



UNIVERSIDAD NACIONAL AUTÓNOMA DE
MÉXICO

FACULTAD DE INGENIERÍA

“MODELADO GEOESTADÍSTICO DE
PROPIEDADES PETROFÍSICAS DE
YACIMIENTOS”

T E S I S

QUE PARA OBTENER EL TÍTULO DE

I N G E N I E R O P E T R O L E R O

P R E S E N T A :

ALBERTO LASTIRI PÉREZ GALLARDO

DIRECTOR DE TESIS: M.C. JAIME ORTIZ RAMÍREZ

CIUDAD UNIVERSITARIA, MÉXICO, D.F., JUNIO 2006





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.

AGRADECIMIENTOS:

Agradezco al M.C. Jaime Ortiz Ramírez, Ing. Manuel Villamar Viguera, M.I. José. A. Gómez Cabrera, Ing. Javier Calderón Navarro y al Ing. Fernando Rosique Naranjo, por sus comentarios y tiempo dedicado al presente trabajo, gracias por sus consejos y compartir sus experiencias.

A mis Padres y Hermanos gracias por confiar en mí.....los quiero.

A la Lic. Alejandra Lastiri López por su apoyo incondicional a lo largo de mi carrera.....gracias por escucharme y apoyarme.

Agradezco a toda la familia Lastiri Frías por siempre preguntar: “Donde anda el Beto!!!”

A toda esa Pandillita de chiquillos y chiquillas que se fue acumulando durante todo este tiempo.....no quiero omitir a nadie.....cada uno de ustedes sabe de las aventuras que sucedieron.....desde el derroche hasta la congoja.....gracias por soportarme.

Al Ing. Oscar Osorio Peralta y al Ing. Ulises Neri Flores por amortiguar mi primeros pasos dentro del mundo real.

Agradezco a la UNAM y a la Facultad de Ingeniería por haberme brindado la oportunidad de estar, vivir, observar, juzgar y escoger libremente la vida.

*Mi abuela Marcelina solía decir: “poderoso caballero es..... Don Dinero”.....
con todo el dolor de mi corazón.....
que razón tenía.*

*Mi abuelo Israel solía decir: “No es por Cantidad.....es por Calidad”....
.y así es.*



ÍNDICE

ÍNDICE	ii
RESUMEN	vii
CAPITULO 1. INTRODUCCIÓN: HETEROGENEIDAD DE UN YACIMIENTO	1
1.1. INTRODUCCIÓN	2
1.2. CARACTERIZACIÓN ESTÁTICA DE YACIMIENTOS	2
1.3. HETEROGENEIDAD DE UN YACIMIENTO	5
1.3.1. DEFINICIÓN DE ESCALAS	5
1.3.2. EFECTO DE LA VARIACIÓN DE ESCALAS	7
1.3.2.1. MEDICIONES	9
1.3.2.2. EFECTO EN EL COMPORTAMIENTO DEL YACIMIENTO	10
1.4. MODELADO DE LA HETEROGENEIDAD DE UN YACIMIENTO	13
1.4.1. MODELADO CONVENCIONAL	13
1.4.2. MODELADO EN PARALELO	17
1.5. USO DE LA GEOESTADÍSTICA EN EL MODELADO DE UN YACIMIENTO	21
CAPITULO 2. PRINCIPIOS DE LA GEOESTADÍSTICA	23
2.1. INTRODUCCIÓN	24
2.2. CONCEPTOS BÁSICOS DE ESTADÍSTICA	24
2.2.1. ESTADÍSTICA DESCRIPTIVA	25
2.2.1.1. DISTRIBUCIÓN DE FRECUENCIAS	25
2.2.1.1.1. VARIABLES CUANTITATIVAS DISCRETAS	25
2.2.1.1.2. VARIABLES CUANTITATIVAS CONTINUAS	26
2.2.1.1.3. HISTOGRAMA	28
2.2.1.1.4. POLÍGONOS DE FRECUENCIA	29
2.2.1.1.5. OJIVA	30
2.2.1.2. ESTADÍSTICA PARA DISTRIBUCIONES UNIDIMENSIONALES	31
2.2.1.3. DISTRIBUCIÓN CONDICIONAL DE FRECUENCIA	34
2.2.1.4. ESTADÍSTICA PARA DISTRIBUCIONES BIDIMENSIONALES	34
2.2.1.4.1. COVARIANZA PARA EL ANÁLISIS ESTADÍSTICO	35
2.2.1.4.2. COEFICIENTE DE CORRELACIÓN PARA EL ANÁLISIS ESTADÍSTICO	35
2.2.1.4.3. REGRESIÓN LINEAL	36
2.2.2. ESTADÍSTICA INFERENCIAL	37
2.2.2.1. MUESTRA ALEATORIA	37
2.2.2.2. ESTADÍSTICO	38
2.2.2.3. ESPACIO MUESTRAL	39
2.2.3. PROBABILIDAD: ENFOQUE FRECUENTISTA	39
2.2.3.1. PROBABILIDAD CONDICIONAL	41
2.2.4. VARIABLE ALEATORIA	43

2.2.4.1. VARIABLE ALEATORIA DISCRETA	45
2.2.4.2. VARIABLE ALEATORIA CONTINUA	47
2.2.4.3. FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA	48
2.2.4.4. VALOR ESPERADO DE UNA VARIABLE ALEATORIA	49
2.2.4.5. MOMENTOS DE LA VARIABLE ALEATORIA	50
2.2.4.6. COEFICIENTE DE VARIACIÓN	52
2.2.4.7. ERROR ESTÁNDAR	52
2.2.5. VARIABLES ALEATORIAS BIDIMENSIONALES	52
2.2.5.1. VARIABLES ALEATORIAS BIDIMENSIONALES DISCRETAS	52
2.2.5.2. VARIABLES ALEATORIAS BIDIMENSIONALES CONTINUAS	54
2.2.5.3. FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA CONJUNTA	54
2.2.5.4. VALORES ESPERADOS PARA FUNCIONES CON DOS VARIABLES ALEATORIAS	55
2.2.5.5. COVARIANZA	56
2.2.5.6. COEFICIENTE DE CORRELACIÓN	57
2.3. OBJETIVO DE LA GEOESTADÍSTICA	57
2.3.1. CONCEPTO DE GEOESTADÍSTICA	58
2.3.2. DIFERENCIA ENTRE ESTADÍSTICA Y GEOESTADÍSTICA	60
2.3.3. VARIABLES ALEATORIAS REGIONALIZADAS	61
2.3.4. REQUERIMIENTOS DE ESTACIONARIEDAD	62
2.3.5. RELACIONES ESPACIALES	65
2.3.5.1. COVARIANZA PARA EL ANÁLISIS DE VARIOGRAMAS	65
2.3.5.2. COEFICIENTE DE CORRELACIÓN PARA EL ANÁLISIS DE VARIOGRAMAS	66
2.3.5.3. VARIOGRAMAS	67
2.3.6. CONOCIMIENTO DEL PROBLEMA	68
CAPITULO 3. ANÁLISIS PETROFÍSICO	69
3.1. INTRODUCCIÓN	70
3.2. PETROFÍSICA Y PARÁMETROS PETROFÍSICOS	70
3.3. DESCRIPCIÓN DE YACIMIENTOS MEDIANTE CARACTERIZACIÓN PETROFÍSICA	72
3.3.1. INTERPRETACIÓN DE REGISTROS GEOFÍSICOS	73
3.3.2. CORRELACIÓN DE REGISTROS	76
3.4. MODELADO PETROFÍSICO	79
3.4.1. MODELADO DETERMINÍSTICO Y ESTOCÁSTICO	80
3.4.2. ANÁLISIS DE DATOS	84
3.4.2.1. TRANSFORMACIÓN DE DATOS	84
3.4.2.2. ANÁLISIS DEL VARIOGRAMA	90
3.4.2.3. CORRELACIÓN DE PARÁMETROS PETROFÍSICOS	90
3.5. ESTIMACIÓN DE VARIOGRAMAS	91
3.5.1. VARIOGRAMA EXPERIMENTAL	93
3.5.2. PROBLEMAS COMUNES ENCONTRADOS EN EL CÁLCULO DE VARIOGRAMAS	97
3.5.3. MODELADO DE VARIOGRAMAS	98
3.5.3.1. PARÁMETROS DEL VARIOGRAMA	98
3.5.3.2. MODELOS TEÓRICOS DE VARIOGRAMAS	100

3.5.3.3. VALIDACIÓN DEL MODELO TEÓRICO	104
3.5.3.4. VALIDACIÓN CRUZADA	105
3.5.4. ANÁLISIS DE ANISOTROPÍA	105
3.5.4.1. EFECTO PROPORCIONAL	106
3.5.4.2. PROBLEMAS EN EL MODELADO DE VARIOGRAMAS	107
3.6. APLICACIÓN PRÁCTICA	109
CAPITULO 4. MODELADO NUMÉRICO DE YACIMIENTOS	121
4.1. INTRODUCCIÓN	122
4.2. MODELADO POR KRIGING	122
4.2.1. KRIGING SIMPLE	130
4.2.2. KRIGING ORDINARIO	134
4.2.3. COKRIGING	136
4.3. MODELADO POR SIMULACIÓN GAUSSIANA SECUENCIAL	140
4.3.1. SIMULACIÓN DE UNA VARIABLE SIMPLE	141
4.3.2. COSIMULACIÓN GAUSSIANA SECUENCIAL	148
4.4. APLICACIÓN PRÁCTICA	150
CAPITULO 5. ESCALAMIENTO DE PROPIEDADES	157
5.1. INTRODUCCIÓN	158
5.2. DEL MODELO ESTÁTICO AL MODELO DINÁMICO	158
5.3. MALLA DEL MODELO DINÁMICO	160
5.4. ESCALAMIENTO DE PROPIEDADES PETROFÍSICAS ESTÁTICAS	160
5.4.1. ESPESOR NETO	161
5.4.2. POROSIDAD	162
5.4.3. SATURACIÓN DE AGUA	163
5.5. ESCALAMIENTO DE PROPIEDADES PETROFÍSICAS DINÁMICAS - FLUJO MONOFÁSICO	165
5.5.1. VALORES DE PERMEABILIDAD PARA CASOS IDEALES	165
5.5.2. MÉTODO DEL ESTRATO INCOMPLETO	168
5.5.3. MÉTODO DE RENORMALIZACIÓN	170
5.5.4. TENSOR DE PERMEABILIDAD	171
5.5.5. FLUJO EN ESTADO ESTACIONARIO - SOLUCIÓN NUMÉRICA	173
5.6. ESCALAMIENTO DE PROPIEDADES PETROFÍSICAS DINÁMICAS - FLUJO MULTIFÁSICO	174
5.6.1. MÉTODOS ANALÍTICOS	175
5.6.2. MÉTODOS NUMÉRICOS	182
5.7. PRESIÓN CAPILAR	185

5.8. APLICACIÓN PRÁCTICA	186
CAPITULO 6. CONCLUSIONES Y RECOMENDACIONES	195
APÉNDICE A: GLOSARIO	A-1
APÉNDICE B: ESTADÍSTICA PARA DISTRIBUCIONES UNIDIMENSIONALES	B-1
B.1. MEDIDAS DE POSICIÓN	B-2
B.1.1. MEDIA	B-2
B.1.2. MEDIANA	B-3
B.1.3. CUARTILES	B-6
B.1.4. QUINTILES	B-7
B.1.5. DECILES	B-8
B.1.6. PERCENTILES	B-9
B.1.7. MODA	B-9
B.2. MEDIDAS DE DISPERSIÓN	B-10
B.2.1. DESVIACIÓN ESTÁNDAR	B-10
B.2.2. VARIANZA	B-12
B.2.3. COEFICIENTE DE VARIACIÓN	B-12
B.2.4. COEFICIENTE DE ASIMETRÍA	B-13
B.2.5. COEFICIENTE DE APUNTAMIENTO	B-14
B.2.6. RANGO INTERCUARTÍLICO	B-15
NOMENCLATURA	N-1
RELACIÓN DE FIGURAS Y TABLAS	RFT-1
REFERENCIAS BIBLIOGRÁFICAS	RB-1

RESUMEN

El objetivo principal del presente trabajo es el de desarrollar sobre fundamentos teóricos y prácticos el modelado de propiedades petrofísicas en un yacimiento basándose en las aplicaciones de la Geoestadística, con el fin de caracterizar de forma estática un yacimiento.

El primer capítulo introduce el concepto de caracterización estática de yacimientos y los aspectos con mayor relevancia que se necesitan para desarrollar dicha disciplina. Se discute la definición de heterogeneidad de un yacimiento así como la asociación de escalas e importancia de las mismas durante la descripción y predicción del comportamiento de yacimientos.

En el capítulo dos, se comentan los principios básicos de la Estadística convencional, necesarios para la comprensión de los procedimientos y análisis geoestadísticos.

En el capítulo tres se hace mención de los conceptos más importantes en lo referente al entorno petrofísico. Adicionalmente se analizan algunas diferencias existentes entre el modelado determinístico y el modelado estocástico, ya que son la estructura trascendental que soporta el principio de construcción de un modelado numérico de yacimientos, así como las técnicas necesarias para el desarrollo de este. El concepto de variograma es la herramienta geoestadística fundamental para obtener los modelados discutidos en el presente trabajo, por lo tanto se detalla un amplio análisis con respecto a este tema, con la finalidad de explotar al máximo las características relevantes de este concepto, aplicadas a un análisis petrofísico.

En el cuarto capítulo se abordan los aspectos y ecuaciones más significativas que se emplean para desarrollar un modelado numérico de yacimientos. El modelado por Kriging y el modelado por simulación Gaussiana secuencial son dos de las técnicas con mayor divulgación y eficiencia para conseguir dicho objetivo. Adicionalmente se menciona bajo que condiciones se pueden emplear estos modelos, así como las ventajas y desventajas de su utilización.

En el capítulo cinco se presenta algunas técnicas de escalamiento utilizadas en el cálculo de propiedades del yacimiento para la construcción de mallas en las cuales se generan las descripciones de propiedades geológicas y petrofísicas, para poder trabajar en un modelo de simulación numérica. Se realiza la descripción de técnicas de escalamiento para propiedades estáticas del yacimiento incluyendo espesor neto, porosidad y saturación. También se describen algunas técnicas de escalamiento para propiedades dinámicas como permeabilidad.

Finalmente en el capítulo seis se presentan las conclusiones y recomendaciones del presente trabajo.

Cabe resaltar, que esta tesis se limita a expresar una metodología de tipo general para obtener un modelado geoestadístico de propiedades petrofísicas de yacimientos y poder transitar de un modelo estático a un modelo dinámico en términos del flujo de trabajo en la caracterización de yacimientos, es decir, dejar una malla lista para la simulación de yacimientos. Existe la posibilidad de que en el futuro, aplicaciones con especificaciones detalladas, de acuerdo al tipo de yacimiento puedan publicarse, y mostrar así tanto la metodología que se desarrolla como la eficiencia de las técnicas descritas con base en el objetivo o meta que se pretende lograr.

CAPITULO 1

INTRODUCCIÓN: HETEROGENEIDAD DE UN YACIMIENTO

1.1. INTRODUCCIÓN.

En el presente capítulo se introduce el concepto de caracterización estática de yacimientos y los aspectos más importantes que se necesitan para desarrollar dicha disciplina. Se discute la definición de heterogeneidad de un yacimiento así como la asociación de escalas e importancia de las mismas durante la descripción y predicción del comportamiento de yacimientos. Los siguientes subcapítulos se enfocan en describir brevemente la diferencia que hay entre el modelado convencional y el modelado en paralelo, enunciando algunas ventajas y desventajas de dichos métodos. La última sección explica de manera muy general la importancia de la Geoestadística en la descripción de yacimientos.

1.2. CARACTERIZACIÓN ESTÁTICA DE YACIMIENTOS.

La caracterización estática de yacimientos puede definirse como el proceso de descripción de las características geológicas de un yacimiento mediante la utilización de todo tipo de datos disponibles. La descripción puede ser cualitativa o cuantitativa, en donde se incluye fundamentalmente la determinación de distribuciones de tamaño de poro y grano, porosidades y permeabilidades del yacimiento, distribución de facies, ambientes de depósito, y descripción de cuencas. Las metas de esta disciplina se enfocan a obtener un modelo geológico – petrofísico a escala de yacimiento. Considerando la variedad de características que necesitan describirse, se recopilan datos disponibles de distintas fuentes de información. Estos tipos de datos incluyen, entre otros: datos de núcleos y registros geofísicos; pruebas en distintos pozos; historiales de producción y en algunos casos, datos sísmicos en 2D y 3D, perfiles sísmicos verticales y tomografía de pozos.

De manera ideal, todas las fuentes de información deben de incluirse en una descripción del yacimiento. Entre más información sea utilizada este proceso se realiza de manera más eficaz. Sin embargo, en la práctica se presentan algunos problemas asociados con el uso simultáneo de la gama de datos, como se enuncia a continuación:

No todos los datos están disponibles al mismo tiempo. Conforme comienza la exploración y delimitación de yacimientos, la disponibilidad de datos es escasa por lo que se tiene mayor incertidumbre en la descripción del yacimiento; en un futuro cuando empieza la producción, se recopila información adicional que puede ser de gran utilidad.

La calidad en los datos colectados puede variar significativamente. El peso e importancia asignada a cada conjunto de datos debe ser diferente; en los análisis de propiedades del yacimiento se requiere realizar un mínimo número de suposiciones para garantizar una mayor confiabilidad en la utilización de datos, permitiendo contabilizar la incertidumbre en el muestreo de los mismos, lo cual es difícil de lograr.

La información colectada de distintas fuentes de información no puede medirse en una sola escala, además de que la resolución de la medida de valores puede diferir abruptamente, por lo que es necesario realizar cálculos adecuados sobre las escalas de medición empleadas, así como denotar la diferencia entre las mismas.

La interpretación de datos suele ser cualitativa. Desafortunadamente, para muchas aplicaciones en yacimientos es necesario describir cuantitativamente las características de éstos. Se realizan intentos para cuantificar algunos datos cualitativos, es decir, contabilizar la distribución de propiedades petrofísicas, pero no existen procedimientos apropiados para usar este tipo de datos, aun en la descripción de propiedades del yacimiento.

INTRODUCCIÓN: HETEROGENEIDAD DE UN YACIMIENTO

El uso de datos de producción como entrada de datos en la descripción del yacimiento presenta importantes problemas debido a las complejas relaciones matemáticas entre la descripción del mismo y los datos de producción. Como se muestra en la figura 1.1, simular con diferentes datos puede conllevar a obtener una representación particular del yacimiento en la que de acuerdo a los mecanismos de producción, sea posible determinar su comportamiento, para posteriormente intuir este en un futuro. Los modelos empleados para simular el yacimiento son invariablemente complejos y no lineales con respecto a los parámetros de entrada.

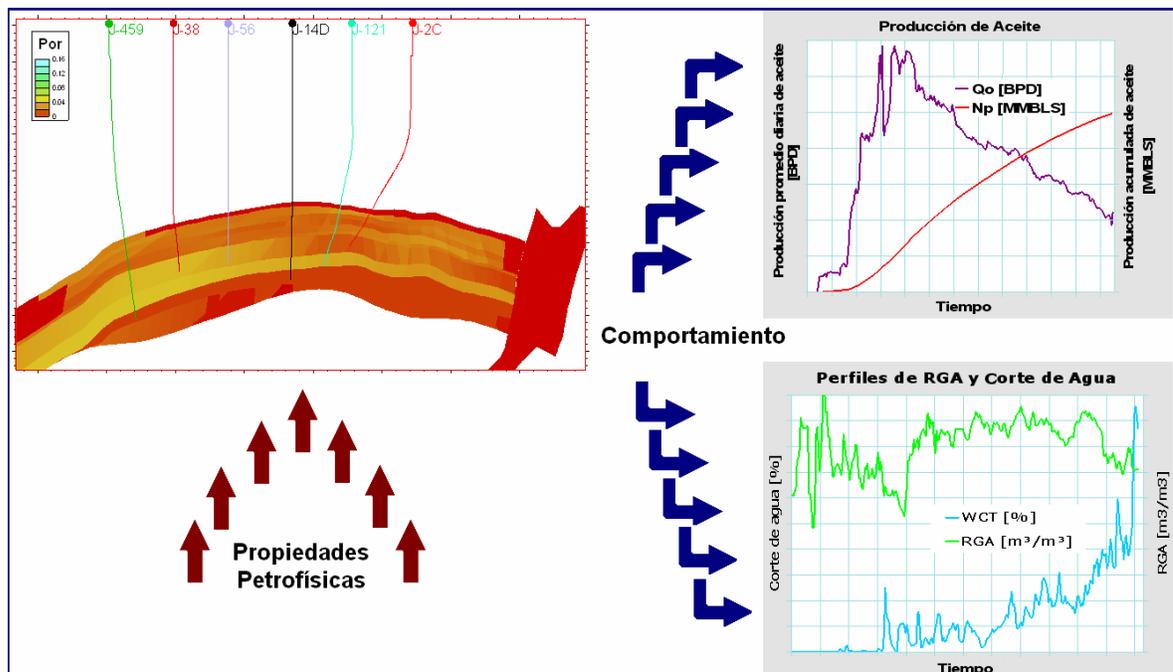


Figura 1.1. Proceso típico de la simulación de yacimientos.

Desde el punto de vista de la Ingeniería Petrolera, los datos de producción son probablemente, los datos con mayor valor; no importa la manera en que se construye la descripción del yacimiento, si en esta no pueden reflejarse los datos de producción. Cuando se tiene en marcha el proceso de producción de un yacimiento, la recolección de datos se regulariza y es más práctica.

Como resultado de las limitaciones antes mencionadas, algunos métodos para describir yacimientos son útiles aun cuando se emplea solo parte de los datos que están al alcance del grupo interdisciplinario.

1.3. HETEROGENEIDAD DE UN YACIMIENTO.

La heterogeneidad de un yacimiento se define como la variación de propiedades en función del espaciamiento que existe entre las zonas que comprenden su totalidad. Idealmente si los yacimientos fuesen homogéneos, la medición de propiedades en cualquier localidad permitiría realizar una minuciosa y detallada descripción de estos. Para interpretar adecuadamente la heterogeneidad del yacimiento es necesario predecir la variación de algunas propiedades de la roca como pueden ser porosidad, saturación de fluidos, espesor, fallas, fracturas y facies, entre otras.

1.3.1. DEFINICIÓN DE ESCALAS.

Es necesario medir la heterogeneidad del yacimiento en distintas escalas. La variación de escalas puede ir desde pequeña escala, en la que se hacen cálculos de la distribución de tamaño de grano a escala microscópica, hasta una descripción geológica a nivel regional. A pesar de que las propiedades del yacimiento varían sobre las distintas escalas continuas, es conveniente dividir dichas escalas en un cierto rango de niveles como se muestra en la figura 1.2, definidas arbitrariamente. En el presente trabajo se considera definir 4 escalas para la agrupación de heterogeneidades. A continuación se resume la clasificación de cada escala.

Heterogeneidades Microscópicas: Estas suelen llamarse en ocasiones heterogeneidades al nivel de poro. Incluyen la identificación de distribución de tamaño de poro (geometría), tamaño de garganta (espacio poroso que comunica los poros de una roca), litología de rocas y mineralogía.

Heterogeneidades Macroscópicas: Son aquellas mediciones que se realizan a nivel de núcleo. En este tipo de interpretación se incluye la medición de permeabilidades, porosidades, saturaciones de fluidos y mojabilidad de rocas mediante el análisis de núcleos, propiedades petrofísicas como lo son permeabilidades relativas y presiones capilares.

Heterogeneidades Megascópicas: Son aquellas heterogeneidades que poseen el mismo orden en cuanto a magnitud, como lo tienen las celdas de malla en un simulador de yacimientos, las cuales comprenden alguna magnitud de distancia considerable a lo largo y ancho. Las propiedades medidas dentro de esta escala incluyen la interpretación de registros geofísicos, datos de pruebas de presión y determinación de saturación residual de aceite. Se considera también la utilización de algunos atributos sísmicos en esta escala.

Heterogeneidades Gigascópicas: Son aquellas heterogeneidades que se miden a escala de yacimiento considerando el espaciamiento entre los pozos. Las propiedades que se miden dentro de esta escala incluyen determinación de permeabilidades con base en pruebas de pozos y pruebas con trazadores entre pozos, además se considera la interpretación de horizontes geológicos para analizar verticalmente la columna estratigráfica de un yacimiento, datos sísmicos de reflexión o contrastes de impedancia acústica que representan la estructura geológica a nivel regional así como localización de fallas para la delimitación de compartimentos de yacimientos.

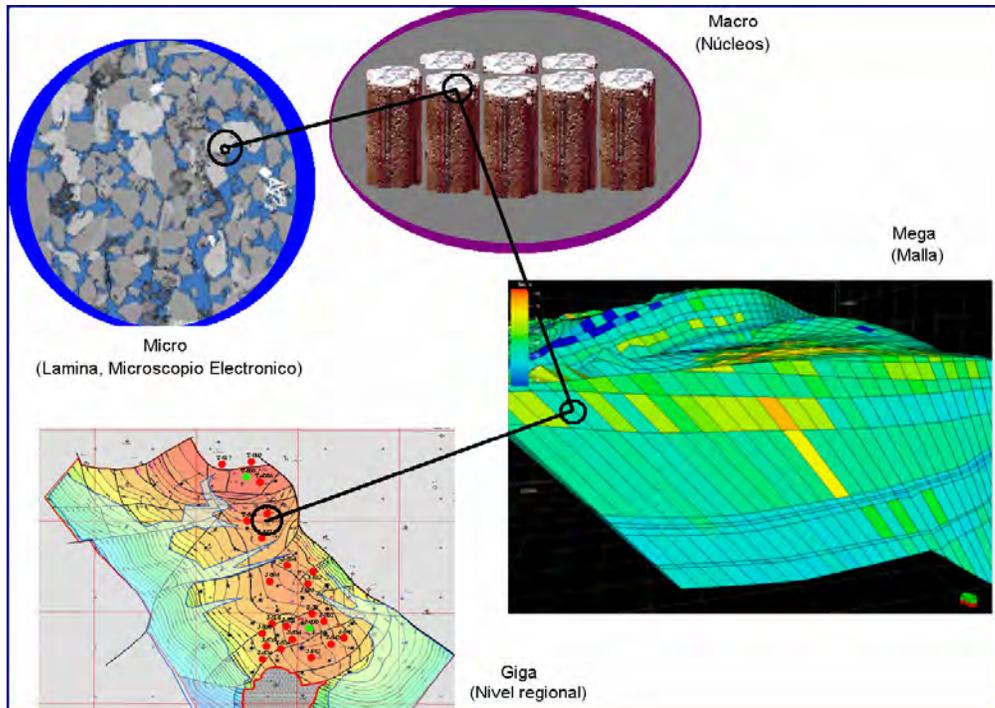


Figura 1.2. Escalas de Heterogeneidad.

1.3.2. EFECTO DE LA VARIACIÓN DE ESCALAS.

La tabla 1.1 proporciona una breve descripción de varios niveles de heterogeneidad así como su efecto en el comportamiento de flujo en el yacimiento. A continuación se explica el significado de las columnas en la tabla:

Escala de Medición: Las escalas de medición indicadas en la tabla 1.1 son arbitrarias. Las propiedades del yacimiento cambian continuamente en todas las escalas. De esta manera, la división de escalas es grande de acuerdo al tipo de medición que se realiza a varios niveles. Obviamente existen algunas superposiciones entre las distintas mediciones, por lo que algunas mediciones podrían caer dentro de dos posibles escalas.

INTRODUCCIÓN: HETEROGENEIDAD DE UN YACIMIENTO

Tipo	Nivel	Escala de Medición	Medición	Observación Geológica	Efecto en el Comportamiento de Flujo	Efecto en el Proceso de Flujo
Micro	Poros	micrómetros	*Geometría de Poro *Tamaño de Grano *Mineralogía	*Textura *Mineralogía *Microfracturas	Eficiencia de Desplazamiento	Complejos Procesos de Recuperación Mejorada
Macro	Núcleo	centímetros	*Permeabilidad *Permeabilidad Relativa *Porosidad *Presión Capilar *Mojabilidad *Capilaridad *Registros de imagen	*Estratificación *Facies *Fracturas *Vúgulos *Cavernas	Eficiencia de Barrido	Recuperación Secundaria (Inyección de Agua y/o Gas) Sistemas Artificiales de Producción
Mega	Celda de malla	metros	*Registros Geofísicos *Trazadores de Pozos *Sísmica	*Propiedades Petrofísicas *Litología *Límites verticales de capas *Tipos de estratificación	Eficiencia de Barrido	Recuperación Primaria y Secundaria (Inyección de Agua y/o Gas) ejemplo: Avance irregular del contacto agua-aceite
Giga	Interpozo	kilómetros	*Pruebas de Pozos *Sísmica de reflexión, impedancia acústica y perfil sísmico vertical a nivel regional *Pruebas con Trazadores Interpozo	*Fallas *Horizontes Geológicos *Volumen de Roca	Eficiencia de Extracción	Recuperación Primaria ejemplo: Optimización de Unidades de flujo

Tabla 1.1. Escalas de Heterogeneidad de Yacimientos¹.

1.3.2.1. MEDICIONES.

Como se muestra en la tabla 1.1 los diferentes tipos de mediciones conducen a diferentes niveles de medición. Normalmente, las propiedades no se miden en todos los niveles. Cuando se describen las propiedades del yacimiento en una localidad particular, es importante saber no solo la localización del yacimiento, también es necesario conocer sobre cual de estas propiedades se define.

Cualquier medición sobre una escala microscópica muestra variaciones significantes en sus propiedades; cuando la medición se realiza a gran escala, la variación observada en las propiedades a un menor nivel se nota debido al promedio aritmético que se emplea a escalas pequeñas. Eventualmente, en el otro extremo, cuando se miden las propiedades a escala de yacimientos, es posible medir un solo valor de porosidad, por ejemplo, que represente al yacimiento. En resumen, la variación de una propiedad medida dada va disminuyendo conforme aumenta la escala de medición. Esta dependencia en las escalas algunos autores la denominan como "efecto de apoyo".

En la figura 1.3, se ejemplifica una comparación entre una prueba de pozo y los datos de permeabilidad de un núcleo para un yacimiento de aceite. Como se indica en la figura, cuando se grafican los datos de permeabilidad como función del porcentaje de probabilidad, los datos de la prueba de pozo muestran una menor variación que los datos obtenidos con núcleos; se tiene un cierto grado de consistencia gracias al efecto de apoyo. Conforme se colectan los datos de la prueba de pozo sobre un volumen mucho más extenso, se promedian representativamente los valores de permeabilidad que se obtienen en todas las pequeñas escalas, es decir, que a medida que el promedio en el volumen se incrementa, la variación entre los valores medidos disminuye. De esta manera, los valores de permeabilidad determinados de la prueba de pozo,

muestran menos variación en comparación con aquellos obtenidos de datos de núcleos².

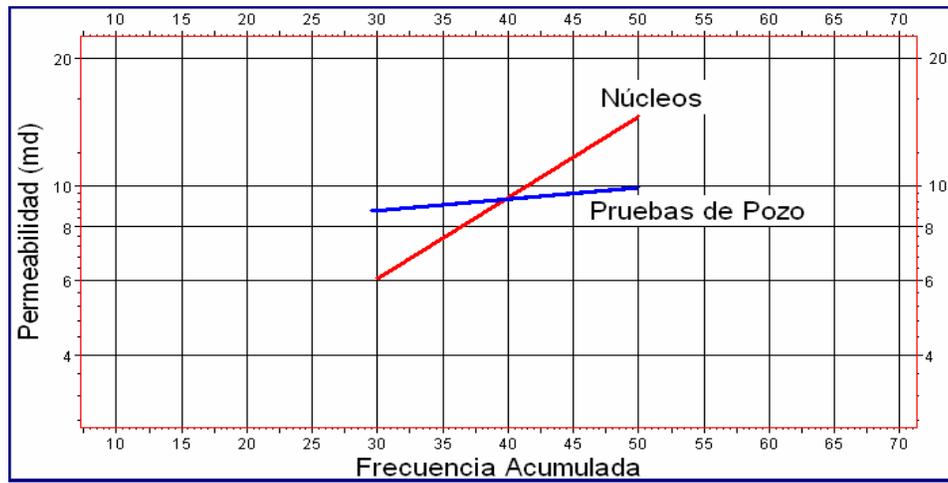


Figura 1.3. Comparación entre los datos de una prueba de pozo y núcleos.

La importancia del "efecto de apoyo" no se restringe únicamente a los valores de permeabilidad. Para la porosidad, los datos pueden medirse con núcleos o registros. Normalmente los datos de núcleos muestran mucho más variación que los datos de registros por lo que esta diferencia tendrá que ajustarse hasta cierto grado de compatibilidad antes de utilizar las mediciones del registro. Una apropiada descripción del yacimiento requiere calcular adecuadamente el "efecto de apoyo" para poder describir las propiedades del mismo.

1.3.2.2. EFECTO EN EL COMPORTAMIENTO DEL YACIMIENTO.

Las escalas de heterogeneidad son sumamente importantes ya que afectan el comportamiento del yacimiento. Las heterogeneidades a nivel microscópico permiten idealizar patrones preferenciales para identificar, por ejemplo, canales de flujo, pero en ocasiones se tienen malos resultados en el momento de cuantificar la eficiencia de desplazamiento de hidrocarburos, es

decir, se tienen altos valores de saturación residual de aceite, lo que impacta de manera considerable la cantidad de aceite recuperado en superficie.

En la figura 1.4, para un yacimiento típico de doble porosidad, donde el agua se comporta como fase continua, podría mejorarse el desplazamiento del aceite remanente mediante la modificación de las fuerzas capilares que existen entre el aceite y el fluido desplazante.

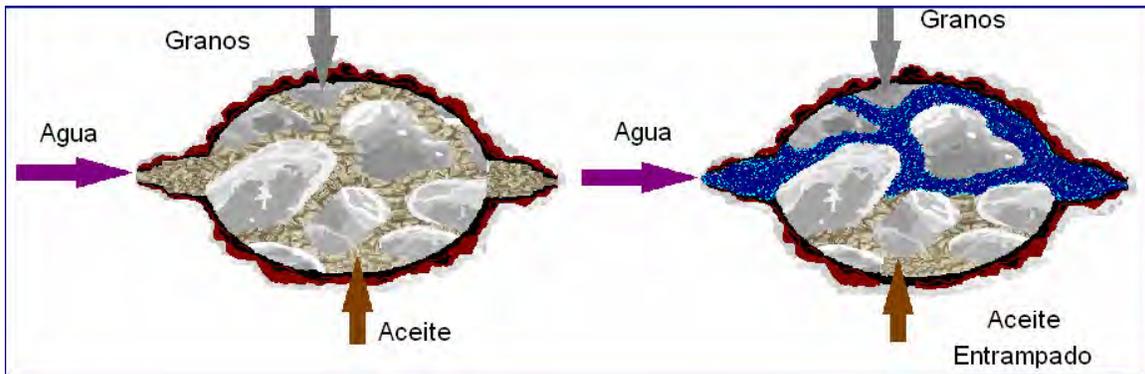


Figura 1.4. Representación de una trampa de aceite.

