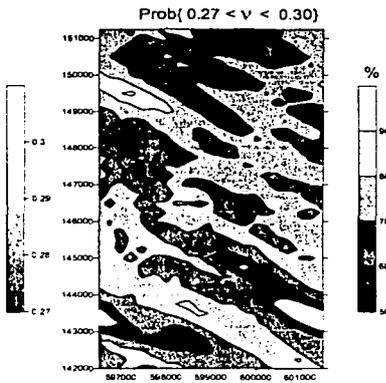
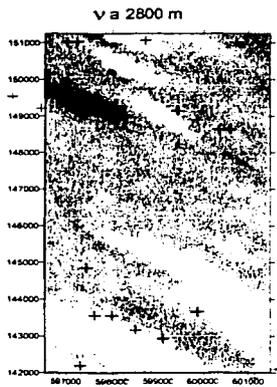
TESIS COMPLETA
FALLA I



TESIS CON
FALLA REVISADA



TERCERA PARTE
 14



TESIS COM
FALLA DE MICHEN