La heterogeneidad en macro y mega escala muestra patrones preferentes para mejorar el desplazamiento de fluidos. La figura 1.5a, muestra el impacto de la heterogeneidad vertical sobre la eficiencia de barrido vertical; entre mayor sea la permeabilidad del estrato, menor será la cantidad de aceite no recuperable, y viceversa. La figura 1.5b, representa la heterogeneidad areal sobre la eficiencia de barrido areal; el análisis de patrones de flujo permite establecer la manera en la que el fluido desplazante alcanza al pozo productor sin necesidad de acceder a todas las partes del yacimiento.

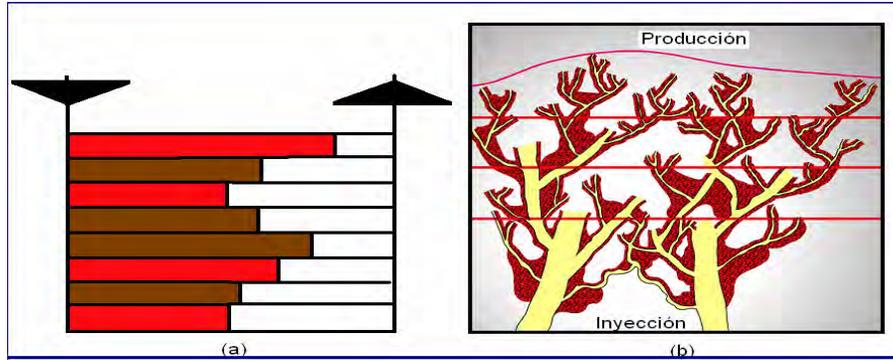


Figura 1.5. Efecto de las heterogeneidades en la eficiencia de barrido.

En lo que respecta a las heterogeneidades gigascópicas, como se muestra en la figura 1.6, estas pueden ayudar a determinar la cantidad de hidrocarburos almacenados en compartimentos aislados que no están conectados unos con otros, y tomar la decisión por ejemplo, de perforar pozos de relleno.

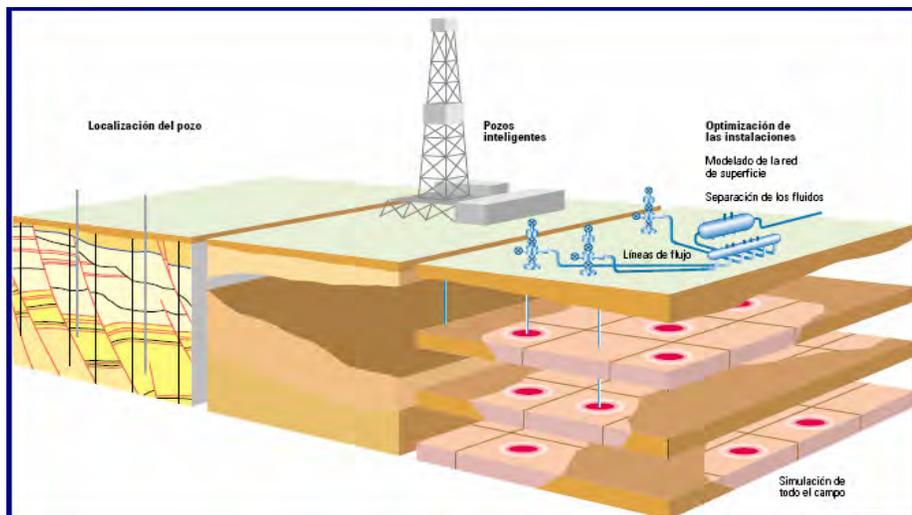


Figura 1.6. Efecto de las heterogeneidades gigascópicas en el comportamiento de un yacimiento.

1.4. MODELADO DE LA HETEROGENEIDAD DE UN YACIMIENTO.

Con el propósito de optimizar la producción, es necesario predecir el comportamiento del yacimiento. Desde hace algunas décadas, una de las técnicas más utilizada en la industria petrolera para predecir el comportamiento de los yacimientos consiste en el uso de simuladores.

Algunas ventajas que muestra esta disciplina sobre herramientas analíticas son:

- a) La habilidad de representar complejos procesos físicos de manera rigurosa.
- b) La habilidad de incorporar heterogeneidades del yacimiento.
- c) La habilidad de predecir futuros comportamientos del yacimiento, modificando sus condiciones.
- d) La habilidad de interpretar los comportamientos del yacimiento bajo distintas condiciones.

Comparando con métodos analíticos, la simulación de flujo en los yacimientos requiere mayor información en la entrada de datos para realizar una adecuada aplicación.

Una limitante que debe tomarse en cuenta para hacer ejercicios de simulación, es la dependencia en gran magnitud de los recursos computacionales.

1.4.1. MODELADO CONVENCIONAL.

Un modelado convencional es la aplicación más común para desarrollar la simulación de yacimientos, donde se intenta predecir el comportamiento futuro de los mismos. El modelo a simular tiene que validarse antes de que este pueda

utilizarse, ya que existen significantes incertidumbres en los parámetros de entrada usados en la simulación.

Matax y Dalton³ enuncian: "La forma más utilizada para construir adecuadamente un modelo que describa las principales características de un yacimiento, consiste en simular su comportamiento pasado y comparar la simulación realizada con el historial del comportamiento del yacimiento".

Al proceso de comparación que se efectúa entre el comportamiento del simulador, se le denomina como "ajuste de historia", cuya finalidad principal es la de agregarle mayor confiabilidad al modelo en la simulación de un comportamiento futuro.

El modelado utilizado comúnmente para representar un ajuste de historia se conoce como modelado en capas (estratos). Este tipo de modelado involucra definir al yacimiento en términos de algunos estratos que poseen diferentes propiedades petrofísicas, en donde se simula numéricamente el comportamiento del yacimiento ya sea en dos o tres dimensiones. Las ventajas de este tipo de modelado son:

- a) Es fácil de comprender la arquitectura física del yacimiento en términos de varios estratos
- b) Es ajustable a la Geología. Por ejemplo Webber y van Geuns⁴, notaron que el modelado en capas puede aplicarse a ciertos ambientes geológicos, incluyendo barreras amplias y cortas, arenas marinas poco profundas, arenas transgresivas, barras costa afuera, depósitos turbidíticos y eólicos, así como ríos trenzados
- c) Es fácil de utilizar ya que los simuladores numéricos son adaptables a mallas ortogonales, y preferentemente rectangulares
- d) Puede imitarse el historial del comportamiento mediante unas cuantas iteraciones

El modelado convencional presenta cuatro importantes desventajas:

- a) Secuencia de Modelado: Comprende cinco escenarios en general
 - 1) Análisis de datos
 - 2) Construcción del Modelo
 - 3) Ajuste de historia
 - 4) Calibración del modelo para un mejor ajuste de resultados
 - 5) Predicción de comportamientos futuros

Las dificultades en la secuencia de modelado, son que el proceso no es interdisciplinario, cualquier retraso en el proceso de modelado causa un correspondiente retraso en la determinación del resultado final y no hay flexibilidad para la incorporación adicional de datos o conocimientos disponibles mientras se tiene el estudio en proceso.

- b) Falta de dependencia en las escalas sobre las propiedades del yacimiento, asociada a la entrada y salida de datos.
- c) Incertidumbre en las estimaciones: Considerando las incertidumbres involucradas en la estimación de propiedades del yacimiento, es razonable suponer que la descripción que se ajusta a la historia del comportamiento no es única tampoco; el modelado convencional no permite predecir incertidumbres en un comportamiento futuro.
- d) Inclusión de características del yacimiento: En muchos casos las características geológicas exhiben complejos patrones que no es conveniente introducir en un modelado en capas. Webber y van Geuns⁴ describen tres categorías de yacimientos clásticos basándose en su arquitectura.
 - 1) Yacimientos en Capas: Estos yacimientos tienen suficientes estructuras continuas que pueden representarse en un modelado en capas

- 2) Yacimientos en Rompecabezas: Estos yacimientos están compuestos por series de cuerpos arenosos encajados fuera de fisuras. Algunos ejemplos de este tipo de yacimientos incluyen barreras en punta y cuerpos arenosos en abanicos turbidíticos
- 3) Yacimientos en Laberinto: Estos yacimientos consisten en complejos arreglos de cuerpos de arena y lentes. Aparecen en forma discontinua, y solo es posible realizar detalladas correlaciones cuando el espaciamiento entre pozos es pequeño, ejemplos de estos yacimientos incluyen rellenos de canal incrustados en barras de arena y arenas lacustres combinadas con canales de arena fluviales

La figura 1.7, muestra una representación esquemática de estos tres tipos de yacimientos.

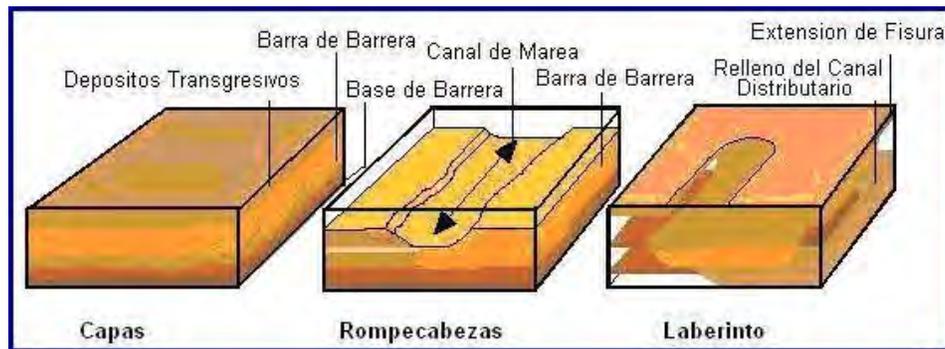


Figura 1.7. Modelos de Yacimientos⁵.

El incremento en la falta de continuidad puede observarse en los yacimientos en forma de rompecabezas y en forma de laberinto.

1.4.2 MODELADO EN PARALELO.

Aun en la presencia de una gran cantidad de datos, quedan muchas incertidumbres con respecto a la descripción de yacimientos. Estas incertidumbres pueden minimizarse si todos los datos disponibles son evaluados adecuadamente para su uso.

El modelado en paralelo ofrece una alternativa para la integración de datos de manera consistente. En el modelado en paralelo, un equipo multidisciplinario que esta conformado por Geólogos, Geofísicos, Ingenieros de Yacimientos e Ingenieros de Producción, realiza en conjunto el estudio del yacimiento.

El nombre de "paralelo" se debe a que la secuencia de pasos para conformar el modelado, se realiza de manera paralela o simultanea (ver figura 1.8.). En un principio, el modelo que respeta las heterogeneidades a gran escala puede construirse con información geológica-geofísica, y el comportamiento de flujo puede simularse con promedios de propiedades petrofísicas. La predicción del comportamiento de flujo puede alimentarse posteriormente con los estudios hechos por los Geoscientistas ya que ellos pueden proporcionar un punto de vista respecto a la posibilidad de realizar modificaciones a gran escala en el modelo original. El modelo ya refinado puede simularse una vez más para propósitos de calibración. Conforme el modelo se acerca progresivamente al comportamiento observado, puede adicionarse un nivel más alto de detalle respecto a las heterogeneidades para conseguir una mejor radiografía del yacimiento.

INTRODUCCIÓN: HETEROGENEIDAD DE UN YACIMIENTO

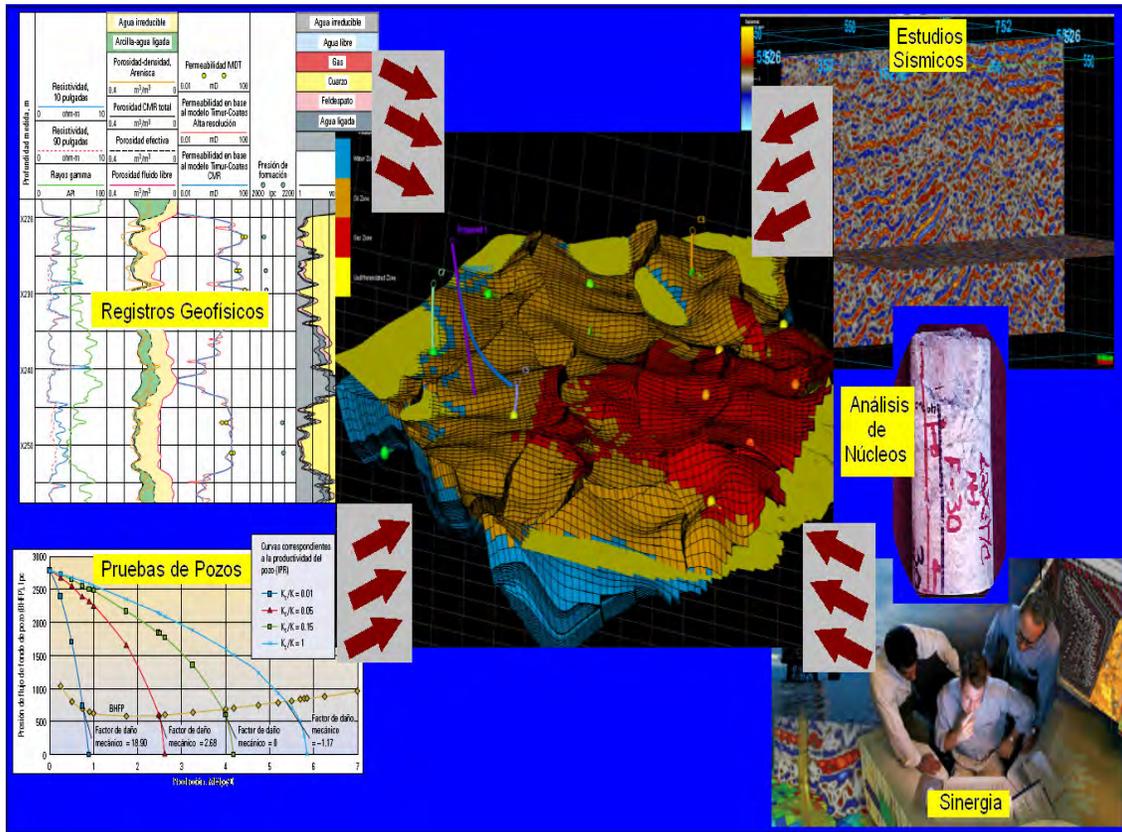


Figura 1.8. Descripción Integral de Yacimientos⁵.

Debido a que la descripción se crea mediante un esfuerzo combinado, puede tomarse en cuenta bastante información en la construcción del descripción final, además, si se consigue nueva información disponible durante el proceso, la misma puede incorporarse fácilmente.

La figura 1.9, muestra el ejemplo de un ajuste de historias en el que se utiliza un modelado en paralelo para la construcción de la descripción de un yacimiento del medio oriente⁶.

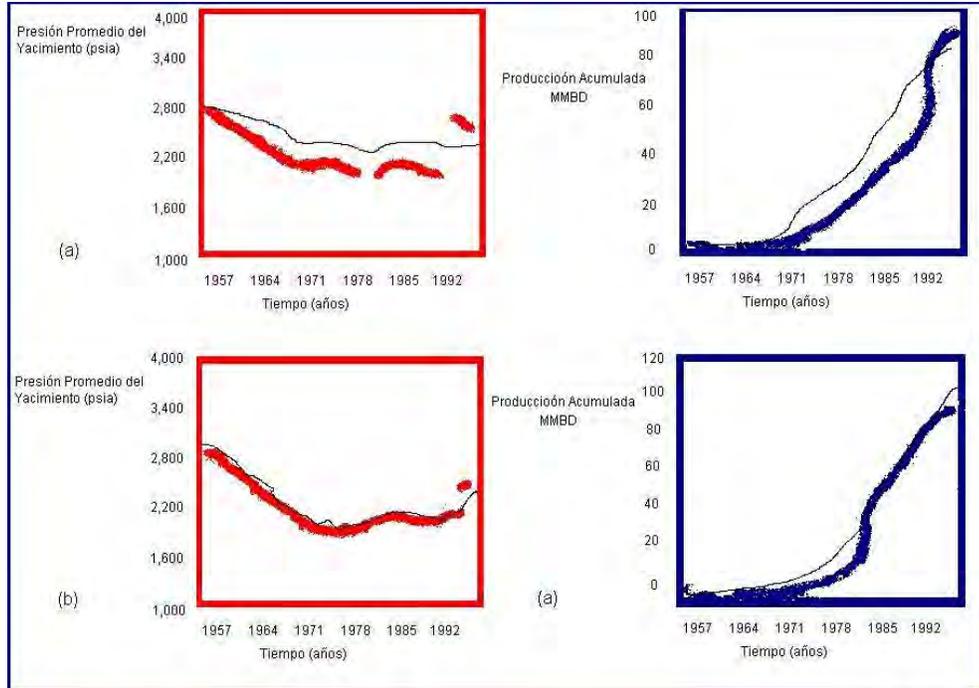


Figura 1.9. (a) Modelado convencional vs (b) Modelado en Paralelo.

Inclusión del efecto de capas: La inclusión de una apropiada escala es necesaria para una exitosa descripción de las propiedades del yacimiento. Para la descripción de propiedades de una celda de malla se necesitan procesos de escalamiento de propiedades como porosidad, permeabilidad y espesor neto, lo cual puede realizarse con la contemplación de promedios aritméticos y utilización de pseudofunciones. Es inevitable tomar en cuenta que cuando se emplean los procesos de escalamiento de propiedades, algunos detalles en el proceso de flujo pueden perderse.

Inclusión de incertidumbres: Debe aceptarse que las incertidumbres involucradas en la estimación de propiedades del yacimiento, permiten predecir y cuantificar incertidumbres en el comportamiento futuro; la figura 1.10, muestra, que en lugar de predecir un solo comportamiento, se ejecuta un arreglo de suposiciones de comportamientos futuros, ajustando razonablemente el historial del comportamiento así como los datos de entrada disponibles. Investigando el rango de posibles comportamientos futuros, pueden representarse de mejor

manera las incertidumbres con respecto a la producción futura y evaluación económica.

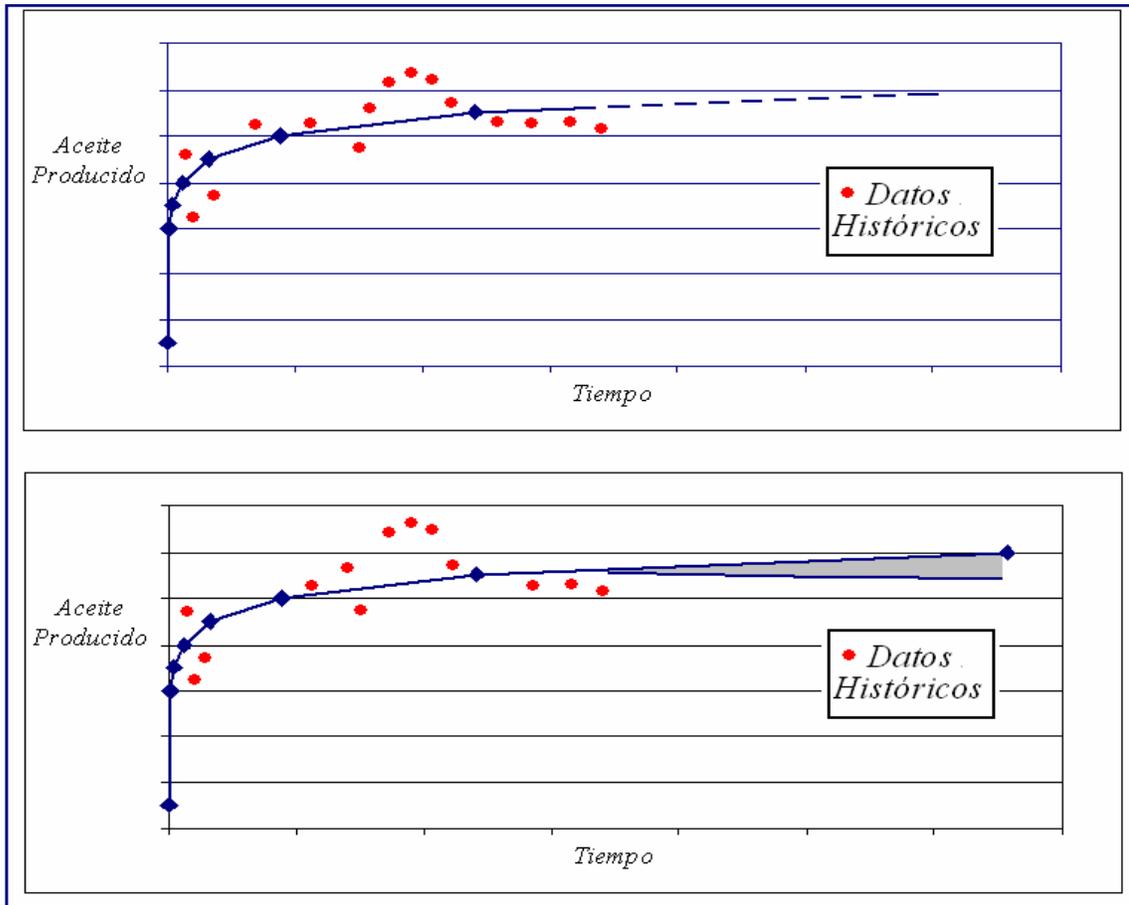


Figura 1.10. Predicción de un comportamiento futuro.

Representación de características geológicas: El proceso de construcción de un apropiado modelo es importante desde dos perspectivas. La de forma (estática), representa el ambiente geológico, tipos de roca y arquitectura del yacimiento. La otra es la de función (dinámica), la cual representa las propiedades petrofísicas resultado del comportamiento de flujo.

Muchos métodos convencionales crean descripciones del yacimiento que no respetan estrictamente la arquitectura geológica del mismo, enfatizando así, la función sobre la forma, por lo que una abundante cantidad de características claves pueden dejar de ser consideradas en la descripción final.

Una forma de superar este problema es mediante del modelado en paralelo, en donde los puntos de vista del Geólogo y alguna información cualitativa del modelo geológico es respetada.

1.5. USO DE LA GEOESTADÍSTICA EN EL MODELADO DE UN YACIMIENTO.

Cuando no se cuenta con suficiente información, en cualquier descripción de yacimientos se emplea el uso de herramientas estadísticas. Aun en simulaciones convencionales de yacimientos, se utilizan distintos esquemas de interpolación para la estimación de propiedades. Estos esquemas consideran relaciones implícitas entre las muestras de datos.

La Geoestadística, en un amplio sentido, puede definirse como una técnica estadística encargada de contabilizar las relaciones entre distintas variables con respecto a una distribución espacial la cual suele ser tridimensional preferentemente, por lo tanto, dicha técnica permite realizar estimaciones de valores de las variables, cuando no es posible obtener muestras en algunas localidades. Dichas variables incluyen propiedades petrofísicas convencionales como porosidad y permeabilidad; información sísmica, como valores de impedancia acústica; o variables geológicas, como arenas o lutitas. Las relaciones pueden incluir información referente a la manera en que los valores de una misma variable se relacionan con otras variables dentro de un área vecina, o información de cómo es que dos variables diferentes tengan la posibilidad de común dependencia en cuanto a espacio, es decir, que estas se relacionen entre si. En el capítulo dos este tema se aborda con mayor detalle.

CAPITULO 2

PRINCIPIOS DE LA GEOESTADÍSTICA

2.1. INTRODUCCIÓN.

Este capítulo, introduce los principios básicos de la Estadística convencional, necesarios para la comprensión de los procedimientos y análisis geoestadísticos que se tratarán en capítulos subsecuentes.

2.2. CONCEPTOS BÁSICOS DE ESTADÍSTICA.

La palabra "Estadística" suele utilizarse bajo significados distintos, a saber:

- a) *Como colección de datos numéricos:* Esto es el significado más común de la palabra Estadística. Se sobrentiende que dichos datos numéricos han de estar presentados de manera ordenada y sistemática. Una información numérica cualquiera, puede no constituir una estadística para merecer este apelativo, los datos deben de constituir un conjunto coherente, establecido de forma sistemática y siguiendo un criterio de ordenación, un ejemplo de este tipo de estadísticas es el anuario estadístico publicado por el Instituto Nacional de Estadística.
- b) *Como ciencia:* Puede definirse como un conjunto de métodos científicos ligados a la toma, organización, recopilación, presentación y análisis de datos¹, tanto para la deducción de conclusiones como para tomar decisiones razonables de acuerdo con tales análisis, en presencia de incertidumbre. En este significado, la Estadística estudia el comportamiento de los fenómenos de masas. Como todas las ciencias, busca las características generales de un colectivo y prescinde de las particulares de cada elemento².

La Estadística se divide principalmente en dos áreas: Estadística descriptiva la cual analiza las características de un grupo de datos sin realizar predicciones o inferencias a futuro y la Estadística inferencial o inductiva la cual realiza proyecciones e inferencias sobre los resultados de una recolección de

datos. En el Apéndice A se describen varios de los términos que se manejan en el ámbito de la Estadística para su mejor comprensión.

2.2.1. ESTADÍSTICA DESCRIPTIVA.

Es la parte de la Estadística que describe las variables aleatorias en las muestras mediante procedimientos empleados para organizar y presentar datos en una forma cómoda y útil que pueda comunicarse¹. Con los avances tecnológicos computacionales, el campo de la Estadística descriptiva se ha vuelto más poderoso; pueden ilustrarse tendencias generales, anomalías y variación en los datos ilustrados con colores contrastantes y mapas a escala.

2.2.1.1. DISTRIBUCIÓN DE FRECUENCIAS.

En la presentación de los datos tabulados de las variables cuantitativas se emplean intervalos de clase, que son la representación de categorías en donde se agrupan valores de la variable según un criterio de secuencia. Una frecuencia de clase, es el número de observaciones que corresponden a cada variable o a cada grupo de variables.

Normalmente el número de intervalos de clase depende de los puntos del muestreo de variables o datos.

2.2.1.1.1. VARIABLES CUANTITATIVAS DISCRETAS.

Las categorías vienen ordenadas de manera natural de menor a mayor, lo que permite introducir el concepto de distribución de frecuencias acumulativas. Para construirla basta ir sumando las frecuencias de cada clase con las de las clases anteriores: si se hace con las frecuencias absolutas, se obtiene la distribución de frecuencias absolutas acumulativas y si se hace con las frecuencias relativas, se tiene la distribución de frecuencias relativas

acumulativas. A continuación se muestra una tabla que ejemplifica los tipos de frecuencias para variables cuantitativas discretas:

Clase	Frecuencia de clase o Absoluta	Frecuencia Relativa	Frecuencia Absoluta Acumulativa	Frecuencia Relativa Acumulativa
c_i	f_i	$f_{Ri} = \frac{f_i}{n}$	$N_i = n_i$	$F_i = \frac{N_i}{n} = f_{Ri}$
c_j	f_j	$f_{Rj} = \frac{f_j}{n}$	$N_j = n_i + \dots + n_j$	$F_j = \frac{N_j}{n} = f_{Ri} + \dots + f_{Rj}$
c_k	f_k	$f_{Rk} = \frac{f_k}{n}$	$N_k = n$	$F_k = 1$
	$f_i + f_j + f_k = n$	$f_{Ri} + f_{Rj} + f_{Rk} = 1$		

Tabla 2.1. Frecuencias para variables cuantitativas discretas.

Donde:

$N_i = \sum_{i=1}^i n_i$	$N_k = \sum_{i=1}^k n_i = n$	$0 \leq N_i \leq 1$
$F_i = \sum_{i=1}^i f_i$	$F_k = \sum_{i=1}^k f_i = 1$	$0 \leq f_i \leq 1$

2.2.1.1.2. VARIABLES CUANTITATIVAS CONTINUAS.

El análisis de la distribución de frecuencias en las variables cuantitativas continuas es más complejo y tiene el interés, de que las categorías mediante las que se ordena la distribución no viene determinado por la variable, sino que es necesario elegirse. El primer paso para construir la tabla de la distribución de frecuencias es dividir el recorrido (conjunto de posibles valores de la variable) en clases o intervalos (preferentemente que no se solapen). Al punto central de cada uno de estos recorridos se le llama marca de clase y se representa por C_i

Intervalo	Clase	Frecuencia de clase o Absoluta	Frecuencia Relativa	Frecuencia Absoluta Acumulativa	Frecuencia Relativa Acumulativa
l_0, l_i	c_i	f_i	$f_{Ri} = \frac{f_i}{n}$	$N_i = n_i$	$F_i = \frac{N_i}{n} = f_{Ri}$
l_{j-1}, l_j	c_j	f_j	$f_{Rj} = \frac{f_j}{n}$	$N_j = n_i + \dots + n_j$	$F_j = \frac{N_j}{n} = f_{Ri} + \dots + f_{Rj}$
l_{k-1}, l_k	c_k	f_k	$f_{Rk} = \frac{f_k}{n}$	$N_k = n$	$F_k = 1$
		$f_i + f_j + f_k = n$	$f_{Ri} + f_{Rj} + f_{Rk} = 1$		

Tabla 2.2. Frecuencias para variables cuantitativas continuas.

La marca de clase queda fijada por $C_i = \frac{l_{i-1} + l_i}{2}$ donde l_{i-1} es el límite inferior del intervalo y l_i es el límite superior del intervalo, se le conoce como amplitud del intervalo a la cantidad de unidades del recorrido de la variable que contiene un intervalo.

Por ejemplo, los valores de cierta propiedad obtenida por algún registro geofísico representan una variable cuantitativa, en la cual se obtiene como resultado, los valores de esta propiedad a determinada profundidad. En la figura 2.1 se muestra un registro de porosidad.

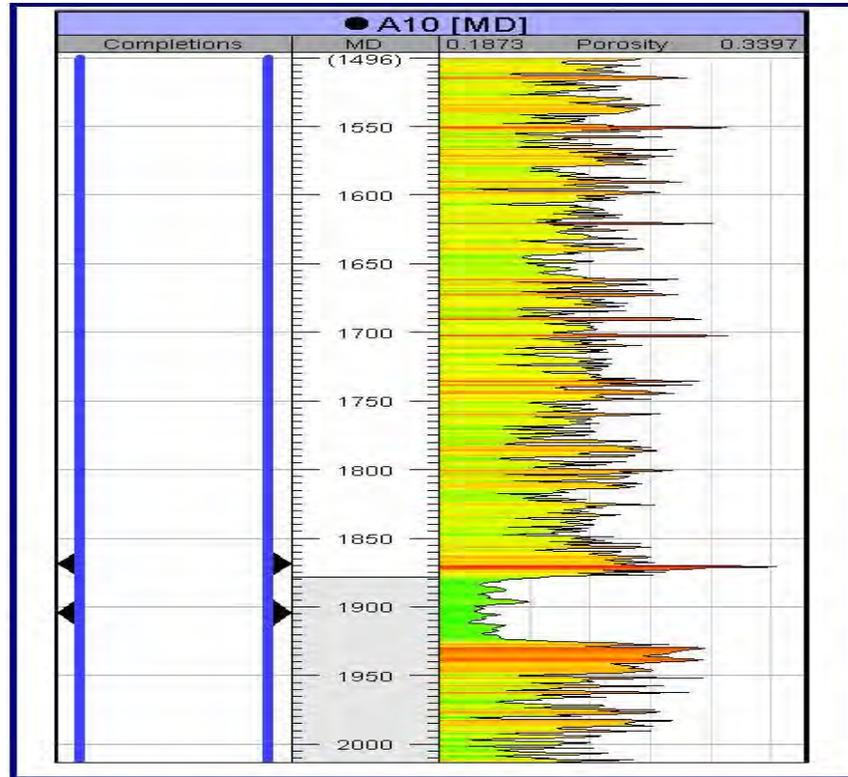


Figura 2.1. Ejemplo de una variable cuantitativa, registro de porosidad.

2.2.1.1.3. HISTOGRAMA.

El histograma es un gráfico para la distribución de una variable cuantitativa discreta y continua que representa frecuencias mediante el volumen de las áreas. Un histograma consiste en un conjunto de rectángulos con (a): bases en el eje horizontal, centros en las marcas de clase y longitudes iguales a los tamaños de los intervalos de clase y (b): áreas proporcionales a las frecuencias de clase³.

En el caso de un histograma para intervalos desiguales sólo se señalizan los valores sobre el eje horizontal, el eje vertical no tiene sentido porque las frecuencias corresponden al área de cada rectángulo.

Un histograma que represente los valores de porosidad de este tipo de registro con intervalos iguales se puede observar en la figura 2.2. En este histograma se presentan intervalos de 15 unidades. Esto puede variar dependiendo de las frecuencias en que se quieran presentar, dependiendo del análisis a realizar.

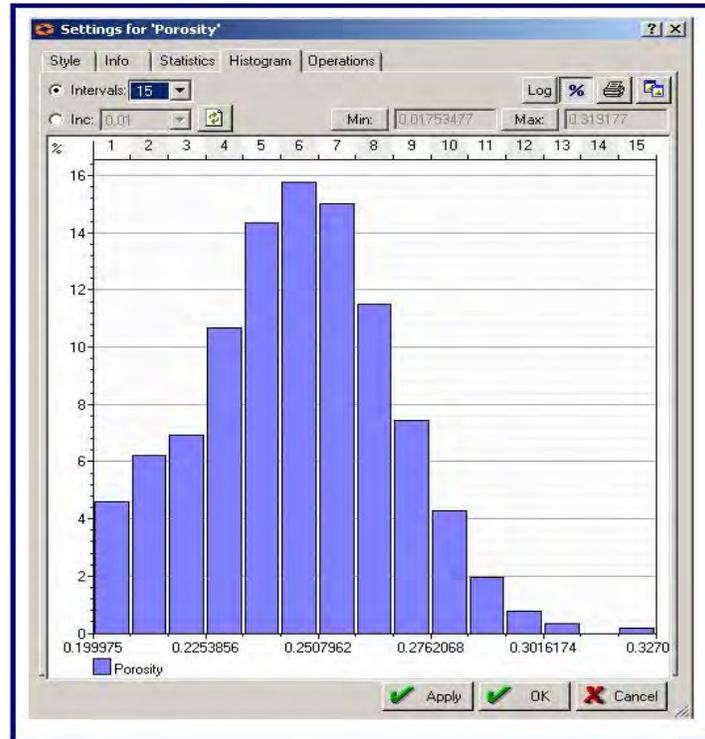


Figura 2.2. Histograma con intervalos iguales, para los valores de porosidad de la figura 2.1.

Si en la distribución se toman clases de la misma longitud, las frecuencias son proporcionales a las alturas de los rectángulos del histograma ya que el área se obtiene multiplicando la base por la altura por lo que queda similar a un diagrama de barras, solo que ahora las barras van una junto a otra por tratarse de una variable continua.

2.2.1.1.4. POLÍGONOS DE FRECUENCIA.

El polígono de frecuencias es una representación gráfica de la distribución de frecuencias que resulta esencialmente equivalente al histograma y se obtiene

uniendo mediante segmentos los centros de las bases superiores de los rectángulos del histograma (es decir, los puntos de las marcas de clase).

En la figura 2.3 se presenta un polígono de frecuencias en el que se observa que para cerrar la figura, se une la línea quebrada con lo que sería la marca de clase (sobre la superficie del eje horizontal) anterior a la primera y posterior a la última registrada.

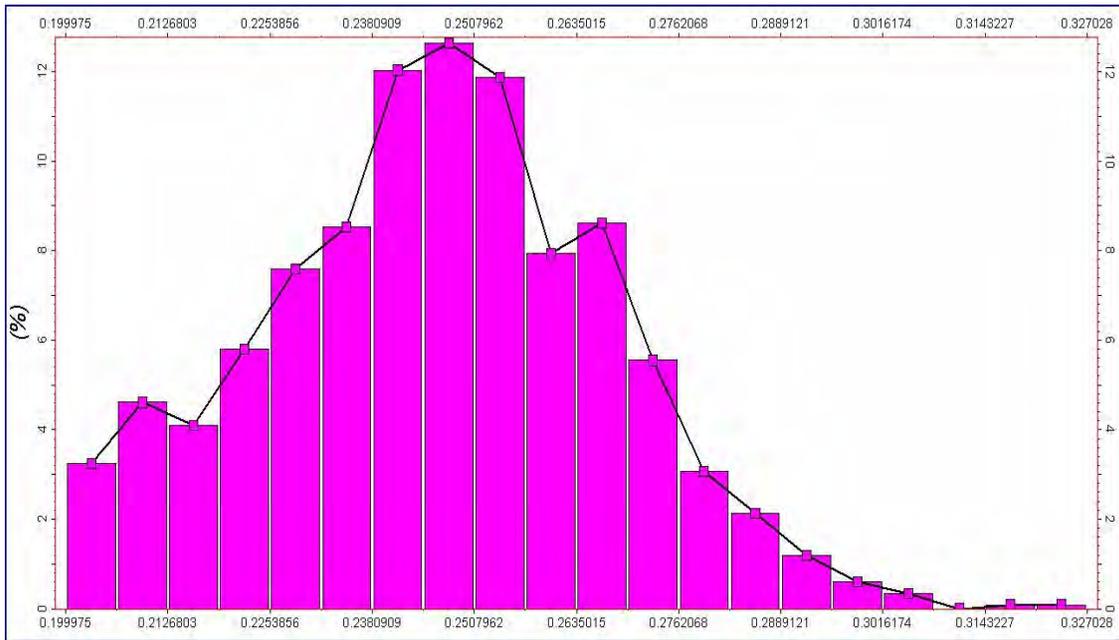


Figura 2.3. Polígono de frecuencias para los valores de porosidad de la figura 2.1.

2.2.1.1.5. OJIVA.

La ojiva tiene el concepto de unir los centros de las bases superiores de los rectángulos de la distribución del histograma de frecuencias acumulativas, dando lugar al polígono de frecuencias acumulativas u ojiva.

En la figura 2.4, se muestra un ejemplo de la ojiva, presentada con el histograma de intervalos iguales en el que presenta la frecuencia acumulativa de la porosidad del registro de la figura 2.1.

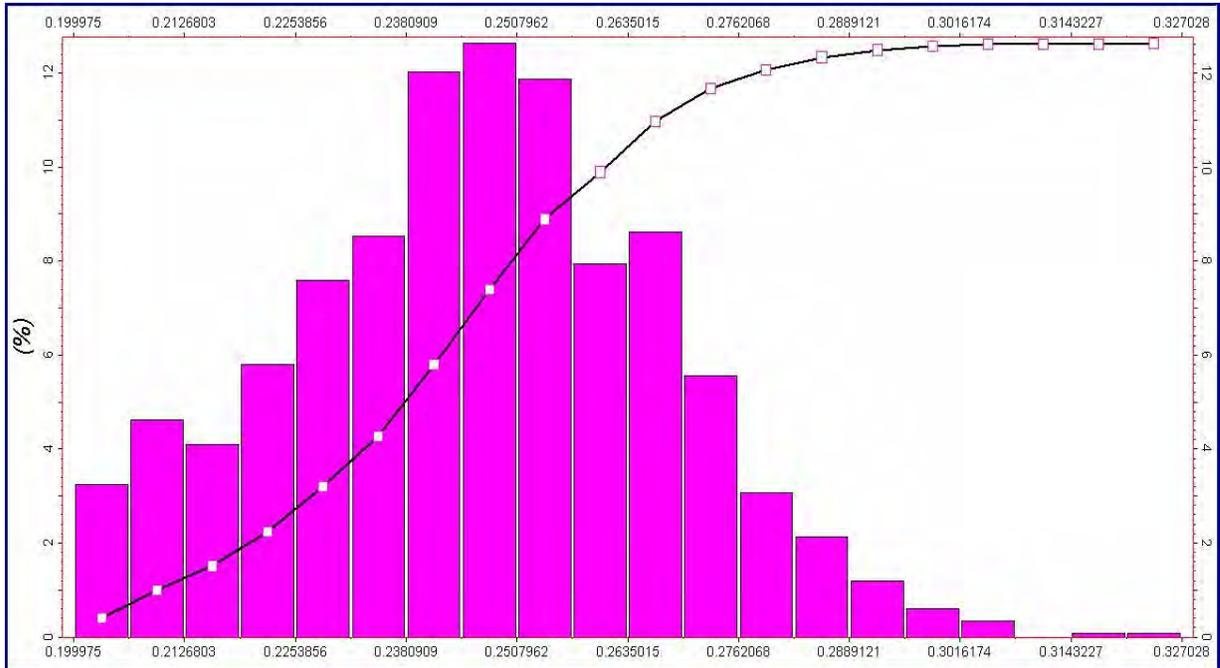


Figura 2.4. Ejemplo de una ojiva para los valores de porosidad de la figura 2.1.

2.2.1.2. ESTADÍSTICA PARA DISTRIBUCIONES UNIDIMENSIONALES.

En adición a la distribución de frecuencias, la cual caracteriza la muestra de manera visual y numérica, el cálculo de valores de la muestra que puedan presentar un valor es también benéfico. En el Apéndice B se detallan las opciones para presentar la muestra en un solo valor como medidas de posición y dispersión.

De un análisis de la variable, si todas las observaciones estuvieran concentradas en un solo valor de la variable, media, mediana y moda coincidirían en el mismo. Si las observaciones se fueran distribuyendo en forma

simétrica, a la izquierda y a la derecha de ese valor central, media, mediana y moda seguirían coincidiendo. Suponiendo que las observaciones de la parte izquierda se alejan del valor central más que las observaciones de la parte derecha, generando una distribución asimétrica hacia la izquierda; en este caso como la media es la suma de los valores de las observaciones dividido por la cantidad total de observaciones, su valor se correrá a la izquierda también y por el mismo motivo, la media será menor que la mediana y ambas menor que la moda. En una distribución asimétrica a la derecha, la media, es mayor que la mediana y que la moda.

Este corrimiento de la media se explica porque si se toma un conjunto de datos cualquiera a los cuales se le calcula media, mediana y moda, y se agrega un dato extremo y se vuelve a calcular la media, la mediana y la moda, se podrá ver que la media puede variar notablemente, mientras que la mediana y la moda permanecen idénticas. Esta falta de variación de la mediana y la moda reciben el nombre de robustez. Las medidas basadas en el orden, como la mediana, gozan de ésta, en tanto que las medidas basadas en la suma, como la media, se ven más afectadas por las observaciones extremas y son, por lo tanto, poco robustas.

Por ejemplo del registro de porosidad mostrado en la figura 2.1, se considera un conjunto de datos sin agrupar, cuyo cálculo de valores de medidas de posición se resume en la siguiente tabla.

<i>Medida de Posición</i>	<i>Considerando 1173 datos de 1173 totales</i>	<i>Considerando 1171 datos de 1173 totales</i>
Media(% ϕ)	0.250	0.244
Mediana(% ϕ)	0.247	0.247
Moda(% ϕ)	0.253	0.253

Tabla 2.3. Medidas de posición para los datos del registro de porosidad de la figura 2.1.

En la tabla anterior se observa un insignificante cambio en la media al dejar de contemplar los datos extremos (inicial y último), mientras que la mediana y moda permanecen idénticas. Es importante recalcar que el conjunto de valores del registro de porosidad varía aproximadamente entre 0.20 y 0.30, por lo que en realidad no se nota una gran transición en la media, si existiera un rango de valores del registro más amplio, la diferencia entre los valores de la media sería más grande. Es de notarse la simetría en la distribución debido a la similitud de los valores de la media y la mediana.

Adicionalmente, del registro de porosidad mostrado en la figura 2.1, se considera al mismo conjunto de datos sin agrupar, cuyo cálculo de valores de medidas de dispersión se resume en la siguiente tabla.

Medida de Dispersión	Considerando 1173 datos de 1173 totales
Desviación Estándar(% ϕ)	0.021339187
Varianza(% ϕ^2)	0.000455361
Coefficiente de Variación(ad)	0.294541563
Coefficiente de Asimetría(ad)	-0.202098148
Coefficiente de Apuntamiento (ad)	0.07799572

Tabla 2.4. Medidas de dispersión para los datos del registro de porosidad de la figura 2.1.

Al analizar los valores de desviación estándar, varianza y coeficiente de variación, se observa poca dispersión por lo que existe mayor homogeneidad en la concentración de datos, con lo que respecta al coeficiente de asimetría, el sesgo de la distribución se inclina hacia la derecha, y finalmente respecto al coeficiente de apuntamiento, la distribución de datos puede clasificarse como platicúrtica.

Para mayores detalles puede consultarse el Apéndice B.

2.2.1.3. DISTRIBUCIÓN CONDICIONAL DE FRECUENCIA.

El concepto de distribución condicional se utiliza para representar la distribución de frecuencia de dos variables. Una distribución condicional, es la distribución de una variable que es condicional a la distribución de otra variable.

En ocasiones no existen suficientes muestras para generar la distribución condicional de una variable, dado un valor de la otra. Como alternativa, es posible considerar un rango de valores para una variable, y la distribución condicional de la otra variable para dicho rango. Conforme el rango se vuelve pequeño, la distribución condicional se acerca a la distribución correspondiente al valor en un punto.

El tipo de información condicional es útil en la determinación de incertidumbres de una variable con respecto al valor de otra variable.

2.2.1.4. ESTADÍSTICA PARA DISTRIBUCIONES BIDIMENSIONALES.

Para esta sección el análisis estadístico esta enfocado a dos variables, en el cual se describen los conceptos de covarianza, el coeficiente de correlación y la regresión lineal. En la figura 2.5 se muestra un ejemplo de distribución de dos variables, como lo es una grafica de porosidad vs permeabilidad.

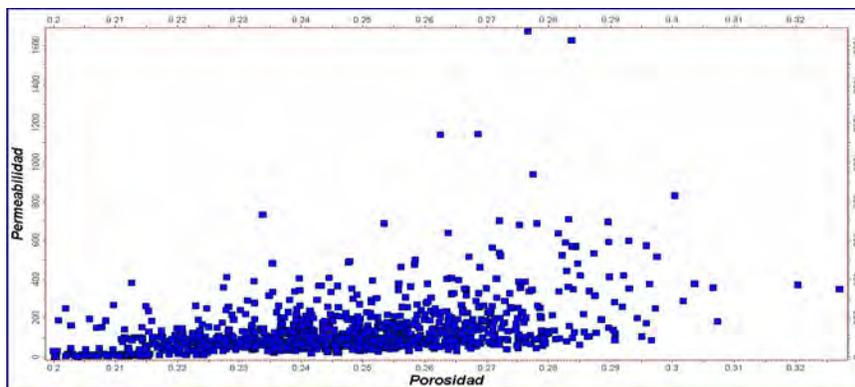


Figura 2.5. Distribución de dos variables.

2.2.1.4.1. COVARIANZA PARA EL ANÁLISIS ESTADÍSTICO.

La covarianza es un indicador de la relación existente entre dos variables.

$$c(x, y) = \frac{1}{n} \sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \frac{1}{n} \sum_{i=1}^n y_i \dots\dots\dots(2.1)$$

Si los valores de x y y son relacionados positivamente, por ejemplo, si ambos incrementan, la covarianza tendrá un valor positivo; si los valores de x y y son relacionados negativamente, por ejemplo x incrementa y y disminuye, la covarianza tendrá un valor negativo; de la misma manera si estos no tienen relación la covarianza tiene un valor muy cercano a cero.

2.2.1.4.2. COEFICIENTE DE CORRELACIÓN PARA EL ANÁLISIS ESTADÍSTICO.

El coeficiente de correlación cuantifica la relación entre dos variables, Similar a la covarianza, es una cantidad adimensional y debe tomar valores entre -1 y +1, se representa mediante la expresión:

$$r(x, y) = \frac{c(x, y)}{s_x s_y} \dots\dots\dots(2.2)$$

Donde s_x es la desviación estándar de la variable x , y s_y es la desviación estándar de la variable y .

Por ejemplo la covarianza y el coeficiente de correlación de la relación de porosidad vs permeabilidad que se muestra en la figura 2.5 son respectivamente:

$$c(x, y) = 0.65272 \quad y \quad r(x, y) = 0.222427$$

2.2.1.4.3. REGRESIÓN LINEAL.

Una extensión lógica para la aplicación de los principios estadísticos, es el establecer una relación lineal entre dos variables. Una relación lineal es de gran ayuda para predecir el valor de una variable cuando el valor de la otra es conocido. La forma más simple para establecer este tipo de relaciones, es mediante la ecuación

$$y = mx + b \dots\dots\dots(2.3)$$

Donde y es la variable a estimar, x es la variable conocida, m es la pendiente de la línea recta, y b es la ordenada al origen o intersección con el eje y .

Es posible obtener los valores de m y $c(x, y)$; primero se utiliza un par de muestras x, y disponibles, para posteriormente obtener el mejor ajuste entre las dos variables. El mejor ajuste se puede obtener con

$$m = \frac{c(x, y)}{s_x^2} \dots\dots\dots(2.4a)$$

Y

$$b = \bar{y} - m\bar{x} \dots\dots\dots(2.4b)$$

Donde $c(x, y)$ es la covarianza entre x e y , s_x^2 es la varianza de x , e \bar{y} y \bar{x} , son las medias aritméticas de las variables x, y respectivamente.

Por ejemplo, en la figura 2.6 se muestra la regresión lineal que existe entre los datos de porosidad y permeabilidad que aparecen en la figura 2.5. De esta relación se tiene que el valor de m es de 2784.95 y el valor de b es de - 550.44, por lo tanto la ecuación lineal que relaciona esta distribución es de la siguiente forma:

$$k = 2784.95\phi - 550.44$$

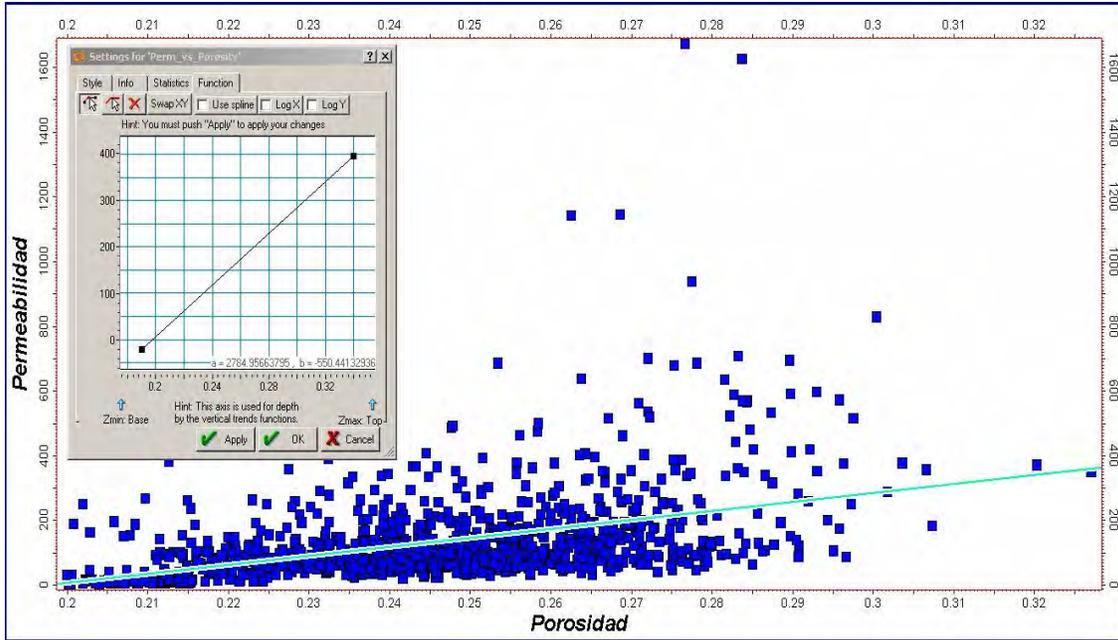


Figura 2.6. Regresión lineal de porosidad vs permeabilidad de la figura 2.5.

2.2.2. ESTADÍSTICA INFERENCIAL.

La Estadística inferencial consiste en conseguir información de una muestra aleatoria en una población mediante la estimación aproximada del verdadero valor en la misma, de acuerdo a la característica de interés. El método que utiliza la inferencia Estadística es inductivo, ya que a partir de resultados particulares en la muestra, se obtienen conclusiones generales para la población^{4, 5}.

2.2.2.1. MUESTRA ALEATORIA.

Sea una población con la característica medible X cuya distribución es $f(x)$. Se dice que

$X = \{x_1, x_2, x_3, \dots, x_n\}$ es una muestra aleatoria simple si:

- a) Cada valor x_i se obtiene observando X de manera independiente bajo las mismas condiciones n veces, es decir, las x_i son variables aleatorias independientes (lo que significa que se ha efectuado un muestreo aleatorio con reemplazo).
- b) Cada observación x_i tiene distribución de probabilidad idéntica a la distribución de la población:

$$f(x_i) = f(x), \forall i = 1, 2, \dots, n$$

Las observaciones x_i son realizaciones de las variables aleatorias X_i , cada una con distribución $f(x_i)$. Si se cumplen las dos condiciones anteriores, entonces se dice que $X_1, X_2, X_3, \dots, X_n$ son variables aleatorias independientes e idénticamente distribuidas.

2.2.2.2. ESTADÍSTICO.

Un estadístico (o estadística) es cualquier función de las observaciones de una muestra aleatoria $X = \{x_1, x_2, x_3, \dots, x_n\}$.

Un estadístico no debe contener cantidades desconocidas o que no se puedan determinar.

Dos ejemplos de estadísticos son la media \bar{x} y la varianza s^2 , que son funciones de las observaciones x_i obtenidas de una muestra aleatoria de tamaño n .

Algunos estadísticos se utilizan para estimar a los parámetros de una población; es decir, se infiere el valor del parámetro poblacional a partir del valor del estadístico, entonces el estadístico se denomina estimador del parámetro⁶.

Generalmente suele representarse a los parámetros de la población con letras griegas minúsculas (θ) y a los estadísticos con letras latinas mayúsculas (T); a los valores particulares que toma un estadístico después de efectuar un muestreo se les representa con letras latinas minúsculas (t). Al estimador de un parámetro se le representa con letras griegas mayúsculas ($\hat{\Theta}$); el valor puntual que adquiere el estimador al efectuar un muestreo se le denomina estimación y se le representa con letras griegas minúsculas ($\hat{\theta}$).

La notación mencionada es frecuente, aunque no es regla general, en este capítulo se utilizará así con algunas excepciones debido a que se usará la notación más generalizada.

2.2.2.3. ESPACIO MUESTRAL.

Es el conjunto de todos los posibles resultados de un experimento; constituye el conjunto universal de todas las observaciones. Se suele denotar por la letra S ⁷.

En general, un evento es un fenómeno aleatorio, es decir, aquel cuyo resultado depende del azar. En la práctica, si bien un evento puede denotar un conjunto total de resultados de experimentos, para el cual se utiliza la notación de conjuntos denominándolo por una letra mayúscula, en ocasiones se denota también un sólo resultado de un fenómeno o experimento, o sea, un elemento de un conjunto⁸.

2.2.3. PROBABILIDAD: ENFOQUE FRECUENTISTA.

Cuando se tiene un experimento E que se repite una gran cantidad de veces, la frecuencia relativa de un experimento podría ser prácticamente igual a un valor p , con un elevado grado de certeza.

Siguiendo este tipo de razonamiento se construye un modelo matemático ideal y abstracto de este experimento que se postula como sigue: dado un experimento E y un evento A , se puede asignar un número p al evento A , el cual se denomina probabilidad del evento A . Ese número p tiene las siguientes características: Cuando el experimento E se repite una gran cantidad de veces (n_C) y el experimento sucede (n_A) veces, la frecuencia relativa $\frac{n_A}{n_C}$ será prácticamente igual a ese número p ⁶.

El número p , que se llama probabilidad del evento A , se denota por $p(A)$.

La $p(A)$ cumple con: $0 \leq \frac{n_A}{n_C} \leq 1$ lo que equivale a $0 \leq p(A) \leq 1$

Puede percibirse que el hecho de que $p(A) = 0$ no asegura que el evento A sea imposible. De la misma forma, $p(A) = 1$ no asegura la ocurrencia cierta del evento A .

Este enfoque posee cuatro características:

- 1) Supone una gran cantidad de ensayos
- 2) Supone la regularidad estadística
- 3) La $p(A)$ se estima por la frecuencia relativa de A
- 4) Está basada en la experiencia

Este enfoque es el principio en el cual se fundamentan los estudios probabilísticos desarrollados en los años cincuenta, principalmente en Inglaterra y los Estados Unidos. No obstante, este enfoque presenta limitaciones, particularmente en lo que a sus valores extremos se refiere, frente a la necesidad de evaluar experimentos que no se producen realmente o no se pueden repetir⁸.

Una corrección a este enfoque, citado en la literatura como segundo enfoque frecuentista, define a la probabilidad como el límite de $\frac{n_A}{n_C}$ cuando n tiende a infinito:

$$p(A) = \lim_{n \rightarrow \infty} \frac{n_A}{n_C} \dots\dots\dots(2.5)$$

2.2.3.1. PROBABILIDAD CONDICIONAL.

Si A y B son dos eventos del espacio muestral S , la probabilidad de que el evento B ocurra con la condición de que previamente haya ocurrido el evento A , se le conoce como probabilidad condicional de B y se representa como $p(B | A)$ a partir de los eventos A y B .

Al realizar el experimento correspondiente, supóngase que ocurre el evento A y una vez ocurrido éste, se desea observar si también ocurrió el B . Puesto que ya ocurrió el evento A , el evento B sólo puede ocurrir si tiene eventos elementales en A .

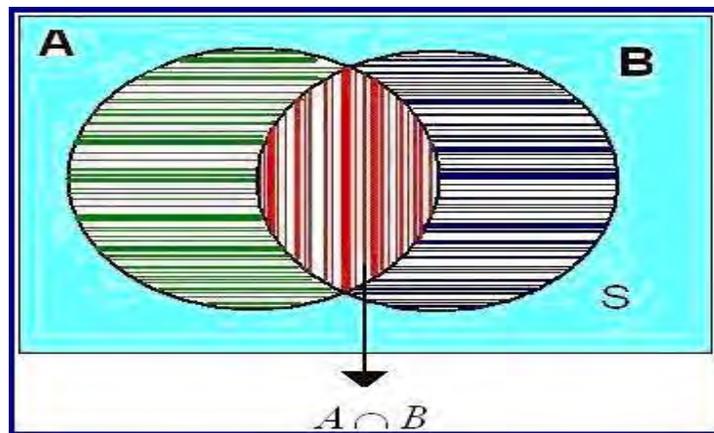


Figura 2.7. Ejemplificación de la operación $A \cap B$.

Además, en ese momento el evento A es el espacio total de eventos posibles.

Luego $p(B | A)$ se puede calcular como

$$p(B | A) = \frac{p(A \cap B)}{p(A)} \dots\dots\dots(2.6)$$

Si $p(A) \neq 0$

Y también

$$p(A | B) = \frac{p(A \cap B)}{p(B)} \dots\dots\dots(2.7)$$

Si $p(B) \neq 0$

La secuencia en que ocurren los eventos puede mostrarse con claridad a través de un diagrama de árbol, siempre y cuando se trate de eventos mutuamente excluyentes, colectivamente exhaustivos con un número finito de resultados.

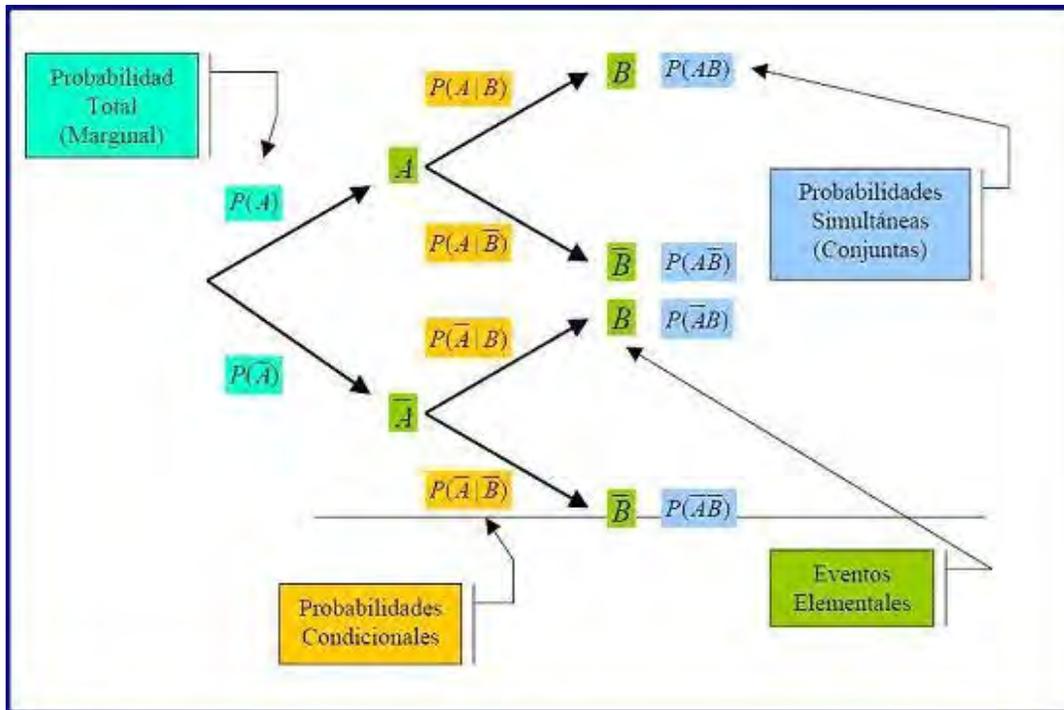


Figura 2.8. Diagrama de árbol⁶.

A partir de la información del experimento a estudiar, primeramente deben especificarse los eventos, posteriormente la secuencia o subordinación entre ellos.

En la generalidad, las probabilidades condicionales se advierten porque los eventos se subordinan con la conjunción *si* (condicional, de diferente significado que el *sí* afirmativo)⁵. De la misma forma, suele detectarse en la definición del experimento que la probabilidad condicional es un dato histórico, tal que cuando es mencionada aún no se produce el evento.

Asimismo, la probabilidad simultánea corresponde a dos eventos que ocurren al mismo tiempo en el instante en que se estudia el experimento, y la subordinación entre ellos se detecta a través de la conjunción *y*.

2.2.4. VARIABLE ALEATORIA.

Es posible incorporar el uso de herramientas matemáticas al cálculo de probabilidades a través de realizar una operación consistente cuyo ejercicio sea buscar una regla de correspondencia que proyecte los resultados de un evento o experimento en la recta de los números reales.

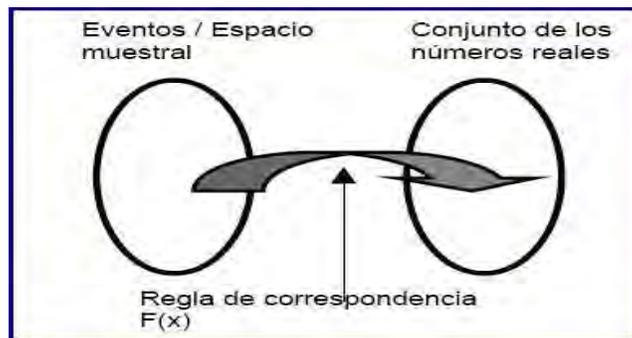


Figura 2.9. Regla de correspondencia: proyección de los resultados de un evento o experimento en la recta de los números reales⁶.

Al asociar los resultados aleatorios a puntos sobre la recta de los reales, es decir, a la variable independiente, esta variable adquiere un carácter aleatorio. De tal forma, se le denomina variable aleatoria.

De acuerdo a la naturaleza del evento, existen dos tipos de variables aleatorias:

- Se dice que una variable aleatoria X es discreta si el número de valores que puede tomar es contable (ya sea finito e infinito) y si estos pueden arreglarse en una secuencia que corresponde con los enteros positivos
- Se dice que una variable aleatoria X es continua si sus valores consisten en uno o más intervalos de la recta de los reales

A la regla de correspondencia $f(x)$ se le denomina función densidad de probabilidad, siempre y cuando cumpla con una condición que se describirá más adelante.

El proceso integral de creación de una variable aleatoria y su respectiva función de probabilidad puede describirse en el siguiente diagrama de bloques:

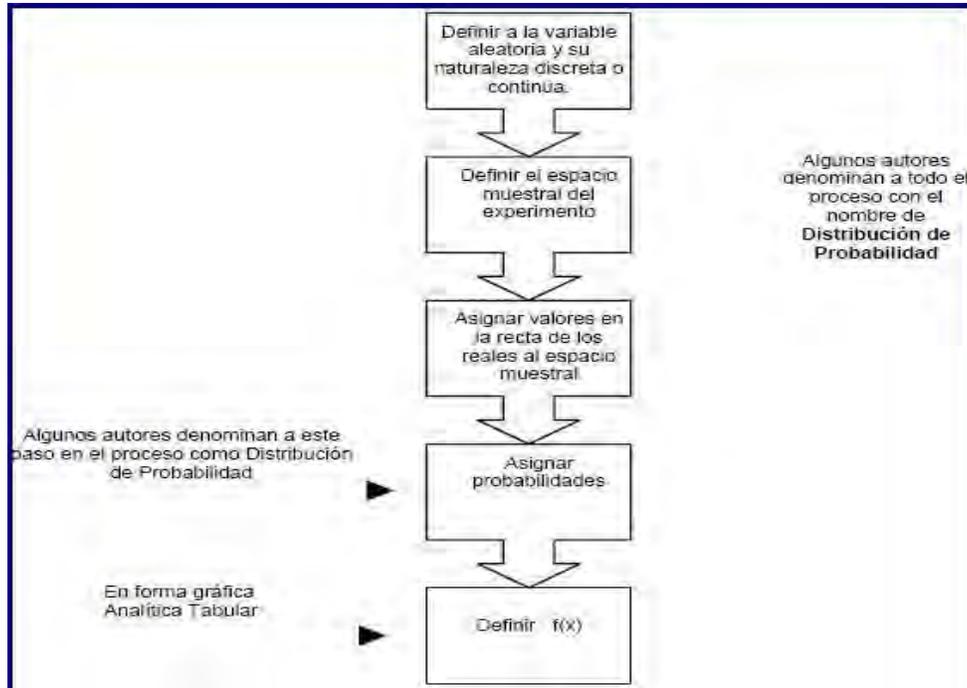


Figura 2.10. El proceso integral de creación de una variable aleatoria⁶.

Como comentario a este procedimiento, la asignación de valores en la recta de los reales para cada punto del experimento o del espacio muestral se hace de acuerdo a la experiencia del diseñador del experimento, no existe una regla escrita para hacerlo. No obstante se recomienda asignarle valores que sean representativos del experimento en ánimo de clasificar el mismo.

2.2.4.1. VARIABLE ALEATORIA DISCRETA.

Sea S un espacio muestral sobre el que se encuentra definida una función de probabilidad. Sea X una función de valor real, definida sobre S de manera que transforme los resultados de S en un punto sobre la recta de los reales, se dice entonces que X es una variable aleatoria (en adelante *v.a*).

En general una variable aleatoria discreta (en adelante *v.a.d*) representa los resultados de un espacio muestral en forma tal que:

$$p(x_i) = P(X = x_i) \dots\dots\dots(2.8)$$

La ecuación (2.8) se entenderá como la probabilidad de que X tome el valor de x_i , de esta forma, al considerar los valores de una *v.a.d* es posible desarrollar una función matemática que asigne una probabilidad a cada realización x_i de la *v.a* X , la cual recibe el nombre de función masa de probabilidad de la *v.a* X .

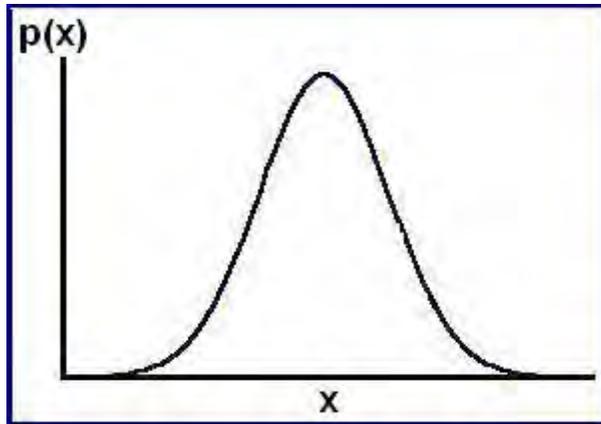


Figura 2.11. Función masa de probabilidad.

Por axiomas de probabilidad se tiene:

$$1) P(X = x_i) \geq 0 \dots\dots\dots(2.9)$$

$$2) \sum_{i=1}^n p(x_i) = 1 \dots\dots\dots(2.10)$$

Distribución de Probabilidad.

De acuerdo a la literatura, se le denomina distribución de probabilidad, a la acción de asignar la probabilidad a cada uno de los valores que puede adquirir la *v.a* ^{5,8}.

2.2.4.2. VARIABLE ALEATORIA CONTINUA.

Dada la naturaleza continua del espacio muestral, resulta imposible determinar el valor preciso que exactamente toma una variable; por este motivo, para fenómenos continuos resulta más apropiado visualizar probabilidades en intervalos.

En el caso discreto se asignan probabilidades positivas a todos los valores puntuales de la *v.a* , pero la suma de todas ellas es 1, aún a pesar de que el conjunto de valores sea infinito contable.

Para el caso continuo lo anterior no es posible, por esta razón la probabilidad de que una variable aleatoria continua (en adelante *v.a.c*) *X* tome un valor específico *x* es cero (0).

La distribución de probabilidad de una *v.a.c* *X* está caracterizada por una función, llamada función densidad de probabilidad:

$$\int_{-\infty}^{\infty} f(x)dx = 1 \dots\dots\dots(2.11)$$

Esta función no es la misma que para el caso discreto.

Dado que existe la probabilidad de que *X* tome el valor específico *x* = 0 , esta función determina la probabilidad en un intervalo.

$$P(a_1 \leq X \leq b_1) = \int_{a_1}^{b_1} f(x)dx \dots\dots\dots(2.12)$$

Para cualquiera *a*₁ y *b*₁ .

Puesto que el área total bajo $f(x)$ es 1, la probabilidad del intervalo $a_1 \leq X \leq b_1$, es el área acotada por la función densidad y las rectas $x = a_1$ y $x = b_1$.

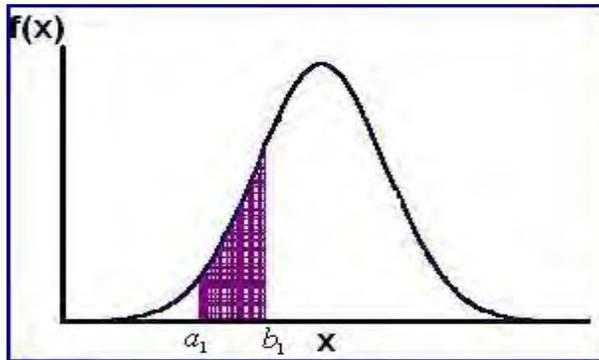


Figura 2.12. Función densidad de probabilidad: Probabilidad de que un valor se aloje dentro de un intervalo (a_1, b_1) .

2.2.4.3. FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA.

Se define a la función de distribución acumulativa de la *v.a* como la probabilidad de que X sea menor o igual a un valor específico de x^5 .

$F(x)$ es la función de distribución acumulativa y

$$F(x) = P(X \leq x) \dots\dots\dots(2.13)$$

Conociendo la definición de función densidad de probabilidad para un intervalo dado, ecuación (2.12), es posible calcular la probabilidad de que una *v.a* se situé dentro de un intervalo. Por ejemplo, para calcular $P(a_1 \leq X \leq b_1)$, se puede escribir $P(X \leq b_1)$ como la inclusión de dos eventos mutuamente excluyentes o

$$P(a_1 \leq X \leq b_1) = P(X \leq a_1) + P(a_1 < X \leq b_1) \dots\dots\dots(2.14a)$$

Y así

$$P(a_1 \leq X \leq b_1) = P(X \leq b_1) - P(X \leq a_1) = F(b_1) - F(a_1) \dots\dots\dots(2.14b)$$

Para una *v.a.d* , la función de distribución acumulativa puede calcularse con la expresión

$$F(a_1) = \sum_{x_i \leq a_1} p(x_i) \dots\dots\dots(2.15)$$

Donde $p(x_i)$ es la función masa de probabilidad de una *v.a* .

Para una *v.a.c* , la función de distribución acumulativa puede calcularse con la expresión:

$$F(a_1) = \int_{-\infty}^{a_1} f(x)dx \dots\dots\dots(2.16)$$

Donde $f(x)$ es la función densidad de probabilidad de una *v.a* .

Es importante comprender que la función de distribución acumulativa esta fuertemente relacionada con la distribución de frecuencia relativa acumulativa. De manera similar a la distribución de frecuencia relativa acumulativa, la función de distribución acumulativa tiene un valor mínimo igual a cero y un valor máximo igual a uno.

2.2.4.4. VALOR ESPERADO DE UNA VARIABLE ALEATORIA.

El valor esperado nace de la práctica de los juegos de azar. El valor esperado (o esperanza) representa para los jugadores la cantidad que ganarán o perderán, después de jugar repetidamente cierto juego.

En general, es posible obtener el valor esperado de una función $g(x)$ de la v.a X

$$E[g(X)] = E[g(x_i)] = \sum_{i=1}^o g(x_i)p(x_i) \quad \text{para una v.a.d} \dots\dots\dots(2.17)$$

Donde $p(x_i) = P[X = x_i]$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x)f(x)dx \quad \text{para una v.a.c} \dots\dots\dots(2.18)$$

Donde $g(x)$ es una función de la variable x .

Para una variable constante

$$E[K] = K \dots\dots\dots(2.19)$$

Es necesario aclarar que el valor esperado de una constante es la constante misma, y es una propiedad de la distribución de probabilidad de la v.a X .

2.2.4.5. MOMENTOS DE LA VARIABLE ALEATORIA.

Son los valores esperados de ciertas funciones de x que forman una colección de medidas descriptivas las cuales pueden emplearse para caracterizar la distribución de probabilidad de x y especificar si todos los momentos de x son conocidos.

A pesar de que los momentos de x pueden definirse alrededor de cualquier punto de referencia, generalmente se definen alrededor del cero o del valor esperado de x .

Los dos valores esperados más importantes para una variable, son la media aritmética y la varianza. La media aritmética se define como

$$\mu = E[X] = \sum_{i=1}^o x_i p(x_i) \quad \text{para una v.a.d} \dots\dots\dots(2.20)$$

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx \quad \text{para una v.a.c} \dots\dots\dots(2.21)$$

μ es el valor esperado de su propia variable. La media aritmética es la media poblacional, no la media muestral; en el presente capítulo, la media muestral se representa mediante \bar{x} y la media poblacional se representa con el símbolo μ .

La varianza se define como

$$\sigma^2 = V[X] = E[(X - \mu)^2] \dots\dots\dots(2.22a)$$

La notación σ^2 es diferente de s^2 (ver apéndice B), la cual representa la varianza muestral. La raíz cuadrada de σ^2 , σ es la desviación estándar. De acuerdo a las características de los valores esperados, se puede mostrar lo siguiente:

$$V[X] = E[X^2] - \{E[X]\}^2 \dots\dots\dots(2.22b)$$

o

$$V[X] = E[X^2] - \mu^2 \dots\dots\dots(2.22c)$$

La varianza de una *v.a* es una medida de dispersión de la distribución de probabilidades de ésta⁷. Por ejemplo, para el caso continuo, si la mayor parte del área por debajo de la curva de distribución se encuentra cercana a la media, la varianza es pequeña; si la mayor parte del área se encuentra muy dispersa al rededor de la media, la varianza es grande.

Algunas características importantes de la varianza son:

$$V[K] = 0 \dots\dots\dots(2.23)$$

Donde K es una constante;

$$V[KX] = K^2V[X] \dots\dots\dots(2.24)$$

2.2.4.6. COEFICIENTE DE VARIACIÓN.

Si se desea comparar la varianza con la desviación estándar de varias variables con diferentes unidades entre sí, la comparación resultará injusta debido a que las distribuciones de variables aleatorias con unidades pequeñas tendrán aparentemente menos variación. Para hacer este tipo de comparaciones se define otro indicador denominado coeficiente de variación, que es el cociente de la desviación estándar entre la media, y tiene las mismas unidades que el coeficiente de variación poblacional por lo tanto es adimensional

$$Cv = \frac{\sigma}{\mu} \dots\dots\dots(2.25)$$

2.2.4.7. ERROR ESTÁNDAR.

El error estándar describe el grado de conocimiento de los datos y se puede calcular por:

$$\varepsilon = \sqrt{\sigma^2 / n} \dots\dots\dots(2.26)$$

2.2.5. VARIABLES ALEATORIAS BIDIMENSIONALES.

Es posible extender las definiciones de funciones de probabilidad para dos variables.

2.2.5.1. VARIABLES ALEATORIAS BIDIMENSIONALES DISCRETAS.

En general una variable aleatoria bidimensional discreta (*v.a.b.d*) representa los resultados de un espacio muestral en forma tal que

$$p(x_i, y_j) = P(X = x_i, Y = y_j) \dots\dots\dots(2.27)$$

Se entenderá como la probabilidad de que X tome el valor de x_i e Y tome el valor de y_j , de esta forma, al considerar los valores de una *v.a.b* es posible desarrollar una función matemática que asigne una probabilidad a cada realización x_i de la *v.a* X , e y_j de la *v.a* Y , la cual recibe el nombre de función masa de probabilidad conjunta de la variables aleatorias X, Y .

De manera similar para las variables aleatorias se tiene que

$$1) P[X = x_i, Y = y_j] \geq 0 \dots\dots\dots(2.28)$$

$$2) \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} p(x_i, y_j) = 1 \dots\dots\dots(2.29)$$

Donde, n_x es el número de posibles eventos de la *v.a* X , y n_y es el número de posibles eventos de la *v.a* Y .

Se llama distribución marginal cuando se considera solo la distribución de una variable mientras se ignora el efecto de la otra, por lo que la distribución marginal de una *v.a* X puede escribirse como

$$P_X(x_i) = \sum_{j=1}^{n_y} p(x_i, y_j) \dots\dots\dots(2.30)$$

Y los posibles valores que adquiera Y se toman en cuenta, mientras el valor de X es siempre el mismo. Así es como también puede escribirse la distribución marginal de una *v.a* Y

$$P_Y(y_j) = \sum_{i=1}^{n_x} p(x_i, y_j) \dots\dots\dots(2.31)$$

2.2.5.2. VARIABLES ALEATORIAS BIDIMENSIONALES CONTINUAS.

La distribución de probabilidad de una variable aleatoria bidimensional continua (*v.a.b.c*) X, Y está caracterizada por la función densidad de probabilidad conjunta

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx dy = 1 \dots\dots\dots(2.32)$$

Dado que existe la probabilidad de que X, Y tome el valor específico $x = 0, y = 0$, esta función determina la probabilidad en un intervalo para cada *v.a*

$$P(a_1 \leq X \leq b_1, c_1 \leq Y \leq d_1) = \int_{a_1}^{b_1} \int_{c_1}^{d_1} f(x, y) dx dy \dots\dots\dots(2.33)$$

Para cualquiera a_1 y b_1, c_1 y d_1 .

Como en el caso de distribuciones discretas, las distribuciones marginales para cada *v.a* X, Y pueden calcularse como

$$f_x(x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dy \dots\dots\dots(2.34)$$

$$f_y(y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dx \dots\dots\dots(2.35)$$

2.2.5.3. FUNCIÓN DE DISTRIBUCIÓN ACUMULATIVA CONJUNTA.

La función de distribución acumulativa conjunta puede escribirse como

$$F(x, y) = P(X \leq x, Y \leq y) \dots\dots\dots(2.36)$$

Para una *v.a.b.d*, la función de distribución acumulativa conjunta puede calcularse con la expresión

$$F(a_1, b_1) = \sum_{x_i \leq a} \sum_{y_j \leq b} p(x_i, y_j) \dots \dots \dots (2.37)$$

Para una *v.a.b.c*, la función de distribución acumulativa conjunta puede calcularse con la expresión

$$F(a_1, b_1) = \int_{-\infty}^{a_1} \int_{-\infty}^{b_1} f(x, y) dx dy \dots \dots \dots (2.38)$$

2.2.5.4. VALORES ESPERADOS PARA FUNCIONES CON DOS VARIABLES ALEATORIAS.

La relación del valor esperado puede extenderse fácilmente para distribuciones bidimensionales

$$E[g(X, Y)] = E[g(x_i, y_j)] = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} g(x_i, y_j) p(x_i, y_j) \quad \text{para cada } v.a.b.d \dots \dots \dots (2.39)$$

Donde

$$p(x_i, y_j) = P[X = x_i, Y = y_j]$$

$E[g(X, Y)]$ es el valor esperado de la función $g(x_i, y_j)$, $p(x_i, y_j)$ es la función masa de probabilidad de las variables X e Y

$$E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f(x, y) dx dy \quad \text{para cada } v.a.b.c \dots \dots \dots (2.40)$$

$g(x, y)$ es una función de las variables aleatorias X e Y , y $f(x, y)$, es la función masa de probabilidad de las dos variables X e Y .

2.2.5.5. COVARIANZA.

La covarianza es uno de los valores esperados más importantes para distribuciones con dos variables, entre mayor sea la relación, mayor será el valor de la covarianza; esta se define como

$$C[X, Y] = E[(X - E[X])(Y - E[Y])] = E[(X - \mu_x)(Y - \mu_y)] \dots\dots\dots(2.41a)$$

Empleando las propiedades de valores esperados se puede mostrar que

$$C[X, Y] = E[X, Y] - E[X]E[Y] \dots\dots\dots(2.41b)$$

Para la varianza

$$V[X \pm Y] = V[X] + E[Y] \pm 2C[X, Y] \dots\dots\dots(2.42)$$

La covarianza es igual a cero para variables independientes, esto es

$$C[X, Y] = 0 \dots\dots\dots(2.43)$$

Si $X = Y$, la ecuación para la covarianza es igual al cálculo de la varianza en una dimensión.

$$C[X, Y] = V[X] = V[Y] \dots\dots\dots(2.44)$$

Algunas propiedades de la covarianza son

$$C[X + Z, Y] = C[X, Y] + C[Z, Y] \dots\dots\dots(2.45)$$

Donde X, Y y Z son variables aleatorias. En general, para variables múltiples

$$C\left[\sum_{i=1}^{n_x} X_i, Y\right] = \sum_{i=1}^{n_x} C[X_i, Y] \dots\dots\dots(2.46)$$

Donde X_i y Y son variables aleatorias. La ecuación 2.46 puede generalizarse como:

$$C \left[\sum_{i=1}^{n_x} X_i, \sum_{j=1}^{n_y} Y_j \right] = \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} C[X_i, Y_j] \dots \dots \dots (2.47)$$

La ecuación 2.47 es usada ampliamente en análisis geoestadísticos.

2.2.5.6. COEFICIENTE DE CORRELACIÓN.

El coeficiente de correlación se puede definir como:

$$\rho[X, Y] = \frac{C[X, Y]}{\sigma_x \sigma_y} \dots \dots \dots (2.48)$$

Donde $\rho[X, Y]$ es el coeficiente de correlación. Similar a la covarianza, $\rho[X, Y]$ es una cantidad adimensional y debe tomar valores entre -1 y +1, $\rho[X, Y]$ también cuantifica la relación entre dos variables.

2.3. OBJETIVO DE LA GEOESTADÍSTICA.

En el campo de las Geociencias es común encontrar variables distribuidas espacialmente. Para el estudio de estas variables son usados diversos procedimientos geoestadísticos de estimación y/o simulación. Esto es, a partir de un conjunto de muestras tomadas en localizaciones del dominio en que se manifiesta un fenómeno a estudiar y consideradas representativas de su realidad, que por lo general es siempre desconocida, estos procedimientos permiten la descripción o caracterización de las variables con dos fines diferentes. Primero, proporcionar valores estimados en localizaciones de interés y segundo, generar valores que en conjunto presenten iguales características de dispersión que los datos originales. La Geología y la Minería son los campos típicos para la aplicación de estos modelos, ya que en este último surge y se desarrolla la Geoestadística como ciencia aplicada. Se hace referencia en este

capítulo a los conceptos fundamentales de la Geoestadística. Para profundizar en el tema puede ser consultada la bibliografía citada.

2.3.1. CONCEPTO DE GEOESTADÍSTICA.

La Geoestadística se define como la aplicación de la teoría de funciones aleatorias al reconocimiento y estimación de fenómenos naturales⁹, o simplemente, el estudio de las variables numéricas distribuidas en el espacio¹⁰, siendo una herramienta útil en el estudio de estas variables¹¹. Su punto de partida es asumir una intuición topo-probabilista¹². Los fenómenos distribuidos en el espacio, presentan un carácter mixto, un comportamiento caótico o aleatorio a escala local (estudiada por Estadísticos), pero a la vez estructural a gran escala (estudiada por Geoscientistas), lo anterior se ilustra en la figura 2.13.

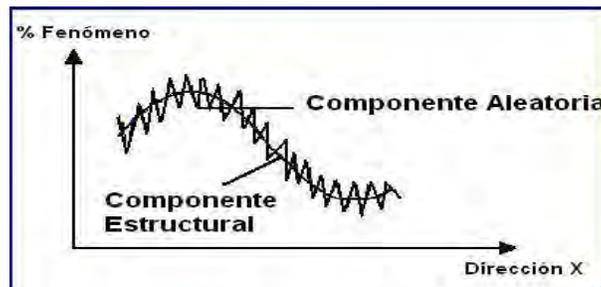


Figura 2.13. Carácter mixto de una función aleatoria.

Todo estudio geoestadístico inicia con un análisis estructural, el cual consiste en ajustar una función denominada variograma a la variabilidad espacial (in-situ) de los parámetros estudiados.

El procedimiento de estimación debe tomar en cuenta la estructura de la variabilidad espacial de cada fenómeno, así como la manera particular en la cual el fenómeno fue muestreado.

Tomando en cuenta estas peculiaridades, es posible asignar a cada valor estimado, un intervalo de confianza. El método de estimación denominado Kriging (en honor al Geoscientista sudafricano Danny G. Krige), toma en cuenta todos estos factores. El Kriging, tiene como objetivo encontrar la mejor estimación posible a partir de la información disponible de las variables estudiadas, y en efecto, el valor estimado obtenido $X^*(\bar{u})$ de un valor real y desconocido $X(\bar{u})$, consiste en una combinación lineal de pesos asociados a cada localización donde fue muestreado un valor $X(\bar{u}_i)$ ($i=1,2,3,\dots,n$) del fenómeno estudiado, observando dos condiciones fundamentales:

- 1) Que el estimador sea insesgado, es decir, que $E[X^*(\bar{u}) - X(\bar{u})] = 0$ y
- 2) Que la varianza $V[X^*(\bar{u}) - X(\bar{u})]$ sea mínima, consiguiéndose de este modo minimizar la varianza de error de estimación.

A diferencia de otros métodos de interpolación, como por ejemplo el inverso de la distancia, el Kriging utiliza en la estimación las características de variabilidad y correlación espacial del fenómeno estudiado, por lo que su uso implica un análisis previo de la información con el objetivo de definir o extraer de esta información inicial, un modelo que represente su continuidad espacial.

Una vez que se logra lo anterior, se presentan las condiciones necesarias para obtener el mejor valor posible en cada localización o bloque a estimar a partir de los datos medidos, acompañada de la varianza de Kriging como medida del error de la estimación que se realiza¹³, lo que distingue al Kriging de otros métodos de interpolación^{14, 15}.

En el capítulo 4 se analiza el modelado por Kriging con mayor detalle.

2.3.2. DIFERENCIA ENTRE ESTADÍSTICA Y GEOESTADÍSTICA.

El objetivo de la Estadística es medir y modelar la variabilidad del proceso mediante un modelo probabilístico, por lo tanto, la Estadística intenta encontrar las regularidades que se encuentran en los fenómenos de masa.

Tradicionalmente, en el estudio de los fenómenos naturales se hace frecuente uso de conceptos básicos de Estadística. Se considera por ejemplo, a cada uno de los valores muestreados, dentro de cierta región R y para un determinado fenómeno natural, como diferentes realizaciones independientes de una misma variable aleatoria X . Más aun, se asume que todos los valores muestreados provienen de una misma distribución. La metodología aplicada consiste en inferir de los valores muestreados, una distribución a partir de la cual, es posible evaluar un fenómeno. Estos métodos, por la simplicidad en las suposiciones, producen, resultados frecuentemente incoherentes. Además, en ellos no se toma en cuenta conceptos tan importantes como el de correlación entre valores muestreados.

Actualmente en Geoestadística, se considera que cada valor muestreado en un punto o soporte x de una región R , representa únicamente una realización de una variable aleatoria $X(\bar{u})$. Si n valores fueran muestreados, ellos representarían realizaciones de n diferentes variables aleatorias $X(\bar{u}_1), X(\bar{u}_2), \dots, X(\bar{u}_n)$, cada una de las cuales tendría su propia función de distribución de probabilidad $f[X(\bar{u}_1)], f[X(\bar{u}_2)], \dots, f[X(\bar{u}_n)]$ como se ilustra en la figura 2.14.

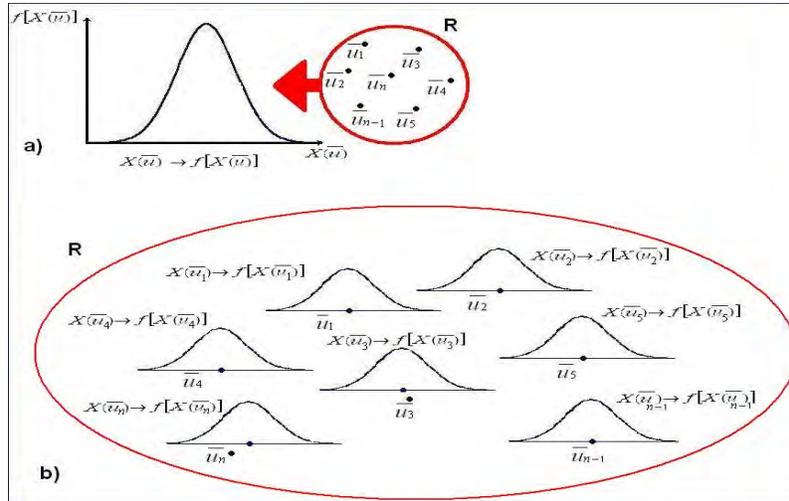


Figura 2.14. Fenómeno natural caracterizado por la distribución espacial de una variable aleatoria regionalizada. (a) Concepto tradicional. (b) Concepto geoestadístico.

2.3.3. VARIABLES ALEATORIAS REGIONALIZADAS.

A las variables aleatorias $X(\bar{u}_1), \dots, X(\bar{u}_n)$ por estar distribuidas en el espacio, se les conoce como variables aleatorias regionalizadas⁹, y al fenómeno representado por las variables aleatorias regionalizadas $X(\bar{u}_i)$ se le denomina regionalización, por ejemplo:

- I. El precio del barril del petróleo, el cual puede ser visto como una variable distribuida en el tiempo (espacio de una dimensión)
- II. El porcentaje de mineral en un bloque de volumen V de un depósito podría ser caracterizado por una distribución espacial en tres dimensiones

Al conjunto de variables aleatorias regionalizadas $\{X(\bar{u}_1), X(\bar{u}_2), \dots, X(\bar{u}_n)\}$, representado de ahora en adelante como $X(\bar{u})$, se le conoce como función aleatoria, en la cual se encuentran implícitos los aspectos estructural y aleatorio, y se distingue, como toda *v.a.*, por tener asociada una función de distribución acumulativa conjunta $\{F[X(\bar{u}_1)], F[X(\bar{u}_2)], \dots, F[X(\bar{u}_n)]\}$.

Sin embargo, es muy frecuente que estas funciones varíen tan irregularmente en el espacio, que impiden un estudio matemático directo, por lo que es necesario realizar un análisis de variabilidad de la información disponible, sugiriendo un estudio profundo de la función variograma como se explicara más adelante.

Conceptos de variable aleatoria regionalizada

En el estudio de las variables aleatorias regionalizadas es importante presentar conceptos que se utilizan por la mayoría de los autores donde se aplican los métodos geoestadísticos como herramienta fundamental de trabajo^{9, 14}.

Estos conceptos son:

- **Región:** Se refiere al espacio en el cual existe y se estudia el fenómeno natural.
- **Localización:** Es el punto de posición dentro de una región en la cual se define una *v.a.r.*
- **Soporte Geométrico:** Está determinado por el elemento físico sobre el cual se realiza la determinación de la *v.a.r.* siguiendo una misma técnica, es decir, la muestra unitaria sobre la cual se estudia el atributo de interés (mismo tamaño relativo, volumen o método de ensaye).

2.3.4. REQUERIMIENTOS DE ESTACIONARIEDAD.

Adicionalmente a la suposición de que todas las localizaciones están descritas por una *v.a.r.*, es necesario considerar restricciones asociadas al uso de datos muestreados disponibles, para predecir valores en localidades donde no es posible obtener muestras. Dicha suposición, en términos cualitativos, requiere que el modelo propuesto basado en los datos muestreados, pueda

describir adecuadamente el comportamiento de la población. Así es como se acepta el cumplimiento de ciertas hipótesis sobre el carácter de una función aleatoria, como son la estacionariedad estricta o estacionariedad de primer orden, y estacionariedad de segundo orden (algunos autores consideran adicionalmente un par de hipótesis adicionales, como son la hipótesis intrínseca y procesos cuasiestacionarios^{9, 14}).

- I. **Estacionariedad Estricta.** Se dice que la función aleatoria $X(\bar{u})$ es estrictamente estacionaria si la función de distribución de probabilidad $f[X(u)]$ permanece constante bajo efectos de translación, lo que requiere que los momentos conjuntos de distinto orden para cada variable aleatoria regionalizada sean iguales y completamente independientes de las localizaciones \bar{u}_i y $\bar{u}_i + \bar{L}$. Esta condición como su nombre lo indica es demasiado restrictiva al estudiar la mayoría de los fenómenos encontrados en la práctica.

En otras palabras equivaldría considerar :

$$\{X(\bar{u}_1), \dots, X(\bar{u}_n)\} = \{X(\bar{u}_1 + \bar{L}), \dots, X(\bar{u}_n + \bar{L})\}$$

Para cualquier valor del vector de translación o incremento \bar{L} . Por lo que:

$$f[X(\bar{u}_i)] = f[X(\bar{u}_i + \bar{L})] \dots \dots \dots (2.49)$$

La función que más se utiliza es el valor esperado. Así es como se puede escribir:

$$E[X(\bar{u}_i)] = E[X(\bar{u}_i + \bar{L})] \dots \dots \dots (2.50)$$

El valor de \bar{L} puede variar desde cero hasta la distancia máxima entre las variables entro de una región de interés.

Esto significa que la media aritmética de una *v.a.r* a lo largo de una región es la misma, es decir, en otras palabras, que dicha media permanece constante.

II. Estacionariedad de Segundo Orden. Una función aleatoria $X(\bar{u})$ es estacionaria de segundo orden cuando ésta es independiente de la localización de las variables aleatorias regionalizadas y solo depende de la dirección y el intervalo del vector de translación \bar{L} que existe entre las variables. En otras palabras:

$$f[X(\bar{u}_i), X(\bar{u}_i + \bar{L})] = f[X(\bar{u}_j), X(\bar{u}_j + \bar{L})] \dots \dots \dots (2.51)$$

En términos de la función covarianza se puede escribir lo siguiente:

$$C[X(\bar{u}_i), X(\bar{u}_i + \bar{L})] = C[X(\bar{u}_j), X(\bar{u}_j + \bar{L})] \dots \dots \dots (2.52)$$

La ecuación 2.52 indica que la covarianza dentro de una región estacionaria es función solo del vector de translación \bar{L} , no de la variable misma. Esta es una importante suposición, en la que se interpreta que, siempre y cuando se conozca la distancia y la dirección entre dos puntos cualesquiera, es posible estimar la covarianza entre las variables aleatorias regionalizadas en esos dos puntos.

Si $\bar{L} = \bar{u}_j - \bar{u}_i$, la ecuación 2.52, puede escribirse como

$$C[\bar{L}] = C[X(\bar{u}_i), X(\bar{u}_i + \bar{L})] \dots \dots \dots (2.53)$$

Y de acuerdo a la definición de covarianza

$$C[X(\bar{u}_i), X(\bar{u}_i + \bar{L})] = E[X(\bar{u}_i), X(\bar{u}_i + \bar{L})] - E[X(\bar{u}_i)]E[X(\bar{u}_i + \bar{L})] \dots \dots \dots (2.54)$$

Sin embargo, la ecuación 2.54 puede simplificarse, ya que, tomando en cuenta la condición de estacionariedad estricta (ver ecuacion2.50).

$$C[X(\bar{u}_i), X(\bar{u}_i + \bar{L})] = E[X(\bar{u}_i), X(\bar{u}_i + \bar{L})] - E\{X(\bar{u}_i)\}^2 \dots \dots \dots (2.55)$$

2.3.5. RELACIONES ESPACIALES.

Las relaciones espaciales entre las variables aleatorias regionalizadas son ampliamente utilizadas para describir la interrelación en la vecindad de los valores muestreados. En esta sección se extenderán los conceptos de covarianza y coeficiente de correlación ya analizados en las secciones 2.3.4.5, y 2.3.4.6, respectivamente, y se introducirá el concepto de variograma.

2.3.5.1. COVARIANZA PARA EL ANÁLISIS DE VARIOGRAMAS

La ecuación 2.55 define a la covarianza bajo condiciones de estacionariedad. Con dichas definiciones, en la practica, la covarianza estimada puede calcularse tomando en cuenta el concepto de valor esperado de una *v.a*, como

$$c[\bar{L}] = \frac{1}{Np[\bar{L}]} \sum_{i=1}^{Np[\bar{L}]} x(\bar{u}_i)x(\bar{u}_i + \bar{L}) - \left[\frac{1}{n} \sum_{i=1}^n x(\bar{u}_i) \right]^2 \dots\dots\dots(2.56)$$

Donde $Np[\bar{L}]$ es el número de pares de valores a la distancia del vector \bar{L} , $x(\bar{u}_i)$ y $x(\bar{u}_i + \bar{L})$ son los valores de la variable en las localizaciones \bar{u}_i y $\bar{u}_i + \bar{L}$, respectivamente, y n es el número total de puntos muestreados. El segundo termino del lado derecho de la ecuación 2.56, representa la media aritmética de todos los datos, y $c[\bar{L}]$ es el valor estimado de la covarianza basado en el muestreo de datos. La ecuación 2.56 se puede reescribir como

$$c[\bar{L}] = \frac{1}{Np[\bar{L}]} \sum_{i=1}^{Np[\bar{L}]} x(\bar{u}_i)x(\bar{u}_i + \bar{L}) - \bar{x}^2 \dots\dots\dots(2.57)$$

Un caso especial se presenta cuando $\bar{L} = 0$, y se conoce como la covarianza al origen $c[0]$, por lo que se tiene

$$c[0] = \frac{1}{Np[\bar{L}]} \sum_{i=1}^{Np[\bar{L}]} x(\bar{u}_i)x(\bar{u}_i) - \bar{x}^2 \dots\dots\dots(2.58)$$

Y en este caso $Np[\bar{L}] = n$

2.3.5.2. COEFICIENTE DE CORRELACIÓN PARA EL ANÁLISIS DE VARIOGRAMAS.

El coeficiente de correlación puede definirse como

$$\rho[\bar{L}] = \frac{C[\bar{L}]}{\sigma_{\bar{u}_i} \sigma_{\bar{u}_i + \bar{L}}} \dots\dots\dots(2.59)$$

Donde $\rho[\bar{L}]$ es el coeficiente de correlación para un intervalo \bar{L} , $\sigma_{\bar{u}_i}$ y $\sigma_{\bar{u}_i + \bar{L}}$ son las desviaciones estándar para los datos en las localizaciones \bar{u}_i y $\bar{u}_i + \bar{L}$ respectivamente. Sin embargo tomando en cuenta la estacionariedad de segundo orden se puede establecer que

$$V[X(\bar{u}_i)] = V[X(\bar{u}_i + \bar{L})] = C[0] \dots\dots\dots(2.60)$$

Y así es como

$$\sigma_{\bar{u}_i} = \sigma_{\bar{u}_i + \bar{L}} = \sqrt{C[0]} \dots\dots\dots(2.61)$$

Sustituyendo la ecuación 2.61 en la ecuación 2.59 se tiene que

$$\rho[\bar{L}] = \frac{C[\bar{L}]}{C[0]} \dots\dots\dots(2.62)$$

Utilizando datos muestreados, puede calcularse el valor estimado del coeficiente de correlación como

$$r[\bar{L}] = \frac{c[\bar{L}]}{c[0]} \dots\dots\dots(2.63)$$

2.3.5.3. VARIOGRAMAS.

La función variograma es la técnica Geoestadística más utilizada para realizar la descripción de relaciones espaciales.

Matemáticamente se define como

$$\gamma[\bar{L}] = \frac{1}{2} V[X(\bar{u}_i) - X(\bar{u}_i + \bar{L})] \dots \dots \dots (2.64)$$

Que es la mitad de la varianza de la diferencia entre dos valores localizados en un intervalo \bar{L} . Es necesario comprender el comportamiento del variograma cualitativamente. De acuerdo a la definición anterior, el valor del variograma cuando $\bar{L} = 0$ es

$$\gamma[0] = \frac{1}{2} V[X(\bar{u}_i) - X(\bar{u}_i)] = 0 \dots \dots \dots (2.65)$$

El variograma empieza con un valor de cero y se incrementa conforme el intervalo entre los dos valores se incrementa. Con la definición de varianza, la ecuación 2.64 puede expandirse a

$$\gamma[\bar{L}] = \frac{1}{2} [E\{[X(\bar{u}_i) - X(\bar{u}_i + \bar{L})]^2\} - E\{[X(\bar{u}_i) - X(\bar{u}_i + \bar{L})]\}^2] \dots \dots \dots (2.66)$$

La condición de estacionariedad de primer orden requiere que

$$E[X(\bar{u}_i)] = E[X(\bar{u}_i + \bar{L})] \dots \dots \dots (2.50)$$

Por lo que el segundo termino de la ecuación 2.66 se vuelve cero, así es como dicha ecuación se simplifica a

$$\gamma[\bar{L}] = \frac{1}{2} E\{[X(\bar{u}_i) - X(\bar{u}_i + \bar{L})]^2\} \dots \dots \dots (2.67)$$

Con la definición de covarianza, la ecuación 2.67 se transforma en

$$\gamma[\bar{L}] = C[0] - C[\bar{L}] \dots \dots \dots (2.68)$$

La función variograma esta fuertemente ligada con la función covarianza, si es que las condiciones de estacionariedad son tomadas en cuenta. El variograma se incrementa conforme el intervalo entre los valores se incrementa; la covarianza se decrementa conforme el intervalo entre los valores se incrementa.

En la practica, la ecuación 2.68 puede escribirse como

$$\hat{\gamma}[\bar{L}] = \frac{1}{2Np[\bar{L}]} \sum_{i=1}^{Np[\bar{L}]} [x(\bar{u}_i) - x(\bar{u}_i + \bar{L})]^2 \dots\dots\dots(2.69)$$

Donde el acento ortográfico ^ sobre γ indica que el variograma es un valor estimado con base en datos muestreados.

2.3.6. CONOCIMIENTO DEL PROBLEMA.

Antes de comenzar un estudio geoestadístico se deben discutir todos los elementos que aporten conocimientos del problema a resolver, la estructura geológica en que se desarrolla el fenómeno en estudio, organización y verificación de la información disponible y finalmente realizar el análisis exploratorio de los datos.

Una vez obtenido los datos, es necesario que se controlen integralmente a fin de verificar por una parte su exactitud y por otra su representatividad. Es importante que se esté familiarizado con los datos, discutir todos los elementos necesarios de acuerdo a los datos que se desean obtener. Cualquier modificación involuntaria en la etapa inicial se refleja sistemáticamente durante todo el estudio¹³.

CAPITULO 3

ANÁLISIS PETROFÍSICO

3.1. INTRODUCCIÓN.

En el presente capítulo se hace mención de los conceptos más importantes en lo referente al entorno petrofísico, ya que son sumamente necesarios para desarrollar un modelado de este tipo. Adicionalmente se analizan algunas diferencias existentes entre el modelado determinístico y el modelado estocástico, ya que son la estructura principal que soporta el principio de construcción de un modelado numérico de yacimientos, así como las técnicas necesarias para el desarrollo de este. El concepto de variograma es fundamental para obtener los modelados discutidos en el presente trabajo, por lo tanto se detalla un amplio análisis con respecto a este tema, con la finalidad de explotar al máximo las características más relevantes de este concepto aplicadas a un análisis petrofísico.

3.2. PETROFÍSICA Y PARÁMETROS PETROFÍSICOS.

La petrofísica puede definirse como la física de las rocas. De manera aun más particular puede definirse como el estudio del espacio poroso de una roca, de sus características intrínsecas así como de las que se originan al estar ocupado el espacio poroso por fluidos que reaccionan, de alguna manera, con los sólidos, o, como el estudio de las propiedades fisicoquímicas de la roca del yacimiento, y su ingerencia en las actividades de Geociencia (exploración petrolera) e Ingeniería (producción de hidrocarburos)¹.

Un amplio porcentaje de la producción de petróleo y gas a nivel mundial, se extrae de acumulaciones en espacios porosos de las rocas de un yacimiento, generalmente areniscas, calizas y dolomías. La cantidad de petróleo o gas contenida en una unidad volumétrica del yacimiento es el producto de su porosidad por la saturación de hidrocarburos.

Además de la porosidad y la saturación de hidrocarburos, se requiere el volumen de la formación que contiene hidrocarburos para calcular las reservas totales y determinar si la reserva es comercial. Es necesario entonces, conocer el espesor y el área del yacimiento para calcular su volumen.

Para evaluar la productividad de un yacimiento, se necesita saber con que facilidad puede fluir el líquido o gas a través del medio poroso. Esta propiedad de la roca que depende de la manera en que los poros estén intercomunicados, se llama permeabilidad.

Por lo tanto es necesario determinar parámetros petrofísicos para evaluar el desempeño de los yacimientos petroleros.

Los parámetros petrofísicos de la roca que comúnmente se determinan son¹:

- Forma, tamaño y empacamiento de grano.
- Geometría y tamaño de poro.
- Porosidades primaria y secundaria.
- Permeabilidades absolutas.
- Tortuosidad.
- Radioactividad natural.
- Transmisión de ondas acústicas (tiempo de transito).
- Compresibilidad.

Los parámetros petrofísicos del sistema roca-fluidos que comúnmente se determinan son¹:

- Saturación de fluidos.
- Resistividad y/o conductividad.
- Capilaridad o presión capilar.
- Permeabilidades a fluidos, efectivas y relativas.
- Compatibilidad roca-agua.
- Densidad.

- Mojabilidad.
- Capacidad de intercambio catiónico.

3.3. DESCRIPCIÓN DE YACIMIENTOS MEDIANTE CARACTERIZACIÓN PETROFÍSICA.

El objetivo de la caracterización petrofísica, es el de determinar cualitativa y cuantitativamente las características propias de la roca y del sistema roca-fluidos¹, definiendo así detalladamente su variación dentro del yacimiento.

En la caracterización petrofísica debe tomarse a consideración el realizar la caracterización dentro de un apropiado modelo geológico del yacimiento o campo petrolero involucrando la sinergia, combinando los procesamientos y las interpretaciones obtenidas con datos de distintas fuentes de información, para finalmente, determinar los parámetros petrofísicos con un alto grado de precisión ya que estos son básicos para llevar a cabo una adecuada explotación de hidrocarburos.

Las fuentes de información más comunes para realizar estudios de caracterización petrofísica son¹:

- Muestras de roca.
- Registros geofísicos de pozos.
- Pruebas en pozos (formación, producción, variación de presión).
- Historias de presión-producción.
- Muestras de fluidos (agua, hidrocarburos).

Cada fuente de datos esta sujeta a limitaciones y errores.

Los datos de muestras y registros describen una región poco profunda alrededor del pozo. El tamaño de una muestra de rutina es muy reducido en

comparación con el de una capa del yacimiento. Las propiedades determinadas con base en muestras presentan mayor variación que los datos promediados sobre volúmenes de roca más grandes. Otras dificultades causadas por la naturaleza puntual de la medición de muestras se encuentran al relacionar la permeabilidad de las muestras con las propiedades de flujo de capas a gran escala y al definir la permeabilidad vertical.

Debido a la dificultad y el costo de efectuar corte de núcleos, solo se obtiene muestras en pozos clave². Dichas muestras son sometidas a un riguroso análisis para desarrollar el modelo geológico del yacimiento y determinar relaciones entre los parámetros petrofísicos de la formación y los parámetros de la formación que pueden determinarse con registros³. Una vez establecida tal relación, los parámetros petrofísicos de la formación (incluyendo las distribuciones de permeabilidad) pueden deducirse en ocasiones, de los datos de registros geofísicos en áreas sin datos de núcleos. Para este propósito se han desarrollado técnicas que utilizan bancos de datos de registros multidimensionales. Estas proporcionan distribuciones continuas de los parámetros petrofísicos de la formación que son coherentes con las muestras, la información geológica, la presión y otros datos para cada pozo del campo y, por lo tanto, son un complemento importante para mejorar las técnicas de descripción de yacimientos.

Los datos de registros y atributos sísmicos también se han usado exitosamente para definir y correlacionar tipos de rocas a través del modelado e interpretación estratigráfica.

3.3.1. INTERPRETACIÓN DE REGISTROS GEOFÍSICOS.

Debido a que solo algunos de los parámetros petrofísicos pueden medirse directamente, es necesario inferir u obtener cálculos para determinar las

mediciones de otros parámetros petrofísicos de las formaciones. Actualmente es posible medir una gran cantidad de parámetros que incluyen, entre otros, la resistividad, la densidad, el tiempo de tránsito, el potencial espontáneo, la radioactividad natural y el contenido de hidrógeno de la roca.

La interpretación de registros es el proceso por el cual dichos parámetros medibles se traducen a los parámetros petrofísicos deseados de porosidad, saturación de hidrocarburos, permeabilidad, litología, etc.

Algunos parámetros geológicos, como composición, textura y estructura, pueden relacionarse de cierta forma con los sensores de un registro geofísico para realizar correlaciones simples hasta el estudio de yacimientos completos, pasando por información estratigráfica.

La composición de una roca puede expresarse de manera mineral o química. La composición mineral proporciona propiedades petrofísicas como dureza, densidad, resistividad y tiempo de tránsito⁴. Algunas herramientas miden la concentración de los elementos, mientras que otras proporcionan una respuesta que depende del porcentaje de la distribución de los minerales de la roca.

La textura se refiere a los aspectos geométricos de los elementos de la roca, tales como clasificación de tamaño, presentación, forma, disposición de los granos (fabrica), disposición de la matriz y el cementante, contactos entre grano y grano, grano y matriz o grano y cementante. Las propiedades petrofísicas de la roca, tales como porosidad y permeabilidad, dependen principalmente de la textura.

Mientras que la textura alude a las relaciones entre los granos de la roca, las estructuras sedimentarias se refieren a las características más grandes que reflejan las condiciones en el momento del depósito, tales como energía y tipo

de corriente. Las estructuras constituyen un importante elemento de las facies de una unidad sedimentaria y ayudan a definir el ambiente de depósito.

La traducción de parámetros se complica aun más debido a los procesos de perforación. Al perforar a través de una formación, los fluidos dentro de los poros de la roca que rodean al agujero pueden verse desplazados o contaminados debido a la invasión de los fluidos de perforación.

Debido a que se requieren los parámetros petrofísicos de la formación virgen y no contaminada, la herramienta de registro debería ser capaz de "ver" más allá de la zona contaminada en la región virgen; como alternativa, las técnicas de interpretación deben ser capaces de compensar los problemas de contaminación en la formación.

En el registro geofísico, se utilizan ambos enfoques. Cuando las propiedades físicas de la medición impiden una investigación a nivel profundo, las técnicas de interpretación deben de tomar en cuenta los problemas ocasionados por el filtrado del lodo.

El propósito de las diferentes herramientas de registro, es el de proporcionar mediciones sobre las que sea posible obtener o inferir las características petrofísicas de las rocas del yacimiento⁵. La meta de la interpretación cuantitativa de registros es proporcionar las ecuaciones y técnicas para que dichos cálculos puedan llevarse a cabo.

Para respaldar el análisis de registros, se calibran los métodos de interpretación con base en las descripciones de muestras de canal y los análisis de núcleos.

3.3.2. CORRELACIÓN DE REGISTROS.

Uno de los primeros usos de los registros geofísicos fue el de realizar correlaciones entre los estratos existentes en pozos vecinos. Los intervalos de registros de distintos pozos se comparan para observar su similitud o las respuestas características de los registros como indicadores litológicos⁶.

Los registros tiene la ventaja, para efectos de correlación, de proporcionar información continua en todo el pozo, a diferencia de la toma de núcleos, registrando la profundidad de la sonda evitando ambigüedad en cuanto a profundidad de los distintos indicadores.

Los estudios de correlación mediante registros geofísicos y atributos sísmicos entre pozos, permiten entonces realizar mapas exactos del subsuelo y determinar:

- a) Las profundidades de las formaciones presentes en el pozo en comparación con otro pozo, así como los afloramientos y las proyecciones geofísicas.
- b) Si es que el pozo forma parte o no de una estructura geológica mayor.
- c) Si la profundidad del pozo ha alcanzado un horizonte productivo desconocido, y de lo contrario, cuanto falta por perforar.
- d) La presencia o ausencia de fallas.
- e) La existencia de inclinaciones, pliegues, discordancias, el hecho de que las unidades litológicas aumenten o disminuyan, o los cambios laterales de sedimentación y/o litología.

Los registros que más se utilizan en la industria petrolera son los de resistividad, potencial espontáneo, rayos gamma y sónico.

MODELADO GEOESTADÍSTICO DE PROPIEDADES PETROFÍSICAS DE YACIMIENTOS

Registro	Parámetros de Correlación	Condiciones para el mejor Desempeño	Usos y Ventajas
SP	Capas permeables vs Capas de lutitas.	Agujero descubierto. Buen contraste entre R_{mf} y R_w . Lodo hecho con agua dulce (aunque puede usarse también lodo hecho con agua salada).	Fácil lectura de despliegue en perfiles lutitas-arenas. Utilizado para correlaciones con el registro de resistividad. No es afectado por agujeros lavados o profundidad de invasiones variables.
Rayos Gamma	Radioactividad asociada al volumen de lutitas. Capas radioactivas.	Tamaño de agujero moderado con zonas lavadas no muy largas.	Insensible a fluidos de perforación. Puede utilizarse en agujeros revestidos
Resistividad Espaciamento Corto Amplificado	Mismos que el registro anterior.	Mismas que el registro anterior.	Utilizado para correlación de secciones de lutitas o de baja resistividad.
Caliper	Variaciones en el tamaño del agujero (lavados de pozo, fracturamientos).	Agujero descubierto	Resuelve ambigüedades sobre otro tipo de registros.
Resistividad Espaciamento Corto (SN, LL8,SFL)	Zonas invadidas y capas porosas. La deflexión en el registro depende del factor de formación, resistividad del agua y del volumen de lutitas. Estrato denso (bajo contenido de agua y matriz no conductiva).	Agujero descubierto. Lodo hecho con agua dulce. Formaciones invadidas, no muy resistivas.	Utilizado para correlación con los registros SP o rayos gamma.
Laterolog Profundo	Mismos que el registro anterior.	Agujero descubierto. Lodo hecho con agua dulce o salada. Altas relaciones R_t/R_w .	Utilizado para Lodo hecho con agua salada y formaciones resistivas.

Tabla 3.1a. Registros geofísicos básicos utilizados para correlación.

Registro	Parámetros de Correlación	Condiciones para el mejor Desempeño	Usos y Ventajas
Inducción	Variaciones de contenido de agua y salinidad en capas con matriz no conductiva. (La respuesta en zonas porosas varía con la porosidad de la formación y la conductividad poro-fluido).	Agujero descubierto. Lodo hecho con agua dulce. Resistividad en la formación por debajo de 100 ohm-m.	La curva de inducción de conductividad se utiliza para correlacionar secciones de lutitas o de baja resistividad
Sónico	Tiempo de transito (Dependiente de la litología y la porosidad).	Agujero lleno de líquido y libre de gas.	Buena correlación de porosidad. Utilizado en formaciones de baja resistividad. Utilizado para la determinación de marcadores geológicos.
Neutrón	Contenido de hidrogeno en la formación.	Depende del tipo de herramienta.	Buena correlación de porosidad. Buena combinación con otros registros detectores de gas. Utilizado para la determinación de marcadores geológicos. Utilizado en agujeros revestidos.
Densidad	Densidad de la formación (Dependiente de la litología y la porosidad).	Agujero descubierto, con poco enjarre y sin rugosidad en las paredes del agujero.	Utilizado para la determinación de marcadores geológicos.

Tabla 3.1b. Registros geofísicos básicos utilizados para correlación (continuación).

La presentación de la información puede realizarse mediante un diagrama de rayas, en el que se representa la correlación entre los pozos, generalmente, por líneas casi verticales.

MODELADO GEOESTADÍSTICO DE PROPIEDADES PETROFÍSICAS DE YACIMIENTOS

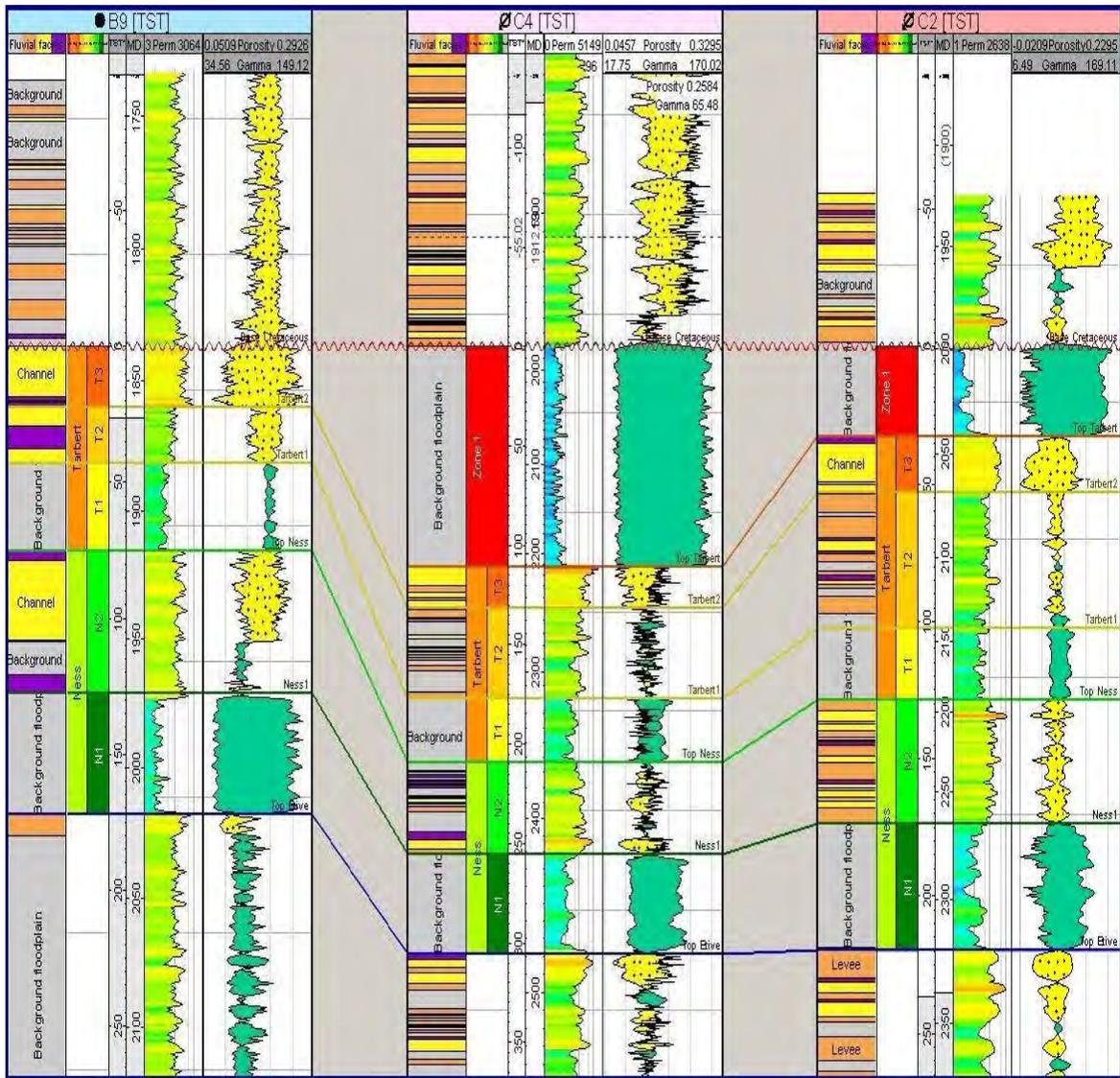


Figura 3.1. Sección transversal que muestra la correlación detallada con un despliegue de registros para tres pozos.

3.4. MODELADO PETROFÍSICO.

Una adecuada descripción de las propiedades petrofísicas, como porosidad y permeabilidad, así como su variación espacial es esencial para comprender la distribución de los fluidos y las propiedades de flujo de un yacimiento.

El objetivo principal de un modelado petrofísico es el de poblar las mallas de simulación con propiedades petrofísicas (porosidad permeabilidad, saturación de agua) provenientes de registros geofísicos de pozos utilizando técnicas de integración de datos, para obtener escenarios pesimistas y optimistas mediante modelos de estimación y/o simulación, comparando así posteriormente dichos modelos con el cálculo de volúmenes por zona, facies y/o unidades estratigráficas, con la finalidad de determinar el modelo que mejor se ajuste a la realidad y a la necesidad del problema a resolver para integrar la información en un modelado de flujo.

Para desarrollar un modelado petrofísico es posible incorporar un conjunto completo de algoritmos de estimación o simulación, como se describe a continuación.

3.4.1. MODELADO DETERMINÍSTICO Y ESTOCÁSTICO.

En un modelado determinístico, se supone que toda la información disponible se conoce con certidumbre. Cada yacimiento es único y está definido, sin embargo, a estos hechos se le asocian algunos problemas. Primero, que un modelado determinístico es el que es más deseado para describir al yacimiento. Si es posible describir cada aspecto del yacimiento con certidumbre, no es necesario conocer la Geoestadística o cualquier otro principio estadístico para describir al yacimiento.

En principio, con suficiente conocimiento sobre el origen del yacimiento, es posible crear un modelado determinístico que permita describir al mismo. Esto requiere saber como es que se formó el yacimiento, bajo que circunstancias se entrampó el aceite, que tipo de presiones y temperaturas prevalecieron cuando se acumularon los hidrocarburos, que tipos de procesos geológicos sucedieron subsecuentemente a la presencia de los hidrocarburos,

que interacciones químicas ocurrieron después de que los mismos se entraparon, y otro tipo de factores posibles. Si se comprenden todos estos procesos con suficiente detalle, y se tiene la habilidad de conjuntar estos mismos en un modelo matemático, su resolución permitirá predecir todas las características necesarias del yacimiento.

Desafortunadamente en la mayoría de las ocasiones, se presenta la incertidumbre en la descripción de yacimientos, por lo que los fenómenos o experimentos a analizar deben estudiarse como si fuesen aleatorios.

A un modelado determinístico suele nombrársele modelado por “estimación” (en este tipo de modelados pueden aplicarse técnicas para modelar por Kriging o Inverso de la distancia, entre otras), y solo es posible obtener una malla, para obtener las estimaciones.

Los algoritmos empleados en un modelado de tipo determinístico, toman literalmente, los valores de los datos, y en ellos se asume que los valores calculados en una malla con relación a los puntos muestreados, tiene por lo menos relación geométrica respecto a dichos puntos. Estos algoritmos no intentan preservar las características en la distribución de los datos de entrada, lo que significa que el histograma de los valores calculados de la malla puede o no puede parecerse al histograma de los datos de entrada.

Ventajas de los algoritmos empleados en un modelado determinístico:

- Existe una gran cantidad de estos en la literatura.
- Son simples de utilizar.
- Son intuitivos.

Desventajas de los algoritmos empleados en un modelado determinístico:

- Son inapropiados para la simulación de flujo.
- No proporcionan ningún tipo de información de probabilidades.
- Tienden a requerir abundantes cantidades de datos para obtener buenos resultados.

La descripción de yacimientos es una combinación de observaciones (componentes determinísticos), conocimientos (componentes geológicos, sedimentológicos, estructurales, estratigráficos) y suposiciones (componentes estocásticos). En un modelado estocástico, los fenómenos o variables se caracterizan por presentar no siempre el mismo resultado (por lo que no hay regularidad determinística) de acuerdo a un conjunto de circunstancias, pero los distintos resultados presentan regularidades estadísticas.

Las técnicas estocásticas para describir yacimientos pueden aplicarse debido a distintas razones:

- Se presenta información incompleta respecto a dimensión, arquitectura, interna (geometría) y variación en las propiedades de las rocas de acuerdo a las diferentes escalas.
- Existen complejas disposiciones espaciales de los bloques o facies del yacimiento.
- La dificultad de capturar la variación en las propiedades de las rocas, así como las estructuras en función de la posición y la dirección espacial.
- No se conocen las relaciones entre el valor de una propiedad y el volumen de roca utilizados para obtener promedios (problemas en el escalamiento de propiedades).
- La relativa abundancia de propiedades estáticas sobre propiedades dinámicas en los datos del yacimiento.
- Conveniencia y rapidez.

El modelado estocástico para la descripción de yacimientos se refiere en términos generales a la generación de una arquitectura geológica sintética y / o datos de propiedades en 1D, 2D o 3D, los cuales están condicionados a observaciones que contemplan un número deseado de características geológicas, con lo que se espera que el modelado sea lo más apegado a la realidad. La meta principal de esta cuantificación es la de exponer los efectos de incertidumbres en la descripción geológica, en los estados original, presente y futuro del yacimiento.

A los modelados estocásticos (probabilísticos), se les suele nombrar modelados de “simulación” (en este tipo de modelados pueden aplicarse técnicas para modelar por simulación Gaussiana secuencial, simulación truncada, simulación fluvial, entre otras), y es posible obtener una o múltiples mallas, para obtener las simulaciones.

Los algoritmos empleados en un modelado de tipo estocástico, además de tomar literalmente los valores de los datos, permiten asumir, que lo que ocurre entre los valores calculados en la malla y los puntos muestreados, es función de mediciones estadísticas, como lo es la distribución de frecuencias, tanto horizontal como vertical. Adicionalmente, los algoritmos utilizan técnicas aleatorias en la selección de datos para cada nodo calculado de la malla. Una característica de las mallas en la que se obtienen los resultados, en donde se utiliza este tipo de algoritmos en modelados estocásticos para un conjunto de datos así como sus parámetros, es la equiprobabilidad entre las mallas. Ya que es común generar múltiples realizaciones (versiones) de una propiedad y luego estudiar sus diferencias y tipos de distribución para determinar probabilidades globales como bases del modelado final.

Ventajas de los algoritmos empleados en un modelado estocástico:

- Funcionan de manera eficiente aun cuando existe dispersión de datos.
- Permiten desarrollar un modelado mucho más realista cuando se presenta dispersión y dificultad de interpretación en los datos.
- Permiten obtener una tasa de probabilidad global.

Desventajas de los algoritmos empleados en un modelado estocástico:

- Son más difíciles de utilizar.
- Consumen demasiado tiempo.

3.4.2. ANÁLISIS DE DATOS.

El análisis de datos petrofísicos es un proceso que tiene la finalidad de obtener un control en la calidad de los datos de entrada, así como explorar y preparar los mismos para su utilización en el modelado petrofísico. A la vez el análisis de datos esta compuesto por la ejecución de tres actividades principales, las cuales son:

1. Transformación de datos.
2. Análisis del variograma.
3. Correlación de parámetros petrofísicos.

3.4.2.1. TRANSFORMACIÓN DE DATOS.

Debido a que en muchas ocasiones se tienen distintos órdenes en la variación de los datos (valores extremos o valores que están fuera de rango en una distribución), se presenta cierta inestabilidad en la estimación del

variograma. También es común que se presenten tendencias en el espacio, ya sea de manera vertical, areal y/o tridimensional, que no reflejan la realidad tal cual es, causando la consideración de valores irreales en la estimación de propiedades petrofísicas, por ejemplo, al contemplar valores negativos. Por lo tanto es necesario minimizar la variación de los datos identificando y removiendo las tendencias poco representativas.

Para minimizar la variación de los datos es posible aplicar transformaciones no lineales que permiten resolver este problema.

Algunas variantes de transformaciones no lineales muy utilizadas son:

- Transformación logarítmica.
- Transformación en potencia o potencial.
- Transformación en rango.
- Transformación por indicadores.
- Transformación normal.

Transformación logarítmica.

La transformación logarítmica es una de las transformaciones utilizadas comúnmente. Consiste en obtener el valor del logaritmo de la muestra, el cual puede ser natural o en base diez; el orden de las variaciones de magnitud se traslada dentro de variaciones en la parte entera del logaritmo de la variable. Esto puede minimizar el efecto de los valores extremos y el orden de las variaciones de magnitud dentro de un conjunto de datos. Los resultados de la transformación logarítmica se reflejan en una mejor identificación de la estructura espacial. Entre mayor sea la variación en el conjunto de datos original, mayor será el impacto de la transformación logarítmica en el variograma estimado.

Transformación en potencia o potencial.

El uso de la transformación en potencia es similar a la transformación logarítmica. El valor estimado se obtiene elevando a una potencia menor que uno el valor de la variable original. La transformación en potencia esta dada por $x_T(\bar{u}_i) = [x(\bar{u}_i)]^p$ (3.1)

Donde $x_T(\bar{u}_i)$ es el valor de la variable transformada, $x(\bar{u}_i)$ es la variable original y p es la potencia a elevar. Entre menor sea la potencia empleada, menor será la variación del valor de los datos. El valor más utilizado de p es 0.5, el cual representa la raíz cuadrada de los valores de los datos muestreados.

En comparación con la transformación logarítmica, la transformación en potencia, es menos efectiva para efectos de transformación de datos, y en general la transformación en potencia no es muy eficiente para eliminar las fluctuaciones o ciclos en el variograma.

Transformación en rango.

Para calcular el rango del valor de una muestra, se hace un arreglo en orden ascendente con todas las muestras de datos, el valor más bajo en magnitud, va primero, y el valor más alto, va al final. El rango del *iésimo* valor en la secuencia se calcula mediante

$$R_i = \frac{i}{n+1} \dots\dots\dots(3.2)$$

Donde n es el número total de puntos muestreados y R_i es el rango del *iésimo* valor en la secuencia. El rango del valor más alto es insignificamente menor a uno, para asegurar que solo existe una pequeña posibilidad de que el valor estimado en una localidad donde no es posible obtener muestras, sea menor que el valor más pequeño, o mayor que el valor más alto.

Sin embargo, la utilización de la transformación en rango, solo proporciona información sobre el relativo orden de los datos muestreados, y no transmite ninguna información acerca de la magnitud de la muestra. Si solo existe interés en el rango relativo de una variable en particular dentro de una localidad donde no es posible obtener muestras, una transformación por rango puede capturar esa información. En general, la transformación en rango puede auxiliar para definir la estructura del variograma. La transformación en rango no retiene ninguna información sobre la magnitud absoluta de los datos muestreados. No obstante, la restricción de los valores entre cero y uno, no elimina la influencia de los valores que están fuera de rango en una distribución (con referencia a la variación de datos y no a la transformación en rango). Adicionalmente, la transformación en rango puede utilizarse para ciertos procedimientos de estimación, en donde se intenta definir incertidumbres en localidades con base en distribuciones de probabilidad⁸.

Transformación por indicadores.

La transformación por indicadores permite transformar una variable continua en una variable discreta. Una gran cantidad de variables del yacimiento, como lo son la permeabilidad y la porosidad, son variables continuas. Dichas variables continuas pueden representarse en términos de variables discretas mediante una transformación por indicadores, que se define como⁹

$$I(\bar{u}, x_L) = 1 \dots\dots\dots(3.3a)$$

Si $X(\bar{u}) \leq x_L$, y

$$I(\bar{u}, x_L) = 0 \dots\dots\dots(3.3b)$$

Si $X(\bar{u}) > x_L$

Donde $I(\bar{u}, x_L)$, es la variable indicadora en la en la localización (\bar{u}) , x_L es el valor limite, y $X(\bar{u})$ es la variable original que será transformada. En la

práctica, cada valor muestreado puede transformarse a un valor indicador mediante⁹

$$i(\bar{u}_j, x_L) = 1 \dots\dots\dots(3.4a)$$

Si $x(\bar{u}_j) \leq x_L$, y

$$i(\bar{u}_j, x_L) = 0 \dots\dots\dots(3.4b)$$

Si $x(\bar{u}_j) > x_L$

Donde $i(\bar{u}_j, x_L)$, es el valor del indicador en el límite x_L , y $x(\bar{u}_j)$ es el valor de la variable en la localización (\bar{u}_j) .

La variable indicadora solo puede tomar los valores de uno y cero. Si el valor de la muestra es menor que el límite propuesto, el indicador toma el valor de uno. Si el valor de la muestra es mayor que el límite propuesto, el indicador toma el valor de cero. Al especificar múltiples valores para el límite, es posible definir múltiples valores para el indicador a cada límite dado.

La selección del valor del límite es subjetiva. En ocasiones pueden tomarse, por ejemplo, los correspondientes valores de los deciles o percentiles como valores para establecer los límites.

Desafortunadamente, la transformación por indicadores pierde sutiles variaciones en los valores muestreados, sin embargo, permite eliminar el impacto de los datos que se encuentran fuera de rango en una distribución, describiendo el valor de la muestras en términos de valores del indicador, lo cual presenta dos grandes ventajas. Primero, la apropiada definición de los valores límites y la apropiada estimación del variograma del indicador para cada límite, permiten examinar como es que los valores muestreados guardan una relación con los diferentes límites. Segundo, la definición del indicador permite que cierta información cualitativa se transforme en información cuantitativa.

Transformación normal.

La transformación normal permite transformar datos muestreados en datos equivalentes que siguen una distribución normal (Gaussiana).

La figura 3.2 muestra la transformación de datos esquemáticamente.

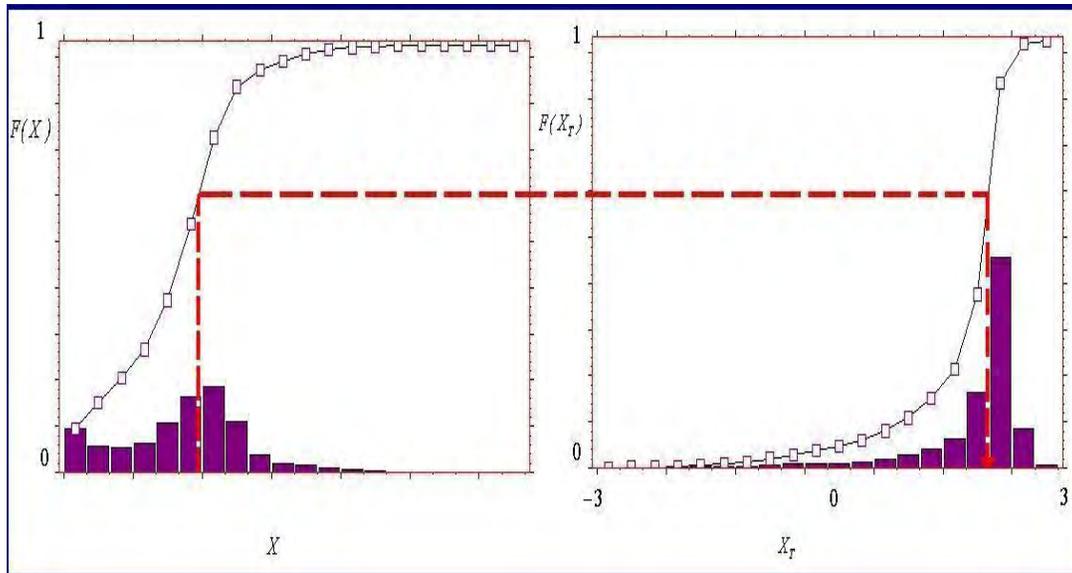


Figura 3.2. Transformación de datos originales en una distribución normal.

De manera típica, los valores transformados de manera normal, adquieren una media de cero y una desviación estándar de uno; por lo tanto estos, varían de -3 a +3.

Con este tipo de transformación no existe un orden en la magnitud de las variaciones dentro del conjunto de datos transformados, por lo que el efecto de los datos extremos en el variograma puede minimizarse, reduciendo las fluctuaciones o ciclos en el mismo, con dicha transformación. Además después de obtener la estimación con los datos transformados, es posible revertir la transformación, es decir, reobtener los valores originales, lo cual también es una ventaja.

3.4.2.2. ANÁLISIS DEL VARIOGRAMA.

La función variograma permite obtener una descripción cuantitativa de la variación de una propiedad que depende de la distancia de separación. Se basa en el principio de que dos puntos muy cercanos, preferentemente, tienen valores similares, que aquellos que se encuentran más lejos entre sí. Existen dos aspectos importantes referentes al variograma:

- Que tanta similitud guardan dos valores cuando estos se encuentran cercanos entre sí.
- Que tan lejos deben estar los puntos de muestreo antes de que no exista relación entre cada uno.

El variograma ayuda a describir la variación natural de una propiedad, en lugar de escalar completamente las tendencias que se observan en los datos. Es necesario considerar que la identificación de cualquier tendencia regional en un análisis de datos debe realizarse previo a comenzar con un análisis del variograma.

Adicionalmente el variograma tiene una gran utilidad como parámetro de entrada cuando se pretende alimentar con muestras de datos de alguna propiedad aleatoriamente distribuida en el espacio, a celdas de malla en 3D, los cuales permitirán el desarrollo de un modelado.

3.4.2.3. CORRELACIÓN DE PARÁMETROS PETROFÍSICOS.

Este tipo de análisis es de gran ayuda en la comprensión de las relaciones existentes entre dos parámetros petrofísicos. El número de pares de datos es importante para definir la correlación entre dos variables. Por ejemplo, un coeficiente de correlación bajo en el caso en que se cuente con cien pares de puntos, es mejor que un coeficiente de correlación alto para el caso en el que se tengan diez pares de puntos.

Para mayores detalles respecto a la interpretación y formulación matemática pueden consultarse las secciones 2.3.4.6 y 2.4.5.2, del capítulo dos.

3.5. ESTIMACIÓN DE VARIOGRAMAS.

El análisis estructural o estudio de variogramas está compuesto por¹⁰:

- El cálculo del variograma experimental.
- El ajuste a este de un modelo teórico conocido.

El cálculo del variograma experimental es la herramienta Geoestadística más importante en la determinación de las características de variabilidad y correlación espacial¹¹, es decir, tener conocimiento de como la variable cambia de una localización a otra¹², representa el útil más importante del cual dispone la Geoestadística para el análisis del fenómeno de la variable de distribución espacial en estudio^{13, 14}. Este análisis tiene como condicionantes: la distribución estadística, la existencia de valores aberrantes o anómalos, y la presencia de zonas homogéneas.

El variograma medio, global u “omnidireccional” (ver variograma experimental), puede calcularse inicialmente para proporcionar una idea preliminar de la variabilidad espacial de los datos, y es el más idóneo para representar u obtener una estructura clara y definida. Posteriormente deben calcularse los variogramas en diferentes direcciones, por ejemplo, pueden calcularse en cuatro direcciones separadas 45° con una tolerancia angular de 22.5°, comenzando por 0° (figura 3.3a) hasta encontrar la dirección de máxima o mínima variabilidad (figura 3.3b), pueden calcularse también, específicamente, en 8 direcciones separadas por 22.5°. Una forma rápida y práctica de visualizar la existencia de anisotropía es mediante el cálculo del “mapa de variogramas”¹⁵,¹⁶ (figura 3.4), lo que permite el análisis adecuado de anisotropía. Posteriormente,

dependiendo de la continuidad espacial, es suficiente sólo calcular dos variogramas separados 90°.

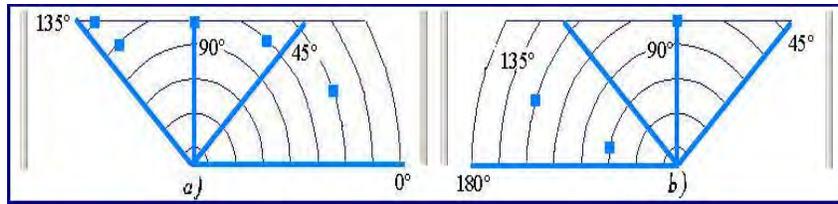


Figura 3.3. Direcciones para el cálculo de los variogramas. (a) Comenzando en 0°. (b) Comenzando en 45°.

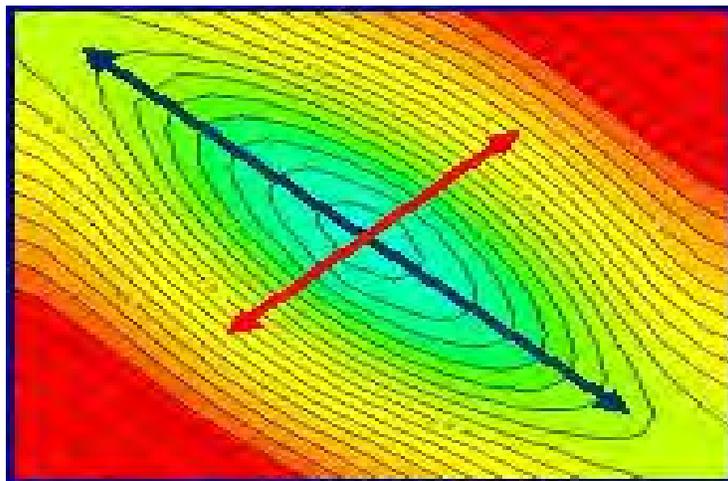


Figura 3.4. Mapa de variogramas. Las flechas muestran las direcciones de anisotropía menor (roja) y mayor (azul).

El variograma experimental resultante no se utiliza en el proceso de estimación, este debe ajustarse a dicho proceso mediante uno a varios modelos teóricos, obteniendo así un modelo o función analítica que caracteriza la continuidad espacial de la variable estudiada. Los modelos de variogramas teóricos que se utilizan en el proceso de estimación o simulación deben satisfacer ciertas condiciones, es decir, tienen que ser “definidos positivos” o de “tipo positivo”^{17, 18, 20}, de lo contrario puede existir el riesgo de encontrar varianzas negativas que no tienen sentido¹⁰. En general el ajuste a modelos

teóricos para la determinación de los parámetros del variograma se realiza de forma visual. En ocasiones se efectúan ajustes polinomiales por el método de mínimos cuadrados u otras variantes, con los que, aunque se encuentra el mejor ajuste, no siempre se verifica la condición de que el variograma obtenido sea siempre de tipo positivo, siendo en ocasiones no satisfactorio²⁰. Finalmente deben obtenerse uno o varios modelos de variogramas con los correspondientes valores de meseta y alcance. El modelo de variograma seleccionado debe representar fielmente los aspectos que se suponen importantes del variograma experimental, que serán usados posteriormente en los procesos de estimación o simulación.

3.5.1. VARIOGRAMA EXPERIMENTAL.

La función variograma $\gamma[\bar{L}] = \frac{1}{2Np[\bar{L}]} \sum_{i=1}^{Np[\bar{L}]} [x(\bar{u}_i) - x(\bar{u}_i + \bar{L})]^2$ representa el útil más importante en todo estudio geoestadístico^{10, 19}. Su cálculo no consiste en una simple evaluación de su expresión, esta operación se relaciona con los siguientes elementos:

- La dirección en la que se calcula el variograma, uno o dos ángulos que definen una dirección en el espacio α y/o β con tolerancias angulares $d\alpha$ y/o $d\beta$.
- El incremento \bar{L} en el cálculo del variograma y su tolerancia lineal $d\bar{L}$; se recomienda que el valor de $d\bar{L}$ sea la mitad del incremento inicial.
- Una distancia, que representa la distancia máxima a la que pueden estar alejados los segundos puntos del par con respecto a la línea que define la dirección de cálculo, que se conoce como ancho de banda.
- La distancia L_{\max} hasta la que se calcula el variograma. Se recomienda que ésta sea la mitad de la distancia entre las muestras más alejadas^{10, 21}, aunque dependiendo de la geometría del fenómeno regionalizado, en algunos casos puede calcularse hasta una distancia superior.

Al definir los elementos anteriores, se evalúa la expresión del variograma para todos los pares de datos de localizaciones separadas a la distancia \bar{L} , que deben cumplir las siguientes condiciones:

1. Que la distancia entre las localizaciones \bar{u}_i y $\bar{u}_i + \bar{L}$ sea mayor que $\bar{L} - d\bar{L}$ y menor que $\bar{L} + d\bar{L}$, o lo que es lo mismo, el segundo punto del par de datos, esté incluido en el espacio definido por $\bar{L} - d\bar{L}$ y $\bar{L} + d\bar{L}$, por lo que el primer punto del par de datos se encuentra en el origen "o" (figura 3.5), este origen se mueve entre las muestras a analizar.

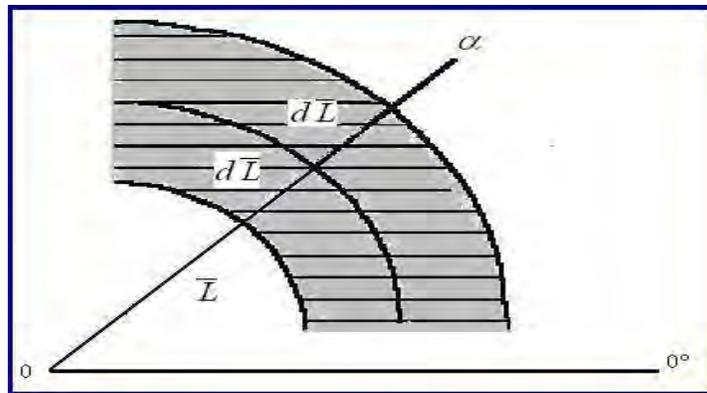


Figura 3.5. Espacio definido por la tolerancia lineal.

2. El ángulo que se forma entre la línea que une los dos puntos del par de datos y la dirección 0° debe incluirse entre $\alpha - d\alpha$ y $\alpha + d\alpha$ (figura 3.6).

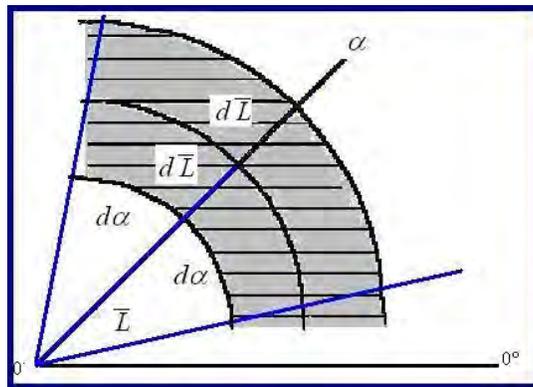


Figura 3.6. Espacio definido por la tolerancia lineal y angular.

3.- La distancia entre el segundo punto del par de datos y la línea que define la dirección de cálculo del variograma no debe superar el ancho de banda¹⁸ (figura 3.7).

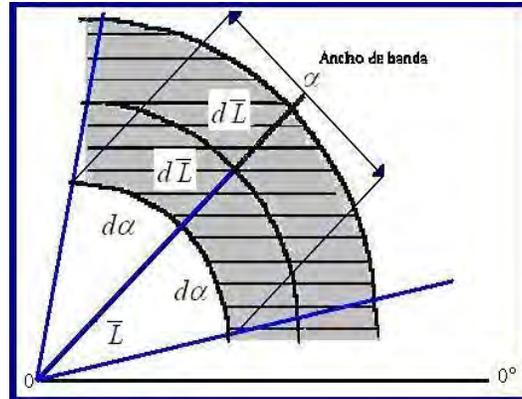


Figura 3.7. Espacio definido por las tolerancias y el ancho de banda.

Finalmente se representan gráficamente los valores de $\gamma[\bar{L}]$ en función de \bar{L} (figura 3.8).

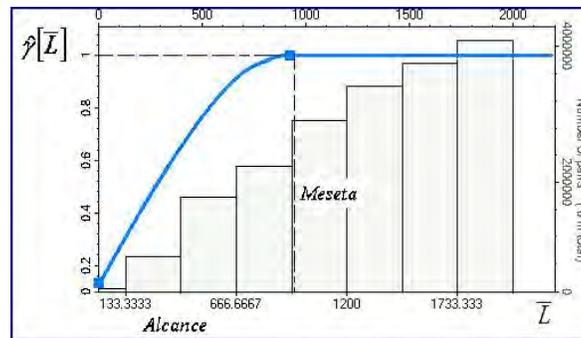


Figura 3.8. Forma típica del variograma.

El gráfico de $\gamma[\bar{L}]$ tiene las siguientes características^{10, 21}.

- Pasa por el origen (para $\bar{L} = 0$, $\gamma[\bar{L}] = 0$)
- Es en general una función creciente de \bar{L} .

En la mayor parte de los casos $\gamma[\bar{L}]$ crece hasta cierto límite que recibe el nombre de meseta, en otros casos puede crecer indefinidamente. El comportamiento en el origen puede tener diferentes formas, las cuales son^{18, 19, 23} (figura 3.9):

- I. **Parabólico:** Caracteriza a una variable muy regular, la cual es continua y diferenciable
- II. **Lineal:** Caracteriza a una variable continua, pero no diferenciable, es decir, menos regular
- III. **Discontinuidad en el origen:** “efecto pepita”o “nugget”, es el caso en que $\gamma[\bar{L}]$ no tiende a cero cuando \bar{L} tiende a cero. Representa a una variable muy irregular
- IV. **Discontinuo puro:** Se le denomina también ruido blanco, representa el caso de mayor discontinuidad, el caso limite es la ausencia de estructura, donde los valores de dos puntos cualesquiera no tienen correlación alguna

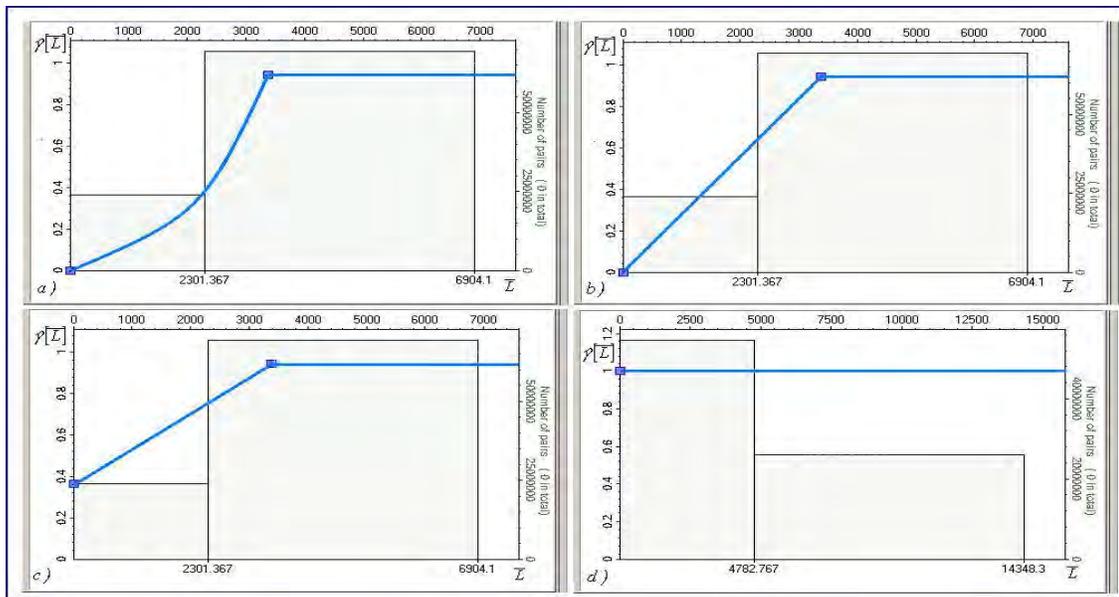


Figura 3.9. Comportamiento del variograma en el origen, (a) Parabólico, (b) Lineal, (c) Efecto Pepita, (d) Discontinuo puro.

3.5.2. PROBLEMAS COMUNES ENCONTRADOS EN EL CÁLCULO DE VARIOGRAMAS.

Algunos de los problemas más comunes que se discuten son¹⁰:

- **El valor idóneo del incremento \bar{L} :** Una inadecuada selección de \bar{L} puede proporcionar un variograma errático, aunque no se puede dar un criterio exacto o aproximado sobre cual es el mejor valor de \bar{L} , por lo que es recomendable recalcular $\gamma[\bar{L}]$ para distintos valores de \bar{L} , hasta encontrar una forma suavizada del variograma.
- **Distribuciones con valores extremos:** La existencia de valores extremos, altos o bajos, en una distribución, puede conducir a la obtención de un variograma fuertemente errático. En este caso la solución puede ser simple, eliminar los datos extremos, porque pueden ser poco representativos. En otros casos existen zonas geográficamente distintas y pueden tratarse de manera separada. Una herramienta útil para detectar valores extremos y encontrar el incremento adecuado puede ser, calculando la “nube de variogramas”, la cual consiste en representar los valores de $\frac{[X(\bar{u}_i + \bar{L}) - X(\bar{u}_i)]^2}{2}$ contra \bar{L} , para cada par posible de la información inicial.
- **La existencia de poblaciones mixtas:** Existen datos que pueden mostrar diferentes poblaciones, las cuales pueden mostrar diferencias estadísticas y se recomienda tratar las zonas de intereses por separado. En otros casos las poblaciones se presenten mezcladas geográficamente, en este caso una solución puede ser una transformación de datos (cambio de escala), con lo que se logra reducir la diferencia de los valores extremos.

Existen otras razones por las que los variogramas son erráticos²¹, las cuales son:

- 1) No hay suficientes muestras
- 2) Las muestras no son representativas del fenómeno
- 3) Las clasificaciones de las muestras no son válidas
- 4) El área estudiada es poco homogénea
- 5) Son necesarios pequeños o largos conjuntos de datos
- 6) Deben ser calculadas pequeñas o largas distancias
- 7) Deben ser calculadas, más o menos distancias
- 8) Son necesarias pequeñas tolerancias
- 9) Las muestras pueden tener localizaciones incorrectas
- 10) Los valores muestreados pueden ser erróneos

El problema fundamental en la obtención de un variograma correcto, es la elección adecuada de los intervalos de distancias para los cuales se calcula el variograma, de modo que en éstos la cantidad de pares encontrados sea suficiente desde el punto de vista estadístico.

3.5.3. MODELADO DE VARIOGRAMAS.

El modelado de variogramas incluye dos etapas fundamentales¹⁹, una vez construido el variograma experimental o empírico es necesario ajustar a éste un modelo teórico, con el objetivo de determinar los parámetros descriptivos del variograma que posteriormente serán usados en la estimación^{12, 21}.

3.5.3.1 PARÁMETROS DEL VARIOGRAMA.

Los parámetros del variograma caracterizan tres elementos importantes en la variabilidad de un atributo que son: la discontinuidad en el origen (existencia del efecto pepita), el valor máximo de variabilidad (meseta), y el área de influencia de correlación (alcance).

Dichos parámetros se presentan a continuación^{12, 21, 23}:

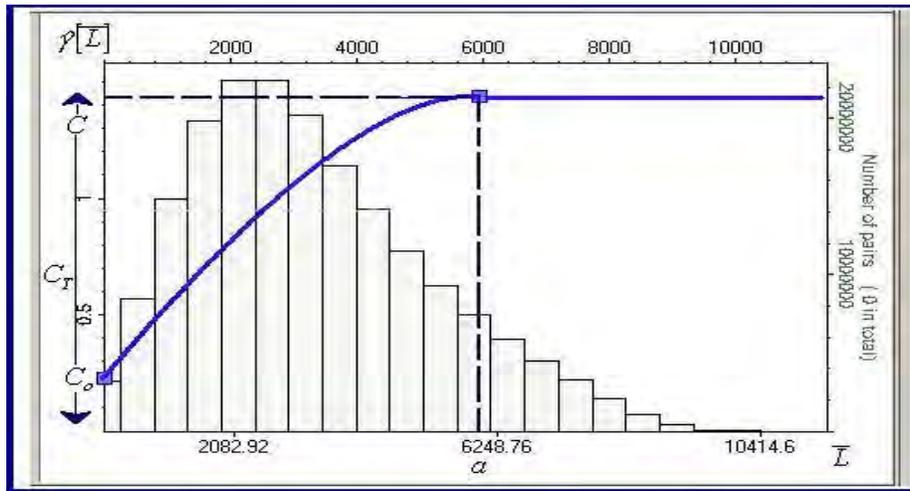


Figura 3.10. Parámetros del variograma.

- **Efecto Pepita (Nugget):** El variograma por definición es nulo en el origen, pero en la práctica las funciones que se obtienen pueden presentar discontinuidad en el origen, a esta discontinuidad se le llama efecto pepita, en inglés nugget effect. Este puede obtenerse al trazar una línea recta entre los primeros puntos del variograma empírico, extendiendo la misma hasta que se intercepte con el eje Y . Si esta intersección ocurre por debajo de cero, el valor asumido por este efecto es cero, pues los valores negativos de $\gamma[0]$ no tienen significado y no son comunes. El efecto pepita se representa como C_o .
- **Meseta (Sill):** Es el valor de $\gamma[\bar{L}]$ para el cual con el aumento de \bar{L} su valor permanece constante, se representa como $C_T = C + C_o$ y se denomina meseta. Puede obtenerse al trazar una línea paralela a la abscisa, ajustando la misma a los puntos de mayor valor del variograma, su valor se lee en la intersección de esta línea con la ordenada.

- **Alcance (Range):** La distancia \bar{L} para la cual las variables $X(\bar{u})$ y $X(\bar{u} + \bar{L})$ son independientes, se denomina alcance y se representa por a , es decir, las distancias para la cual los valores de la variable dejan de tener relación, o lo que es lo mismo, la distancia para la cual el variograma alcanza su meseta. El alcance siempre tiene valor positivo y puede obtenerse a partir de la intersección de las líneas descritas en los puntos anteriores, ese punto que se lee en la abscisa, es una fracción del propio alcance, fracción que se detallara posteriormente en la explicación de los modelos teóricos.

3.5.3.2. MODELOS TEÓRICOS DE VARIOGRAMAS.

Las características más importantes en el modelado de variogramas son:

1. Su comportamiento en el origen, el cual puede ser lineal, parabólico y con efecto pepita
2. La presencia o ausencia de meseta

Estos modelos son:

Modelo Efecto Pepita: Corresponde a un fenómeno netamente aleatorio (ruido blanco), sin correlación entre las muestras, cualquiera que sea la distancia que las separe, (figura 3.11), donde C representa el valor de la meseta.

El variograma al considerar el efecto pepita, se obtiene con las siguientes expresiones

$$\gamma[\bar{L}] = 0 \dots\dots\dots(3.5a)$$

Si $\bar{L} = 0$

Y

$$\gamma[\bar{L}] = C \dots\dots\dots(3.6a)$$

Si $|\bar{L}| > 0$

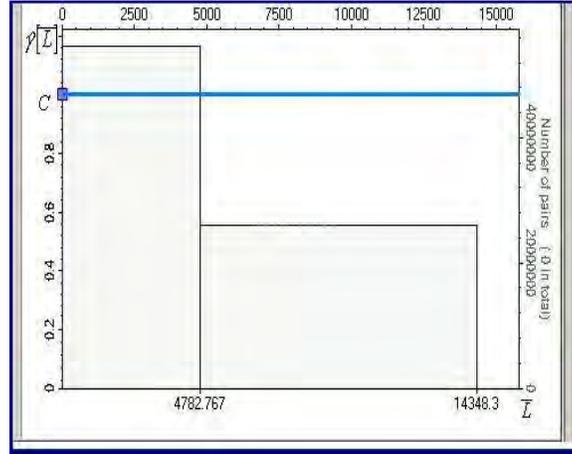


Figura 3.11. Modelo Efecto Pepita.

Modelo Esférico: Este modelo es probablemente el que más se utiliza, es una expresión polinomial simple cuya forma se representa en la figura 3.12, en donde se puede observar un crecimiento casi lineal y después a cierta distancia finita del origen se alcanza una estabilización (meseta). La tangente en el origen encuentra a la meseta en el punto de abscisa $\frac{2}{3}a$, donde a representa el valor del alcance.

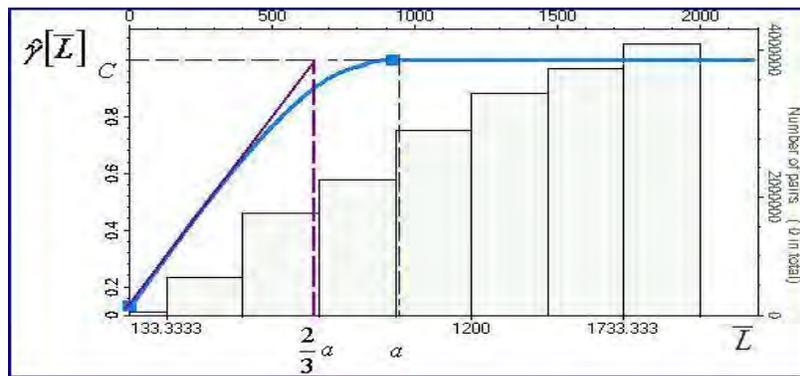


Figura 3.12. Modelo Esférico.

El variograma tomando en cuenta el modelo esférico se calcula mediante

$$\hat{\gamma}[\bar{L}] = C \left[\frac{3}{2} \left(\frac{\bar{L}}{a} \right) - \frac{1}{2} \left(\frac{\bar{L}}{a} \right)^3 \right]; \bar{L} \leq a \dots\dots\dots(3.7a)$$

O

$$\hat{\gamma}[\bar{L}] = C; \bar{L} > a \dots\dots\dots(3.7b)$$

Modelo Exponencial: Este modelo a diferencia del esférico crece inicialmente más rápido y después se estabiliza de forma asintótica (figura 3.13). Como la meseta no se alcanza a una distancia finita, se usa un valor con fines prácticos denominado “alcance efectivo” o “alcance práctico” a' , el cual se obtiene en el punto de abscisa para el cual el modelo obtiene el 95% de la meseta, con un valor $a' = 3a$, donde a es el parámetro de escala. La tangente en el origen encuentra a la meseta en el punto $a = \frac{1}{3} a'$.

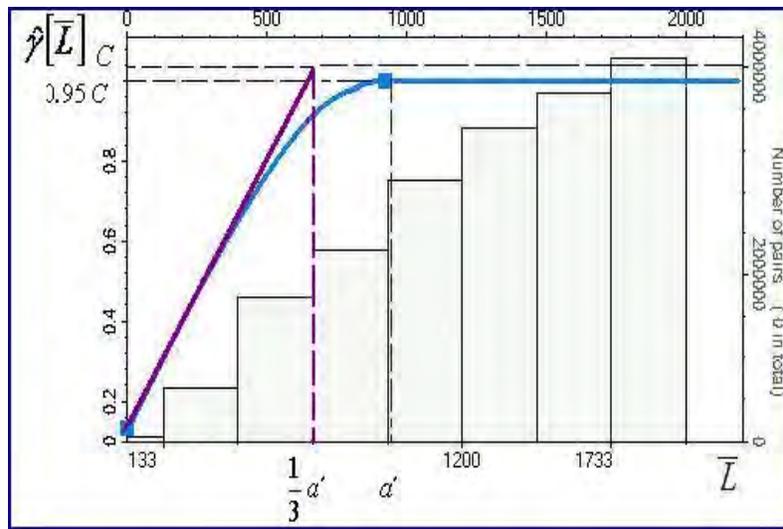


Figura 3.13. Modelo Exponencial.

La expresión para calcular el variograma con base en el modelo exponencial es

$$\gamma[\bar{L}] = C \left[1 - \text{Exp} \left(-\frac{|\bar{L}|}{a} \right) \right]; |\bar{L}| > 0 \dots\dots\dots(3.8)$$

Modelo Gaussiano: Este es un modelo extremadamente continuo (figura 3.14), inicialmente presenta un comportamiento parabólico en el origen, después al igual que en el modelo Exponencial se alcanza la meseta de forma asintótica. El alcance práctico tiene un valor de $a' = 1.73a$, que es el valor de la abscisa donde se alcanza el 95% de la meseta.

El variograma tomando en cuenta el modelo Gaussiano se determina con la formula

$$\gamma[\bar{L}] = C \left[1 - \text{Exp} \left(-\frac{|\bar{L}|^2}{a^2} \right) \right]; |\bar{L}| > 0 \dots\dots\dots(3.9)$$

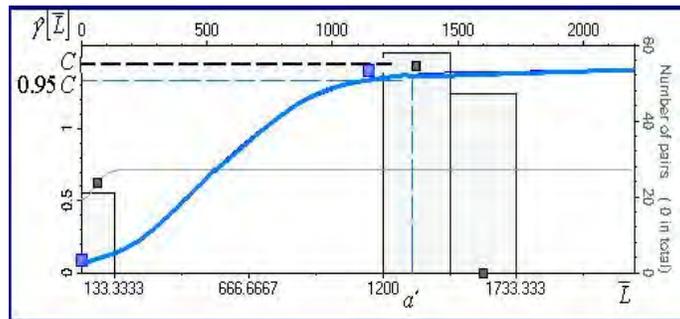


Figura 3.14. Modelo Gaussiano.

Modelo Potencial o en Potencia: Este es un modelo sin meseta, su forma se representa en la figura 3.15, para valores de α correspondientes a $0.5, 1.0$ y 1.5 .

El variograma puede calcularse como

$$\gamma[\bar{L}] = |\bar{L}|^\alpha; \alpha \in]0,2[\dots\dots\dots(3.10)$$

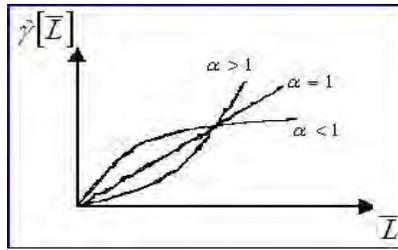


Figura 3.15. Modelo con función Potencia.

Para el valor de $\alpha = 1$ en el modelo anterior se obtiene el modelo lineal, el cual no tiene ni meseta ni alcance. Para efectos prácticos, muchos programas informáticos denotan la pendiente del modelo lineal con la relación $\frac{C}{a}$ (figura 3.16).

$$\gamma[\bar{L}] = \bar{L} \frac{C}{a} \dots\dots\dots(3.11)$$

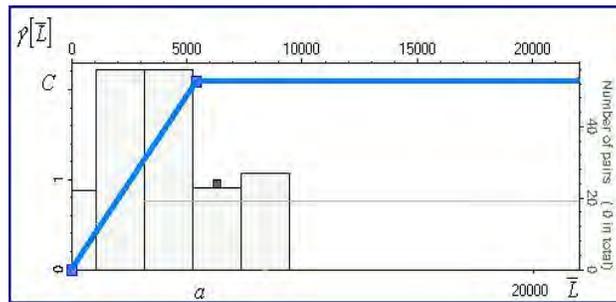


Figura 3.16. Modelo Lineal.

En esta sección se presentan los modelos más usados en la práctica, aunque cabe señalar, que existen otros modelos que emplean ampliamente en el sistema geoestadístico.

3.5.3.3. VALIDACIÓN DEL MODELO TEÓRICO.

Debido a que el ajuste de los modelos teóricos al variograma experimental, se realiza de forma visual o interactiva, variando los valores C_0 ,

(efecto de pepita), $C + C_o$ (meseta) y a (alcance), hasta coincidir con los parámetros que mejor se ajustan, es conveniente validar el modelo seleccionado, así como los parámetros meseta y alcance que se escogen. Al respecto algunos autores como Journel y Huijbregts¹⁸, y Deutsch y Journel²³, discuten la validación cruzada.

3.5.3.4. VALIDACIÓN CRUZADA.

El método de validación cruzada ha sido ampliamente utilizado para evaluar el grado de bondad de un modelado de variogramas y reconocido como un método óptimo de estimación de sus parámetros. La operación de validar un variograma teórico ajustado a uno experimental siempre toma mucho tiempo, éste se considera como el último de los pasos importantes del análisis de variabilidad, debido a que una vez obtenido este resultado, será utilizado en la estimación por Kriging en cualquiera de sus variantes^{10, 18, 19, 22, 23}. En el capítulo cuatro la validación cruzada se explicara con mayor detalle.

3.5.4. ANÁLISIS DE ANISOTROPÍA.

El variograma describe las características de continuidad espacial de la variable regionalizada en una dirección, pero este comportamiento puede variar según la dirección que se analice. Se exige por este motivo un análisis del comportamiento de la continuidad en distintas direcciones, conocido en la literatura Geoestadística como análisis de anisotropía^{21, 23}.

Cuando el variograma calculado en diferentes direcciones (norte-sur, este-oeste, y en direcciones intermedias de 45° o de 22.5°, con tolerancia de 22.5°), muestra similar comportamiento, se dice que el fenómeno es isótropo, cuando muestran diferentes comportamientos es anisótropo. Los tipos de anisotropías más comunes son la geométrica y la zonal^{10, 21, 23}.

- **Anisotropía Geométrica:** Está presente cuando los variogramas en diferentes direcciones, tienen la misma meseta pero distintos alcances (figura 3.17a).
- **Anisotropía Zonal:** Está presente cuando los variogramas en diferentes direcciones, tiene diferentes mesetas y diferentes alcances (figura 3.17b).

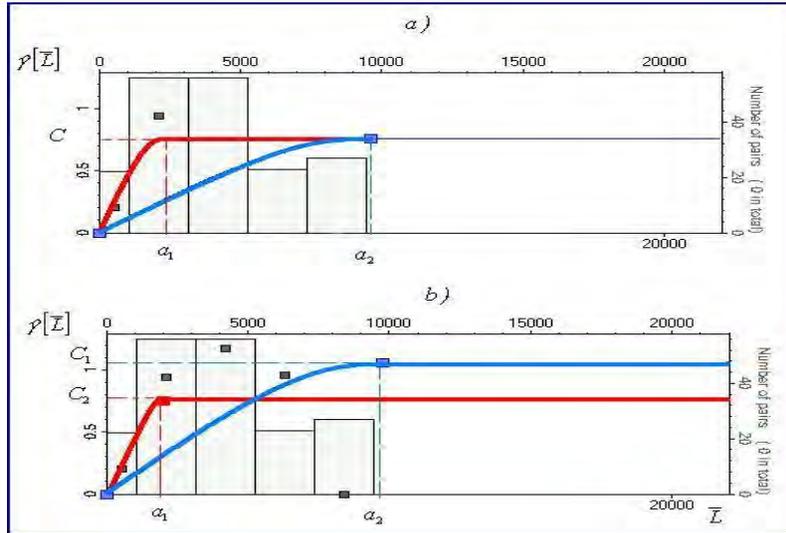


Figura 3.17. (a) Anisotropía Geométrica. (b) Anisotropía Zonal.

3.5.4.1. EFECTO PROPORCIONAL.

Quando en el cálculo del variograma se detecta que existe una relación lineal entre el valor medio de las muestras usadas en el cálculo de cada $\gamma[L]$ y la desviación estándar correspondiente, se dice que existe un efecto proporcional (heteroscedasticidad). Este efecto se puede detectar al graficar los valores de la media contra la desviación estándar, es decir, que el coeficiente de variación sea aproximadamente constante, lo cual ocurre cuando los datos presentan una distribución Log normal (figura 3.18). La solución a este problema consiste en dividir cada valor del variograma local por el cuadrado de la media local, y obtener lo que se conoce como variograma relativo.

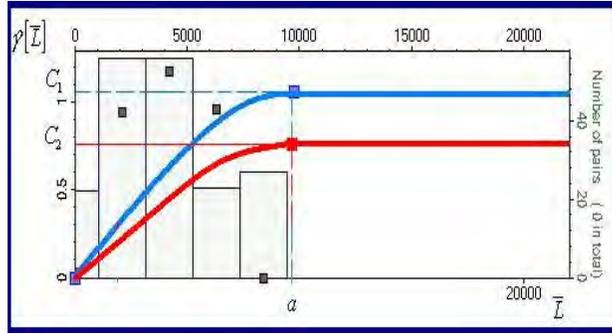


Figura 3.18. Efecto Proporcional.

El variograma relativo en términos de $X(\bar{u}_i)$ y \bar{L} se puede escribir como

$$\hat{F}[\bar{L}] = \frac{1}{2Np(\bar{L})} \sum_{i=1}^{Np(\bar{L})} \frac{[X(\bar{u}_i) - X(\bar{u}_i + \bar{L})]^2}{[(X(\bar{u}_i) - X(\bar{u}_i + \bar{L}))/2]^2} \dots\dots\dots(3.12)$$

Y puede ser calculado usando los pasos anteriormente presentados para el cálculo de los variogramas tradicionales.

3.5.4.2. PROBLEMAS EN EL MODELADO DE VARIOGRAMAS.

Se analizan los siguientes casos que tienen relación con los problemas más comunes al modelar variogramas²¹.

- I. **La anisotropía geométrica está presente:** Indica que los variogramas direccionales tienen la misma meseta pero diferentes alcances, esto puede ser corregido a través de una transformación lineal de coordenadas que permita reducir una elipse a un círculo

- II. **La anisotropía zonal está presente:** Indica que tanto las mesetas como los alcances son diferentes para los variogramas direccionales, lo cual puede corregirse separando el variograma en el componente isótropo horizontal, y el anisótropo vertical

- III. **La tendencia de los datos está presente:** Indica que los valores medidos aumentan o disminuyen dramáticamente en la zona de estudio con el aumento de la distancia. Esto puede ser resuelto al aplicar polinomios a la ecuación del variograma, es decir, aplicar un análisis de tendencia
- IV. **El efecto proporcional está presente:** Indica que la desviación estándar local es proporcional al cuadrado de la media local y que los datos presentan una distribución Log normal, lo anterior puede resolverse al dividir cada valor del variograma local por el cuadrado de la media local, es decir, usar variogramas relativos
- V. **Existencia de estructuras anidadas:** Indica que diferentes procesos operan a diferentes escalas, por ejemplo, los siguientes:
- A muy pequeñas distancias, la variabilidad puede estar presente debido a cambios en los que se presenta el fenómeno.
 - A pequeñas distancias la variabilidad en un fenómeno puede estar presente debido a errores.
- VI. **Existencia de efecto hueco:** Indica que muy pocos pares están disponibles para la comparación a una distancia específica. Lo anterior puede resolverse recuperando más casos para la distancia definida.
- VII. **La periodicidad está presente:** Indica que el comportamiento del variograma repite por sí mismo periodicidades, por ejemplo:
- El valor de la meseta puede aumentar o disminuir sistemáticamente. Esto puede ser resuelto si es un problema real y no un antifaz del análisis.

3.6. APLICACIÓN PRÁCTICA.

El objetivo de este ejercicio es el de describir como se determinan algunos parámetros óptimos para la construcción de un mapa de variograma con ayuda de un variograma experimental. Adicionalmente se ejemplifica el ajuste de un modelo de variograma experimental a un modelo teórico. Se asume la utilización del sistema métrico de unidades en el cálculo de todos los parámetros involucrados.

Un atributo sísmico es una característica muy particular de sísmica que permite mejorar la interpretación de cierta información que en la sísmica tradicional o global puede ser muy suave o fina. Los atributos pueden dividirse en cuatro colecciones con tareas específicas, estas son:

- Métodos de procesamiento de señales
- Trazas de complejos atributos
- Atributos estructurales
- Atributos estratigráficos

Dentro de los atributos estratigráficos se encuentra la impedancia acústica, que se calcula a partir de registros de densidad o sísmicos y mide la facilidad con la que las ondas sísmicas pueden pasar a través del subsuelo. Se calcula como la densidad de la formación multiplicada por la velocidad de onda acústica.

Para realizar este ejercicio se emplearon 41 datos de impedancia acústica que permitieron la construcción de distintos mapas de variogramas y variogramas experimentales. Los datos se ilustran en la siguiente figura

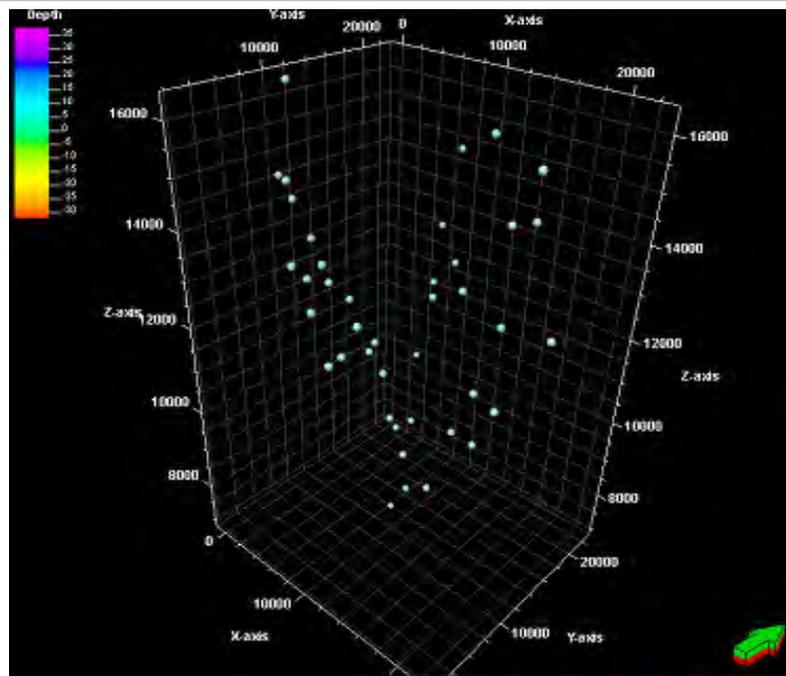


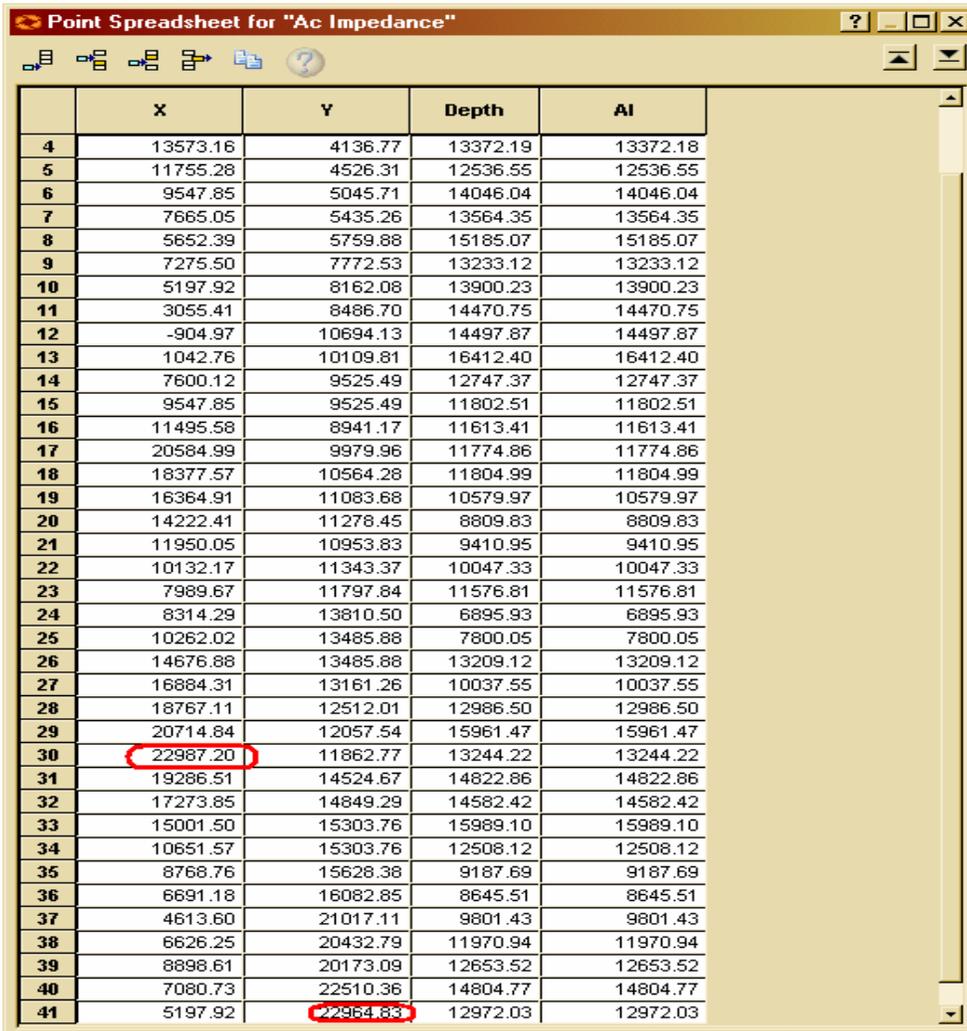
Figura 3.19. Datos de impedancia acústica distribuidos en 3D.

Se eligen tres distancias máximas diferentes de búsqueda, cada una con tres intervalos distintos, con una tolerancia angular común, para visualizar en total nueve mapas de variograma.

Para determinar los parámetros más representativos del variograma experimental, se elige el mapa con mayor definición en cuanto a visualización mediante una simple inspección.

Con base en el conjunto de datos de impedancia acústica, y los valores más grandes en cuanto a ubicación de coordenadas x, y de los mismos, se eligen las tres distancias máximas.

El conjunto de datos de impedancia acústica se muestra a continuación



	X	Y	Depth	AI
4	13573.16	4136.77	13372.19	13372.18
5	11755.28	4526.31	12536.55	12536.55
6	9547.85	5045.71	14046.04	14046.04
7	7665.05	5435.26	13564.35	13564.35
8	5652.39	5759.88	15185.07	15185.07
9	7275.50	7772.53	13233.12	13233.12
10	5197.92	8162.08	13900.23	13900.23
11	3055.41	8486.70	14470.75	14470.75
12	-904.97	10694.13	14497.87	14497.87
13	1042.76	10109.81	16412.40	16412.40
14	7600.12	9525.49	12747.37	12747.37
15	9547.85	9525.49	11802.51	11802.51
16	11495.58	8941.17	11613.41	11613.41
17	20584.99	9979.96	11774.86	11774.86
18	18377.57	10564.28	11804.99	11804.99
19	16364.91	11083.68	10579.97	10579.97
20	14222.41	11278.45	8809.83	8809.83
21	11950.05	10953.83	9410.95	9410.95
22	10132.17	11343.37	10047.33	10047.33
23	7989.67	11797.84	11576.81	11576.81
24	8314.29	13810.50	6895.93	6895.93
25	10262.02	13485.88	7800.05	7800.05
26	14676.88	13485.88	13209.12	13209.12
27	16884.31	13161.26	10037.55	10037.55
28	18767.11	12512.01	12986.50	12986.50
29	20714.84	12057.54	15961.47	15961.47
30	22987.20	11862.77	13244.22	13244.22
31	19286.51	14524.67	14822.86	14822.86
32	17273.85	14849.29	14582.42	14582.42
33	15001.50	15303.76	15989.10	15989.10
34	10651.57	15303.76	12508.12	12508.12
35	8768.76	15628.38	9187.69	9187.69
36	6691.18	16082.85	8645.51	8645.51
37	4613.60	21017.11	9801.43	9801.43
38	6626.25	20432.79	11970.94	11970.94
39	8898.61	20173.09	12653.52	12653.52
40	7080.73	22510.36	14804.77	14804.77
41	5197.92	22964.86	12972.03	12972.03

Figura 3.20. Datos de impedancia acústica: coordenadas x,y,z.

Las cantidades que están encerradas con rojo son las de mayor valor, mismas que se toman como referencia para la elección de las distancias máximas. Las distancias máximas se escogen de acuerdo al siguiente criterio:

- 1.- L_{\max} en la dirección $x = 14,000\ m$ y L_{\max} en la dirección $y = 14,000\ m$ que son aproximadamente la mitad de la distancia máxima en cuanto a ubicación
- 2.- L_{\max} en la dirección $x = 26,000\ m$ y L_{\max} en la dirección $y = 26,000\ m$ que son un tanto mayores que la distancia máxima en cuanto a ubicación

- 3.- L_{\max} en la dirección $x = 30,000\text{ m}$ y L_{\max} en la dirección $y = 30,000\text{ m}$ que son considerablemente mayores que la distancia máxima en cuanto a ubicación

La tolerancia angular es de $d\alpha = 30^\circ$ y los intervalos para cada distancia máxima son los siguientes:

$$\bar{L} = 10\text{ m}, \quad \bar{L} = 20\text{ m}, \quad \bar{L} = 30\text{ m}$$

A continuación se muestran los resultados obtenidos en los nueve mapas construidos con el software comercial PETREL 2004^{MR}.

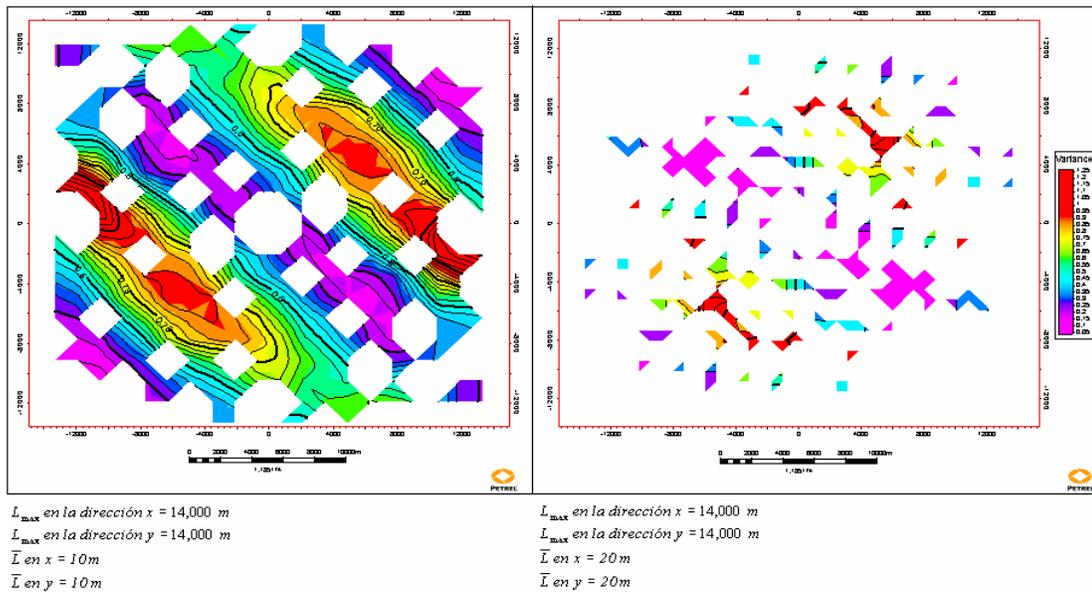
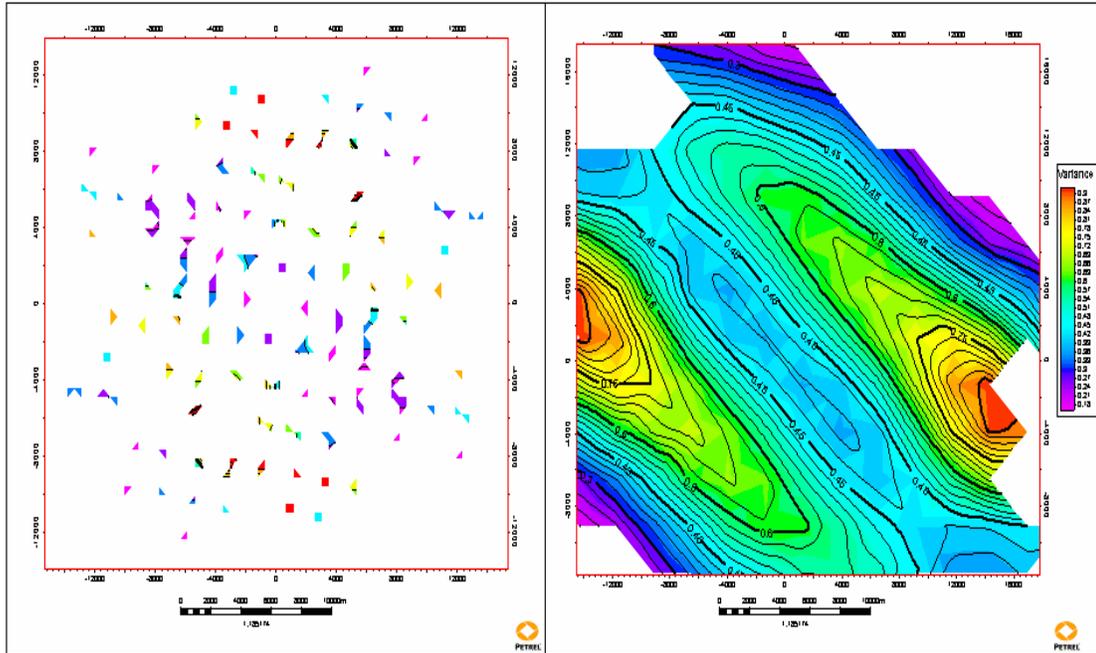


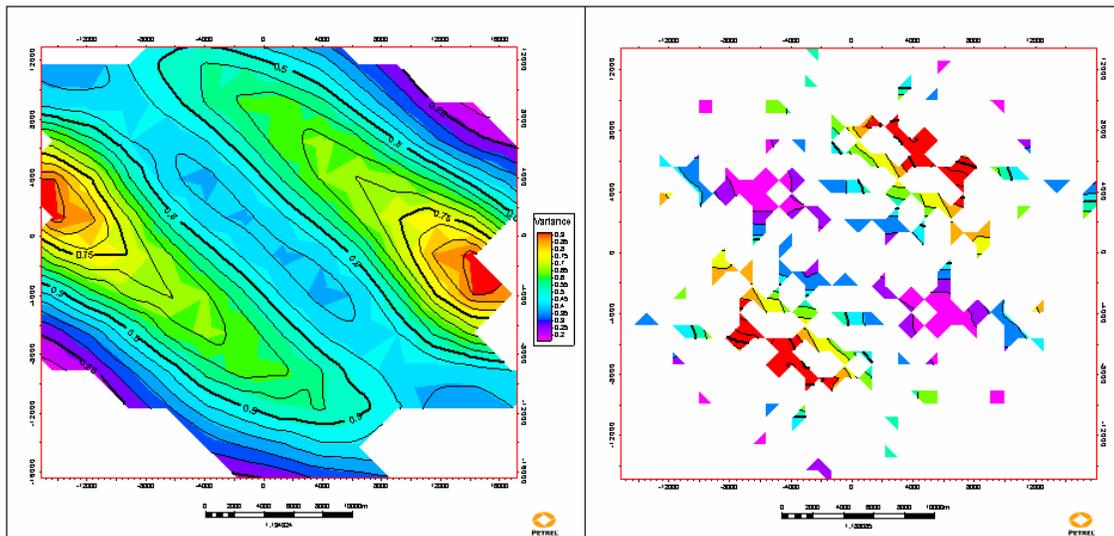
Figura 3.21a. Mapas de Variograma.



L_{\max} en la dirección $x = 14,000$ m
 L_{\max} en la dirección $y = 14,000$ m
 \bar{L} en $x = 30$ m
 \bar{L} en $y = 30$ m

L_{\max} en la dirección $x = 26,000$ m
 L_{\max} en la dirección $y = 26,000$ m
 \bar{L} en $x = 10$ m
 \bar{L} en $y = 10$ m

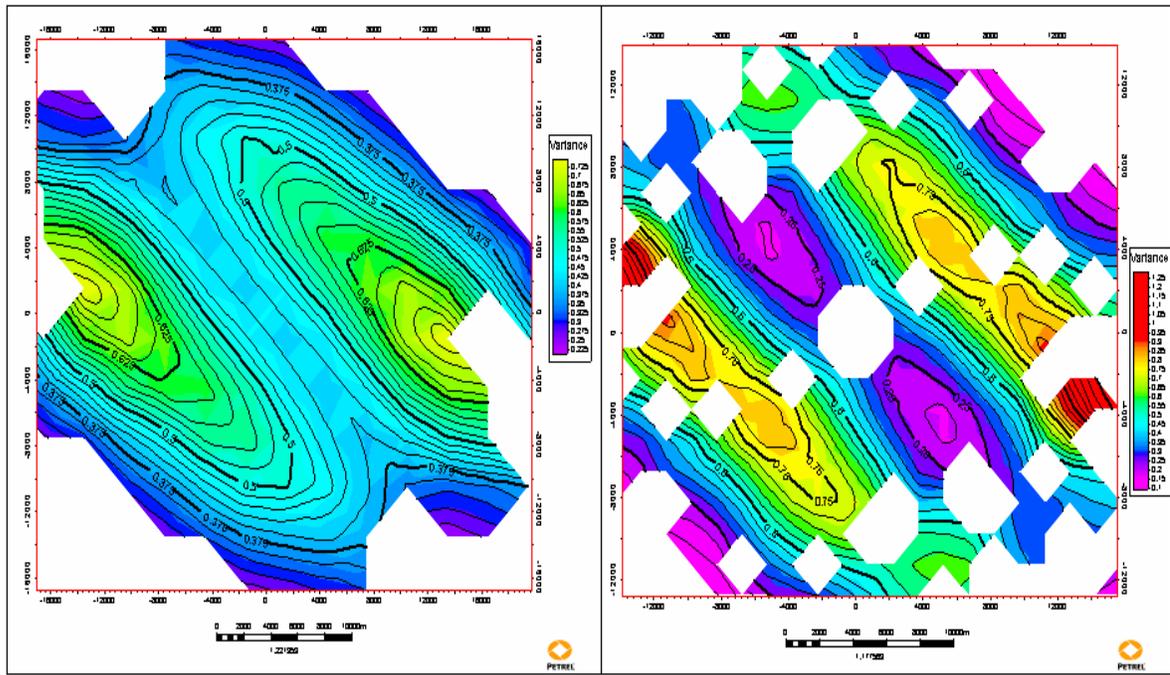
Figura 3.21b. Mapas de Variograma.



L_{\max} en la dirección $x = 26,000$ m
 L_{\max} en la dirección $y = 26,000$ m
 \bar{L} en $x = 20$ m
 \bar{L} en $y = 20$ m

L_{\max} en la dirección $x = 26,000$ m
 L_{\max} en la dirección $y = 26,000$ m
 \bar{L} en $x = 30$ m
 \bar{L} en $y = 30$ m

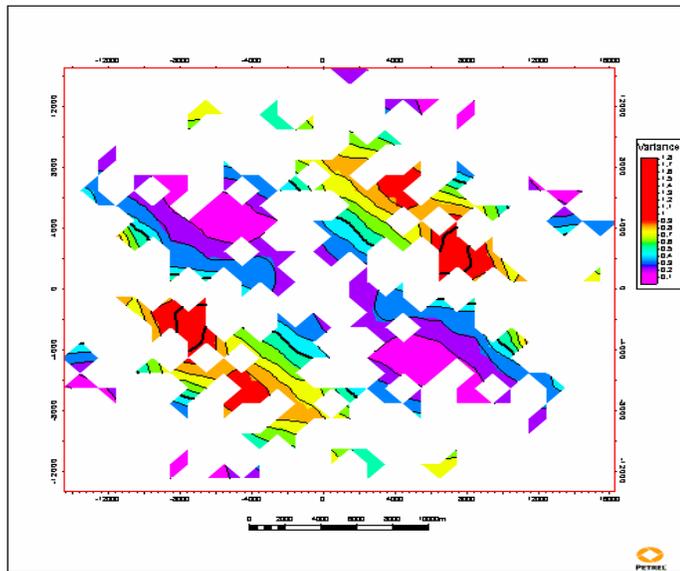
Figura 3.21c. Mapas de Variograma.



L_{max} en la dirección $x = 30,000\ m$
 L_{max} en la dirección $y = 30,000\ m$
 \bar{L} en $x = 10\ m$
 \bar{L} en $y = 10\ m$

L_{max} en la dirección $x = 30,000\ m$
 L_{max} en la dirección $y = 30,000\ m$
 \bar{L} en $x = 20\ m$
 \bar{L} en $y = 20\ m$

Figura 3.21d. Mapas de Variograma.



L_{max} en la dirección $x = 30,000\ m$
 L_{max} en la dirección $y = 30,000\ m$
 \bar{L} en $x = 30\ m$
 \bar{L} en $y = 30\ m$

Figura 3.21e. Mapa de Variograma.

Los mapas en los que se observan “huecos” se descartan, debido a su inestabilidad, por consiguiente solo se analizan tres mapas.

Al analizar los mapas de variograma se determinan las zonas de anisotropía menor representada por la línea azul y anisotropía mayor representada por la línea roja.

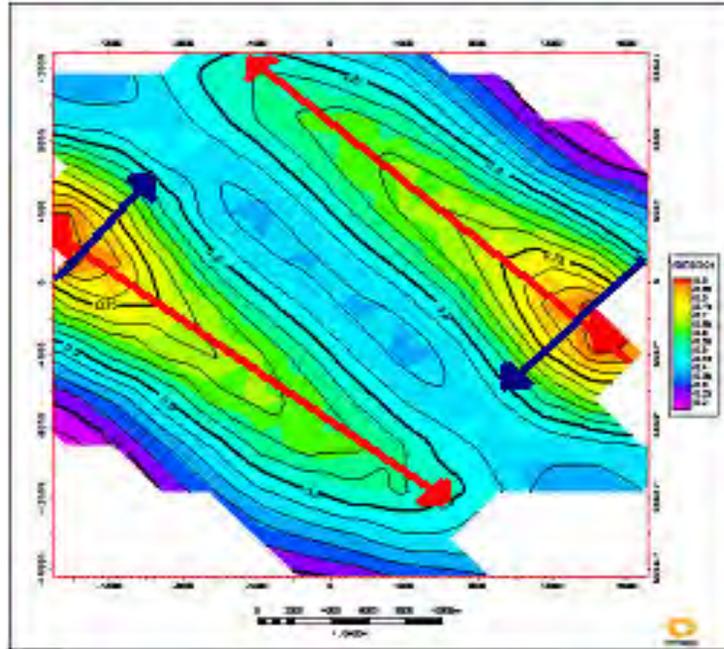


Figura 3.22. Determinación de zonas de anisotropía.

Del análisis de estos mapas con PETREL 2004^{MR}, se determina la orientación de anisotropía mayor y menor, la cual es la misma para los tres mapas. Los valores son:

Orientación de la zona de mayor anisotropía = -51°

Orientación de la zona de menor anisotropía = 44°

Además como segundo criterio para la determinación de los parámetros óptimos que se obtienen de un mapa de variograma, se utilizan las graficas del variograma experimental, realizadas en PETREL 2004^{MR}, en las que fue

necesario encontrar su forma más suavizada de acuerdo a los datos de distancia máxima e intervalos antes descritos.

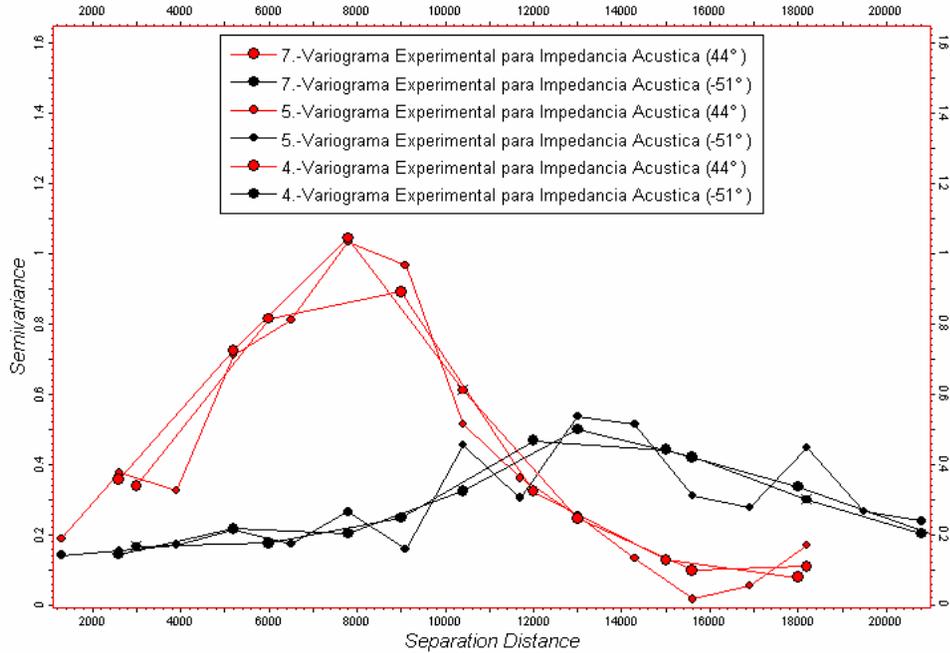


Figura 3.23. Variogramas experimentales de los tres mapas de variograma con mayor definición en cuanto a visualización.

De los tres mapas restantes se obtiene lo siguiente:

El mapa con los datos (ver figura 3.21d)

$$L_{\max} \text{ en la dirección } x = 30,000 \text{ m}$$

$$L_{\max} \text{ en la dirección } y = 30,000 \text{ m}$$

$$\bar{L} \text{ en } x = 10 \text{ m}$$

$$\bar{L} \text{ en } y = 10 \text{ m}$$

Queda descartado debido a que hay muy pocos puntos en la grafica del variograma experimental como se muestra en la siguiente figura

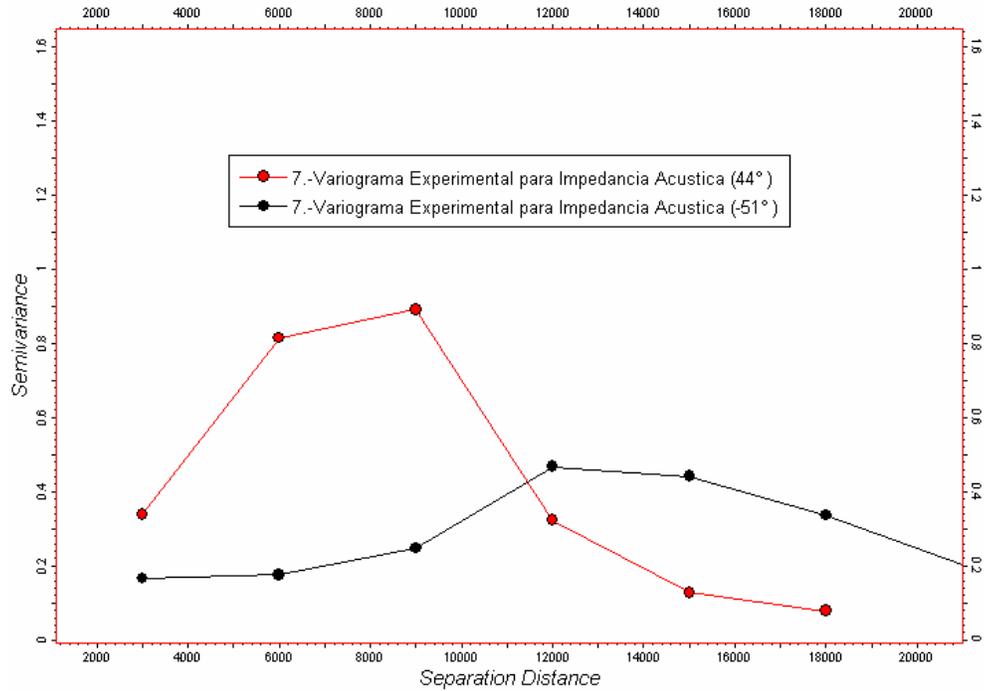


Figura 3.24. Variograma experimental para una distancia maxima x,y de 30,000 m y un intervalo x,y de 10 m.

El mapa con los datos (ver figura 3.21b)

L_{\max} en la dirección $x = 26,000\ m$

L_{\max} en la dirección $y = 26,000\ m$

\bar{L} en $x = 10\ m$

\bar{L} en $y = 10\ m$

A pesar de tener una forma bien definida, se descarta también por la falta de puntos graficados en el variograma experimental como se muestra en la siguiente figura

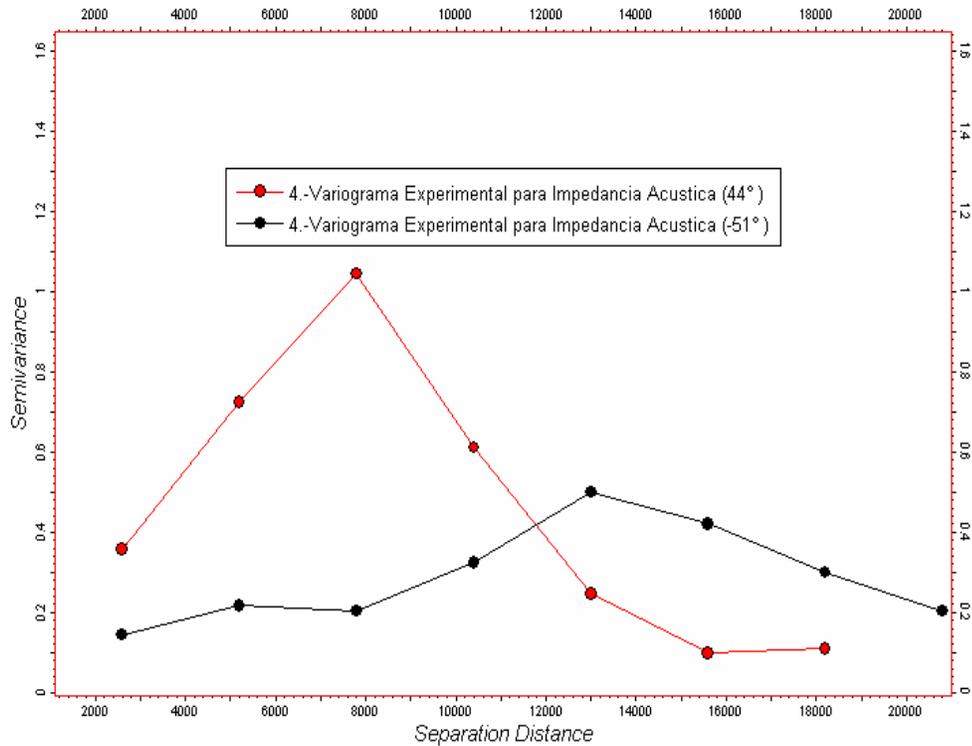


Figura 3.25. Variograma experimental para una distancia maxima x,y de 26,000 m y un intervalo x,y de 10 m.

Finalmente se elige el mapa con mayor definición y el variograma experimental con la forma más suavizada con el mayor número de puntos en la grafica a comparación de los dos variogramas anteriores.

Por lo tanto los datos óptimos que permiten construir un mapa de variograma son

$$L_{\max} \text{ en la dirección } x = 26,000 \text{ m}$$

$$L_{\max} \text{ en la dirección } y = 26,000 \text{ m}$$

$$\bar{L} \text{ en } x = 20 \text{ m}$$

$$\bar{L} \text{ en } y = 20 \text{ m}$$

En la siguiente grafica se ilustran los variogramas experimentales para las zonas de mayor y menor anisotropía, así como el variograma ajustado a un modelo teórico de tipo esférico.

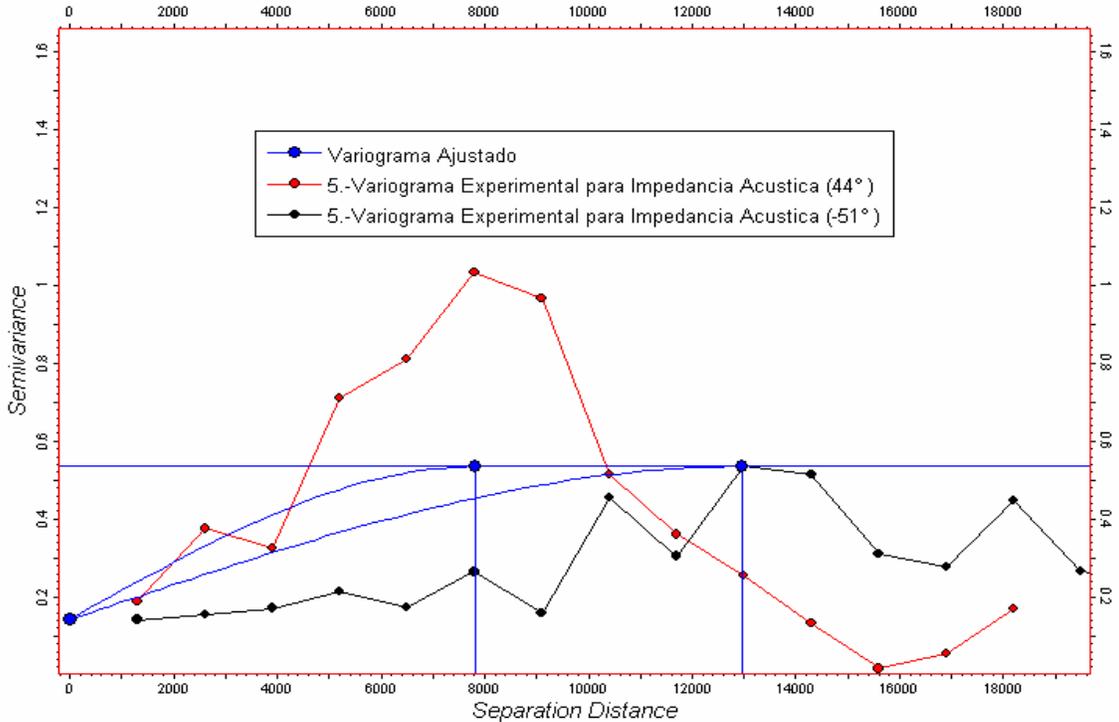


Figura 3.26. Variograma experimental ajustado a un modelo esférico para una distancia máxima x,y de 26,000 m y un intervalo x,y de 20 m.

Los parámetros obtenidos de los variograma ya ajustado son.

$$Meseta = 0.54$$

$$Nugget = 0.14$$

$$Alcance \text{ M\u00ednimo (Menor anisotrop\u00eda)} = 7,820 \text{ m}$$

$$Alcance \text{ M\u00e1ximo (Mayor anisotrop\u00eda)} = 12,965 \text{ m}$$

Se determina la existencia de anisotropía geométrica ya que la meseta es la misma pero el alcance es distinto.

Los datos de meseta, nugget, alcance máximo y mínimo se utilizan como parámetros de entrada en la construcción de un modelado petrofísico, ya sea de tipo determinístico o estocástico como se vera en el siguiente capitulo.

CAPITULO 4

MODELADO NUMÉRICO DE YACIMIENTOS

4.1. INTRODUCCIÓN.

En este capítulo se abordan los aspectos y ecuaciones más importantes que se emplean para desarrollar un modelado numérico de yacimientos. El modelado por Kriging y el modelado por simulación Gaussiana secuencial son dos de las técnicas más populares y eficientes para conseguir dicho objetivo, ya que actualmente los simuladores empleados en la industria petrolera, utilizan dichos procedimientos para obtener modelados petrofísicos. Adicionalmente se menciona bajo que condiciones se pueden emplear estos modelos, así como las ventajas y desventajas de su utilización. No obstante existe otro tipo de metodologías para obtener modelados numéricos, las cuales se mencionan con la intención de hacer notar que los modelos mencionados en este trabajo no son únicos y que estos dependen de la información con que se alimente a dichos modelos así como del tipo de resultados que se espere de los mismos.

4.2. MODELADO POR KRIGING.

El modelado por Kriging es una técnica convencional de interpolación que emplea procedimientos de estimación lineal para la determinación de valores en localidades donde no es posible obtener muestras con base en la media y la varianza de los datos disponibles. Como se mencionó en el capítulo dos el término Kriging se debe al Geoscientista sudafricano Danny G. Krige, quien aplicó dicha técnica en la Minería, más tarde los fundamentos y validaciones matemáticas fueron propuestas por Matheron¹.

En principio, el modelado por Kriging asume que el valor en la localidad donde no es posible obtener muestras se estima mediante

$$X^*(\bar{u}_o) = \sum_{i=1}^n \lambda_i X(\bar{u}_i) \dots\dots\dots(4.1)$$

Donde $X^*(\bar{u}_o)$ es el valor estimado en la localidad donde no es posible obtener muestras, $X(\bar{u}_i)$ es el valor de la muestra en la localidad vecina (\bar{u}_i), y λ_i es el “peso” o importancia asignado (a) a el valor en dicha localidad. Todos los algoritmos de Kriging, usan la ecuación 4.1 con diferentes variaciones dependiendo de la aplicación particular.

El objetivo del modelado por Kriging, es el de calcular los pesos asignados a los puntos individuales que existen dentro de una vecindad. El valor de dichos pesos dependerá de las relaciones espaciales existentes entre la localidad donde no es posible obtener muestras y la relación de los valores vecinos.

Dichas relaciones se pueden obtener del modelado de variogramas, respetando las condiciones de estacionariedad de primer y segundo orden.

Antes de iniciar un modelado por Kriging se recomienda tomar en cuenta dos consideraciones básicas:

1) Determinar la vecindad entre los valores.

En la determinación de la vecindad de los valores, se tiene que tomar en cuenta lo siguiente:

Número de muestras.

Esta consideración se refiere a que si se emplea una gran cantidad de datos, el tamaño de la matriz de datos aumentara de igual manera, lo que implica un aumento en la memoria y tiempo de cálculo en el sistema de ecuaciones de acuerdo a los recursos computacionales disponibles. Por otro lado si escogen pocos valores, la estimación puede ser poco representativa. Por lo tanto es necesario mantener un balance que permita emplear suficientes muestras, lo cual puede resolverse mediante la división de zonas vecinas, en

secciones o cuadrantes, en donde se seleccione cierto número de puntos dentro de cada sección (ver figura 4.1).

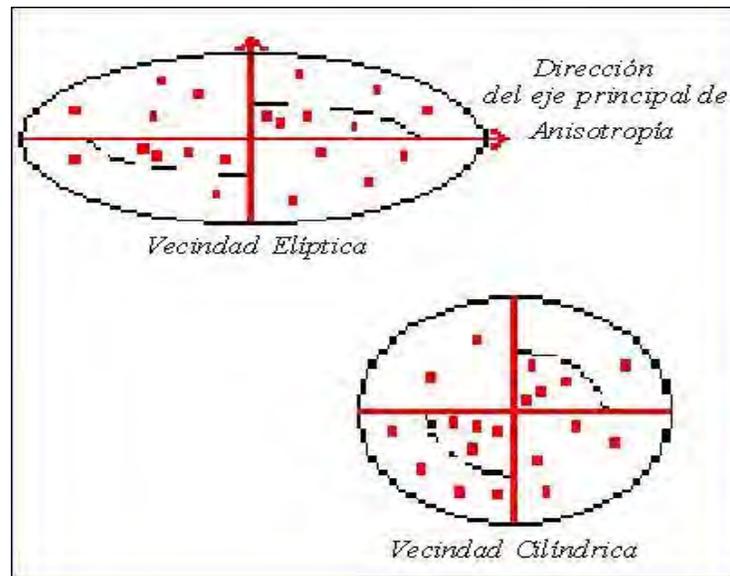


Figura 4.1. Determinación de tipos de vecindad.

Tamaño de la vecindad.

El tamaño de la vecindad debe ser lo suficientemente grande para incluir suficientes puntos; al mismo tiempo, debe ser lo suficientemente pequeño para que las muestras en los puntos seleccionados estén en la región de estacionariedad local.

Tipo de la vecindad.

El tipo de vecindad se rige en un principio, por los modelos del variograma. Deutsch y Journel² recomiendan lo siguiente:

Si se usa un modelo de variograma isotrópico en 2D, el tipo de vecindad debe ser circular. Si se usa un modelo de variograma anisotrópico en 2D, el tipo

de vecindad debe ser elíptica. La figura 4.1 adicionalmente ejemplifica vecindades de tipo circular y elíptica.

Si la continuidad espacial es descrita por un modelo de variograma isotrópico en 3D, el tipo de vecindad debe ser esférica. Si la continuidad espacial es descrita por un variograma anisotrópico en 3D, el tipo de vecindad debe ser elipsoidal.

También, una practica común es la de hacer coincidir la orientación del eje mayor de la elipse o elipsoide con la dirección principal de la continuidad espacial. Lo anterior asegura que la selección de los puntos de muestreo más cercanos, que tienen una relación espacial fuerte con la localidad donde no es posible obtener muestras, se realiza adecuadamente.

2) Realizar una validación cruzada.

En lo que respecta a la validación cruzada, esta involucra la estimación de valores en la localización donde se realiza muestreo por lo que la estimación puede compararse con valores muestreados. La versión más común para la validación cruzada consiste en remover un punto muestreado de los datos, y utilizando las muestras restantes, el valor de la variable desplazada, se calcula, considerándolo ahora como si este correspondiera a una variable en una localización donde no es posible obtener muestras³. Posteriormente el valor estimado se compara con el valor observado previamente. Este procedimiento se repite para todas las muestras, estimando el valor para cada una de las localizaciones. Cuando el proceso de validación cruzada ha concluido, se tienen dos valores ahora en todas las localizaciones muestreadas, un valor real u observado, y otro valor estimado. Al comparar el valor real con el valor estimado, es posible detectar problemas relacionados con el proceso de estimación.

Definiendo

$$e(\bar{u}_i) = x(\bar{u}_i) - x^*(\bar{u}_i) \dots\dots\dots(4.2)$$

Donde $e(\bar{u}_i)$ es el error de estimación en la localización (\bar{u}_i) , $x(\bar{u}_i)$ es el valor real $x^*(\bar{u}_i)$ es el valor estimado. Se deben examinar algunas observaciones respecto a $e(\bar{u}_i)$, $x(\bar{u}_i)$ y $x^*(\bar{u}_i)$, como parte del proceso de validación cruzada. A continuación se enuncian algunas:

Graficar el valor real vs el valor estimado.

- Cuando se tiene una propagación de error uniforme a lo largo de una línea a 45°, se tiene una indicación de que el número aproximado de valores se encuentran bajo o sobre estimados con respecto a la región de interés. Lo anterior indica un comportamiento ideal (ver la figura 4.2a).
- Cuando la propagación del error (desviación respecto a la línea recta) se incrementa conforme el valor verdadero se incrementa, existe una indicación de que el error depende de la magnitud de la muestra. Este tipo de comportamiento refleja la heteroscedasticidad del error en la varianza (ver la figura 4.2b).
- Cuando se presenta una acumulación considerable de la propagación del error por encima de la línea a 45°, el valor de las muestras es bajoestimado; y si al mismo tiempo se presenta acumulación por debajo de la línea a 45°, el valor de las muestra es sobrestimado. Este tipo de estimación se denomina estimación condicional parcial ideal (ver la figura 4.2c).

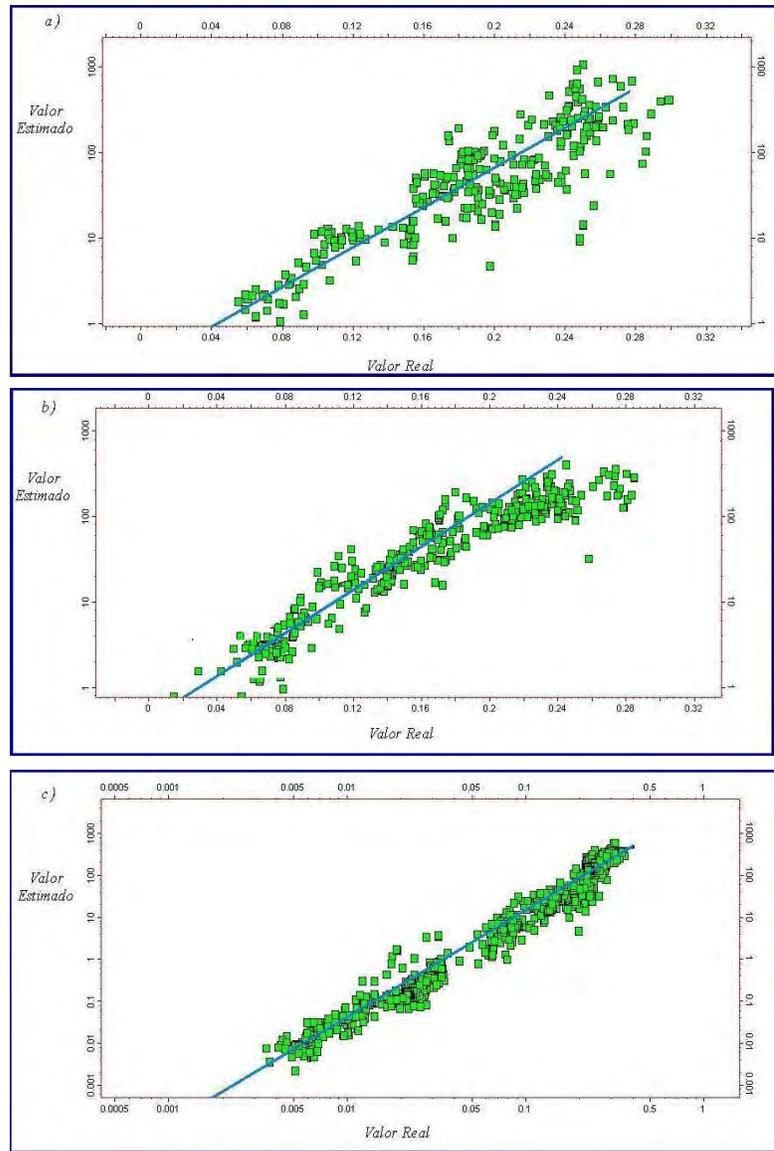


Figura 4.2. Valor real vs valor estimado durante el proceso de validación cruzada.

Las figuras 4.2b y 4.2c muestran características no deseables. La figura 4.2b puede reflejar que la determinación de la vecindad es demasiado amplia, resultado de la incorporación de datos que no se encuentran dentro de la región de estacionariedad local. La figura 4.2c puede reflejar la mezcla de dos regiones de estacionariedad, por lo tanto la suposición de estacionariedad deberá ser reexaminada para asegurar que se está incluyendo un adecuado conjunto de datos dentro de la región de interés.

Graficar el error de estimación $e(\bar{u})$ vs el valor estimado.

La figura 4.3a muestra el ejemplo donde el error está centrado alrededor de una línea denominada “cero” con igual propagación. Idealmente, para n muestras,

$$\frac{1}{n} \sum_{i=1}^n e(\bar{u}_i) \approx 0 \dots\dots\dots(4.3)$$

Lo que indica que la estimación es condicional imparcial. La misma propagación alrededor de la línea cero, indica la homoscedasticidad del error en la varianza, por lo tanto el error es independiente al valor estimado. En contraste, en la figura 4.3b, aunque la estimación es condicional imparcial, la magnitud del error se incrementa conforme el valor muestreado se incrementa. Esto es un reflejo del comportamiento heteroscedástico. Como ya se mencionó, el tamaño y la pendiente de la vecindad deben examinarse para asegurar que el error en la varianza tenga características homoscedásticas.

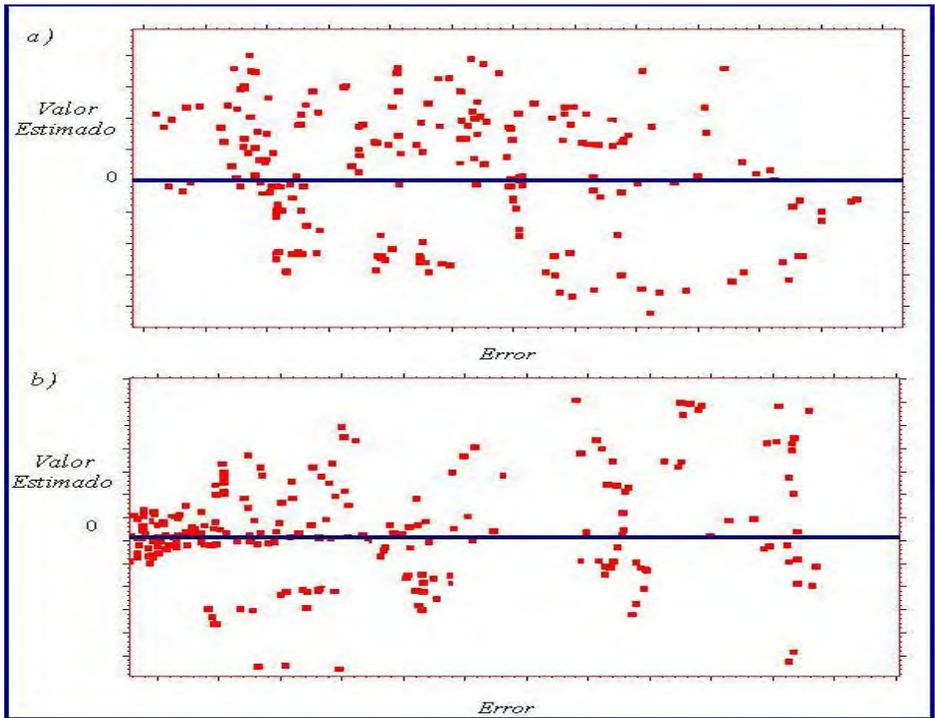


Figura 4.3. Error (real-estimado) vs valor real durante el proceso de validación cruzada.

En la práctica, la condición de homoscedasticidad es difícil de alcanzar; el error debe ser independiente de la estimación. En términos de magnitud absoluta, el error no debe cambiar.

La validación cruzada es un procedimiento de ensayo y error. Envuelve el ajuste de algunos parámetros operacionales tales como tamaño y tipo de vecindad determinada, así como la suposición de estacionariedad en una región, o la contemplación de un modelo de variograma para capturar tendencias. Estos parámetros se ajustan hasta que se obtienen los resultados esperados (por ejemplo condiciones de imparcialidad respecto a la estimación, error de homoscedasticidad, y falta de relaciones espaciales entre los errores) lo más cercano posible a las condiciones ideales. Debido a que la validación cruzada es un procedimiento de ensayo y error, este no depende de la demanda computacional, y con unas cuantas iteraciones, los resultados deseados pueden obtenerse.

Una característica común en todas las técnicas para modelar por Kriging es la creación de figuras aparentemente suavizadas.

El modelado por Kriging es una técnica de filtrado en donde las pequeñas variaciones se eliminan y las características a gran escala se conservan, por lo que la varianza de las mallas es minimizada mediante técnicas de estimación imparcial, es así como es posible asignar un mejor peso para el agrupamiento de dos datos cercanos y visualizar de mejor manera la anisotropía.

Una gran dificultad que se observa en el modelado por Kriging es la inhabilidad para capturar adecuadamente las incertidumbres locales, así como su limitación por el manejo de grandes conjuntos de datos. Otro problema que se tiene es al no respetar la arquitectura del yacimiento, por ejemplo, cuando se presentan fallas.

A continuación se describen algunas técnicas para modelar por Kriging las cuales se realizan de acuerdo a una aplicación específica.

4.2.1. KRIGING SIMPLE.

El Kriging simple es la técnica para modelar por Kriging más simple, pero no la más practica; esta comienza con la suposición de que un valor en una localidad donde no pueden obtenerse muestras, puede estimarse por

$$X^*(\bar{u}_o) = \lambda_o + \sum_{i=1}^n \lambda_i X(\bar{u}_i) \dots\dots\dots(4.4)$$

Donde $X^*(\bar{u}_o)$ es el valor a estimar, $X(\bar{u}_i)$ es el valor de la muestra en la localidad vecina (\bar{u}_i) , y n es el número total de muestras seleccionadas dentro de la vecindad determinada, λ_i es el “peso” asignado a cada muestra, y λ_o es una constante.

Para estimar el valor de λ_i , se utilizan las técnicas de estimación imparcial para minimizar la varianza. La condición de imparcialidad enuncia que $E[X^*(\bar{u}_o) - X(\bar{u}_o)] = 0 \dots\dots\dots(4.5)$

Sustituyendo la ecuación 4.4 en la ecuación 4.5 se tiene

$$\lambda_o + \sum_{i=1}^n \lambda_i E[X(\bar{u}_i)] = E[X(\bar{u}_o)] \dots\dots\dots(4.6)$$

Asumiendo la condición de estacionariedad de primer orden, $E[X(\bar{u}_i)] = E[X(\bar{u}_o)]$, entonces puede escribirse

$$\lambda_o = E[X(\bar{u}_o)] - \sum_{i=1}^n \lambda_i E[X(\bar{u}_i)] \dots\dots\dots(4.7a)$$

Por lo que

$$\lambda_o = m \left[1 - \sum_{i=1}^n \lambda_i \right] \dots\dots\dots(4.7b)$$

La ecuación 4.6 se puede escribir como

$$E[X(\bar{u}_o)] = \sum_{i=1}^n \lambda_i E[X(\bar{u}_i)] - m \left[1 - \sum_{i=1}^n \lambda_i \right] \dots\dots\dots(4.8)$$

Adicionalmente al requerimiento de imparcialidad, la condición de mínima varianza debe satisfacerse. Matemáticamente deben asignarse los pesos, así que

$$V[X^*(\bar{u}_o) - X(\bar{u}_o)] \dots\dots\dots(4.9)$$

Se minimiza.

El resultado de la condición anterior es la ecuación

$$\sum_{j=1}^n \lambda_j C(\bar{u}_i, \bar{u}_j) = C(\bar{u}_i, \bar{u}_o) \text{ para } i = 1, \dots, n \dots\dots\dots(4.10)$$

Donde $C(\bar{u}_i, \bar{u}_j)$ es el valor de la covarianza entre los puntos de localización \bar{u}_i, \bar{u}_j , respectivamente, y $C(\bar{u}_i, \bar{u}_o)$ es la covarianza entre la muestra en la localización \bar{u}_i , y la muestra que no puede obtenerse en la localización \bar{u}_o . Los valores de covarianza se obtienen de acuerdo al modelo espacial. Para detalles de la derivación de ecuaciones, puede consultarse Kelkar, Perez y Chopra⁴.

En forma matricial la ecuación 4.10 puede escribirse como:

$$\begin{bmatrix} C(\bar{u}_1, \bar{u}_1) & \dots & C(\bar{u}_1, \bar{u}_n) \\ \vdots & & \vdots \\ C(\bar{u}_n, \bar{u}_1) & \dots & C(\bar{u}_n, \bar{u}_n) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C(\bar{u}_1, \bar{u}_o) \\ \vdots \\ C(\bar{u}_n, \bar{u}_o) \end{bmatrix} \dots\dots\dots(4.11)$$

O

$$|C|\Lambda = |c|$$

Para calcular el valor de $|\Lambda|$, se puede escribir

$$|\Lambda| = |C|^{-1}|c| \dots\dots\dots(4.12)$$

Donde $|C|^{-1}$ es el inverso de la matriz $|C|$.

El peso asignado a una muestra individual depende de dos factores. El primero es la relación espacial para la localidad donde no es posible obtener muestras. Entre más fuerte sea la relación, mayor será el peso asignado. El segundo es la relación espacial del punto muestreado con otros puntos. Entre más fuerte sea la relación, menor será la información independiente que un punto proporcione. Como resultado, el peso asignado a un punto muestreado será menor.

Una vez que los pesos se han estimado, la ecuación 4.4 se utiliza para determinar el valor de $X^*(\bar{u}_o)$. Adicionalmente a dicha estimación, es necesario determinar la varianza de error de estimación, la cual esta dada por la siguiente expresión

$$\sigma^2_E = C(\bar{u}_o, \bar{u}_o) - \sum_{i=1}^n \lambda_i C(\bar{u}_i, \bar{u}_o) \dots\dots\dots(4.13)$$

Para detalles de la derivación de ecuaciones, puede consultarse Kelkar, Perez y Chopra⁴.

Al examinar la ecuación 4.13, se observa que el máximo valor de la varianza de error de estimación, $C(\bar{u}_o, \bar{u}_o)$, es la varianza misma. En ausencia de información espacial, la incertidumbre respecto a la estimación, debe representarse por la varianza de los datos. Conforme se tiene más información disponible respecto a las relaciones espaciales, la varianza de error de estimación se reduce. Así mismo la reducción tiene vínculo, tanto con el peso asignado a los puntos muestreados, con la covarianza entre los mismos y con la localización donde no es posible obtener muestras. La varianza de error se reduce cuando el valor de $C(\bar{u}_i, \bar{u}_o)$ es grande, y se asigna un valor grande al peso. Como se discutió previamente, un punto muestreado, con una muy fuerte relación en la localidad donde no es posible obtener muestras, esta aislado de otros puntos muestreados y tiene una mucha mayor asignación en cuanto al valor del peso asignado. Esto, en combinación con el alto valor de la covarianza, resulta en una alta reducción de la varianza de error. Por otra parte, si un punto muestreado tiene una fuerte relación espacial respecto a la localización donde no es posible obtener muestras, pero esta rodeado por otras muestras, el peso asignado a dicho punto será menor, por lo que la varianza de error también será menor.

Todas las ecuaciones previas están escritas en términos de la covarianza en lugar del variograma. Aunque se pueden escribir ecuaciones en términos del variograma, es mucho más difícil resolver la matriz de ecuaciones cuando esta es escrita en términos de la covarianza. Por lo tanto, es preferible manejar valores de la covarianza en lugar de emplear valores del variograma.

4.2.2. KRIGING ORDINARIO.

En la técnica para modelar por Kriging simple, se supone que el valor de la media, $m(\bar{u})$, se conoce. Al asumir la condición de estacionariedad de primer orden, $m(\bar{u})$ se reduce a m . Por lo que es necesario conocer el valor de m antes de utilizar la expresión para el Kriging simple. Sin embargo, en la práctica el valor real de la media rara vez se conoce, a menos de que se asuma que el valor de la media muestral sea el mismo, que el valor de la media global. Aun más, como se comentó antes, la media local durante la determinación de la vecindad, puede variar sobre la región de interés. Como resultado, la suposición de estacionariedad de primer orden no puede ser estrictamente válida.

La técnica para modelar por Kriging ordinario puede resolver este problema redefiniendo las ecuaciones empleadas para realizar la estimación.

Considerando la ecuación 4.4 utilizada para el Kriging simple

$$X^*(\bar{u}_o) = \lambda_o + \sum_{i=1}^n \lambda_i X(\bar{u}_i) \dots\dots\dots(4.4)$$

La condición de imparcialidad requiere que

$$E[X^*(\bar{u}_o) - X(\bar{u}_o)] = 0 \dots\dots\dots(4.5)$$

Si se supone $E[X^*(\bar{u}_o)] = E[X(\bar{u}_o)] = m(\bar{u}_o)$, donde $m(\bar{u}_o)$ representa la media dentro de la localización (\bar{u}_o) , se puede escribir lo siguiente

$$\lambda_o = m(\bar{u}_o) \left[1 - \sum_{i=1}^n \lambda_i \right] \dots\dots\dots(4.14)$$

Si se fuerza a que $\lambda_o = 0$, y se asume que

$$\sum_{i=1}^n \lambda_i = 1 \dots\dots\dots(4.15)$$

La estimación entonces puede escribirse como

$$X^*(\bar{u}_o) = \sum_{i=1}^n \lambda_i X(\bar{u}_i) \dots\dots\dots(4.16)$$

Con la limitación definida en la ecuación 4.15. La necesidad del requerimiento de un valor de la media se elimina forzando a que el valor de λ_o sea cero. Aun más, ya no es necesario satisfacer estrictamente la condición de estacionariedad de primer orden.

Adicionalmente a la condición de imparcialidad, se requiere satisfacer la condición de varianza mínima. La minimización de la varianza al contemplar la limitación de la ecuación 4.15, arroja la siguiente expresión

$$\sum_{j=1}^n \lambda_j C(\bar{u}_i, \bar{u}_j) + \mu = C(\bar{u}_i, \bar{u}_o) \text{ para } i = 1, \dots, n \dots\dots\dots(4.17)$$

Donde μ es denominado el multiplicador de Lagrange, y C representa la covarianza. Para detalles de la derivación de ecuaciones, puede consultarse Kelkar, Perez y Chopra⁴.

En forma matricial la ecuación 4.17 puede escribirse como

$$\begin{bmatrix} C(\bar{u}_1, \bar{u}_1) & \dots & C(\bar{u}_1, \bar{u}_n) & 1 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ C(\bar{u}_n, \bar{u}_1) & \dots & C(\bar{u}_n, \bar{u}_n) & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \cdot \\ \cdot \\ \cdot \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} C(\bar{u}_1, \bar{u}_o) \\ \cdot \\ \cdot \\ \cdot \\ C(\bar{u}_n, \bar{u}_o) \\ 1 \end{bmatrix} \dots\dots\dots(4.18)$$

Para resolver la matriz respecto λ_i , se puede escribir

$$|\Lambda| = |C|^{-1} |c| \dots\dots\dots(4.19)$$

Una vez que λ_i se calcula, el estimador $X^*(\bar{u}_o)$, se obtiene con la ecuación 4.18, y la varianza de error se determina con

$$\sigma^2_E = C(\bar{u}_o, \bar{u}_o) - \sum_{i=1}^n \lambda_i C(\bar{u}_i, \bar{u}_o) - \mu \dots\dots\dots(4.20)$$

4.2.3. COKRIGING.

El Cokriging es una técnica de modelado utilizada para estimar el valor de una variable en localizaciones donde no es posible realizar muestreo al emplear muestras que se relacionan y se encuentran rodeando a la variable en cuestión. El objetivo al aplicar la técnica de Cokriging consiste en mejorar la estimación y reducir la incertidumbre en dicha estimación con ayuda de información espacial disponible de otras variables. Existe una suposición implícita en el proceso. La variable de interés y las otras variables se relacionan mutuamente en cuanto a espacio.

La estimación de datos de permeabilidad al utilizar datos de porosidad, y la estimación de datos de porosidad al emplear datos sísmicos, son dos ejemplos claros y comúnmente aplicados, que se realizan durante la descripción del yacimiento en donde el Cokriging se utiliza para mejorar las estimaciones.

Normalmente solo se obtienen muestras con núcleos en unos cuantos pozos, y en casi todos los pozos se realiza toma de registros geofísicos. Si se establece una relación espacial entre la permeabilidad y los datos de porosidad, la permeabilidad estimada, en la localización donde no es posible obtener muestreo, se puede mejorar mediante los datos de porosidad que se encuentran alrededor de dichas localidades no muestreables.

Antes de discutir las ecuaciones necesarias para realizar un modelado empleando la técnica de Cokriging, es necesario comprender sus limitaciones:

1. La técnica de Cokriging presume una relación mutua de linealidad entre las variables. Si la relación es lineal, la covarianza solo cuantifica la relación espacial. Si la relación no es lineal, la covarianza no puede proporcionar información significativa. La relación entre la variable principal y las variables de soporte o covariables, debe de ser lineal antes

de comenzar con el proceso de estimación. Por ejemplo, al utilizar el logaritmo de la permeabilidad con los datos de porosidad se asume normalmente una relación lineal.

2. La relación entre la variable principal y la (s) covariable (s) debe ser “fuerte”. La fortaleza depende del coeficiente de correlación entre las variables, y el número de puntos muestreados que se utilizan en la estimación de dicho coeficiente. Entre mayor sea el coeficiente de correlación, y mayor sea el número de puntos muestreados mejor será la relación. También es necesario conocer si la relación se basa en factores físicos y existe evidencia histórica cuando la covariable se utiliza para estimar la variable principal. Por ejemplo puede que exista una fuerte evidencia empírica entre el logaritmo de la permeabilidad y la porosidad, sin embargo no existe una evidencia física entre dicha relación. Antes de decidir utilizar una covariable para mejorar la estimación de la variable principal, es necesario asegurar que tan favorables son los factores antes mencionados.
3. La aplicación de los procedimientos del Cokriging, requieren de un esfuerzo substancial en el modelado espacial. Entre más covariables se empleen, mayor será el esfuerzo empeñado en el modelo.
4. Resolver el sistema de ecuaciones que se involucra en la técnica para modelar por Cokriging requiere un esfuerzo computacional adicional mucho mayor comparado con la técnica para modelar por Kriging. Conforme el número de variables se incrementa, la matriz que resuelve el sistema de ecuaciones también se incrementa.

Con base en lo ya descrito, la técnica de modelado por Cokriging puede aplicarse si es que existe una buena relación entre las variables de interés y las covariables.

Para el desarrollo de ecuaciones apropiadas, es necesario primero determinar si se seguirá una aproximación por Kriging simple o por Kriging ordinario. Las ecuaciones finales difieren insignificadamente dependiendo el modelado por Kriging base, que se escoge.

El modelado por Kriging ordinario se practica con más frecuencia, en comparación con el modelado por Kriging simple debido a las consideraciones discutidas en los subcapitulos anteriores. Para simplificar algunas ecuaciones, se tomará solo una covariable para la estimación del valor de la variable principal en la localidad donde no es posible obtener muestras. Algunos autores proporcionan expresiones más generalizadas para múltiples covariables^{5, 6}.

Para estimar un valor $x^*(\bar{u}_o)$, en una localidad no muestreable, \bar{u}_o , dentro de una vecindad ya determinada, se tienen n muestras de la variable principal y m muestras de la covariable Y . La ecuación para obtener la estimación se puede escribir como

$$X^*(\bar{u}_o) = \sum_{i=1}^n \lambda_{X_i}^* X(\bar{u}_{X_i}) + \sum_{k=1}^m \lambda_{Y_k}^* Y(\bar{u}_{Y_k}) \dots\dots\dots(4.21)$$

Donde λ_{X_i} es el peso asignado a la muestra, $X(\bar{u}_{X_i})$, localizada en (\bar{u}_{X_i}) , y λ_{Y_k} es el peso asignado a la muestra $Y(\bar{u}_{Y_k})$, en la localización (\bar{u}_{Y_k}) . Es posible observar la variable principal y la covariable en la misma localidad. Sin embargo, para mantener la generalidad, se asumen diferentes notaciones para indicar las localidades de las dos variables.

Al aplicar la condición de imparcialidad

$$E[X^*(\bar{u}_o) - X(\bar{u}_o)] = 0 \dots\dots\dots(4.5)$$

Sustituyendo la ecuación 4.21 en la ecuación 4.5, se observa que

$$m_X \sum_{i=1}^n \lambda_{X_i} + m_Y \sum_{k=1}^m \lambda_{Y_k} - m_X = 0 \dots\dots\dots(4.22)$$

Donde m_x y m_y son los valores esperados de las variables X e Y , respectivamente.

Para satisfacer la ecuación 4.22, se puede escribir

$$\sum_{i=1}^n \lambda_{x_i} = 1 \text{ y } \sum_{k=1}^n \lambda_{y_k} = 0 \dots\dots\dots(4.23)$$

La ecuación 4.23 asegura que la condición de imparcialidad se satisface. Pueden utilizarse otros métodos alternativos para satisfacer dicha condición. Para detalles de la derivación de ecuaciones, puede consultarse Kelkar, Perez y Chopra⁴.

Adicionalmente, debe satisfacerse la condición de minimización en la varianza. Matemáticamente se tiene

$$V \left[X(\bar{u}_o) - \sum_{i=1}^n \lambda_{x_i}^* X(\bar{u}_{x_i}) - \sum_{k=1}^n \lambda_{y_k}^* Y(\bar{u}_{y_k}) \right] \dots\dots\dots(4.24)$$

Al minimizar la ecuación 4.24, con las dos restricciones definidas en la ecuación 4.23 se tiene que

$$\sum_{j=1}^n \lambda_{x_j} C_X(\bar{u}_{x_i}, \bar{u}_{x_j}) + \sum_{k=1}^m \lambda_{y_k} C_C(\bar{u}_{x_i}, \bar{u}_{y_k}) + \mu_X = C_X(\bar{u}_o, \bar{u}_{x_i}) \text{ para } i = 1, \dots, n \dots\dots(4.25)$$

y

$$\sum_{i=1}^n \lambda_{x_i} C_C(\bar{u}_{x_i}, \bar{u}_{y_k}) + \sum_{l=1}^m \lambda_{y_l} C_Y(\bar{u}_{y_k}, \bar{u}_{y_l}) + \mu_Y = C_C(\bar{u}_o, \bar{u}_{y_k}) \text{ para } k = 1, \dots, m \dots\dots(4.26)$$

En las ecuaciones 4.25 y 4.26, C_X y C_Y son la covarianzas de las variables X e Y , respectivamente, y C_C es la covarianza cruzada entre las dos variables, μ_X y μ_Y son los multiplicadores de Lagrange respectivamente.

Las ecuaciones anteriores pueden representarse en forma matricial

$$\begin{bmatrix}
 C_X(\bar{u}_{X_1}, \bar{u}_{X_1}) & \dots & C_X(\bar{u}_{X_1}, \bar{u}_{X_n}) & C_X(\bar{u}_{X_1}, \bar{u}_{Y_1}) & \dots & C_X(\bar{u}_{X_1}, \bar{u}_{Y_m}) & 1 & 0 \\
 \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\
 \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\
 \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\
 C_X(\bar{u}_{X_n}, \bar{u}_{X_n}) & \dots & C_X(\bar{u}_{X_n}, \bar{u}_{X_n}) & C_X(\bar{u}_{X_n}, \bar{u}_{Y_n}) & \dots & C_X(\bar{u}_{X_n}, \bar{u}_{Y_m}) & 1 & 0 \\
 C_C(\bar{u}_{X_1}, \bar{u}_{Y_1}) & \dots & C_C(\bar{u}_{X_n}, \bar{u}_{Y_1}) & C_C(\bar{u}_{Y_1}, \bar{u}_{Y_1}) & \dots & C_C(\bar{u}_{Y_1}, \bar{u}_{Y_m}) & 0 & 1 \\
 \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\
 \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\
 \cdot & \dots & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\
 C_C(\bar{u}_{X_1}, \bar{u}_{Y_m}) & \dots & C_C(\bar{u}_{X_n}, \bar{u}_{Y_m}) & C_C(\bar{u}_{Y_m}, \bar{u}_{Y_1}) & \dots & C_C(\bar{u}_{Y_m}, \bar{u}_{Y_m}) & 0 & 1 \\
 1 & \dots & 1 & 0 & \dots & 0 & 0 & 0 \\
 0 & \dots & 0 & 1 & \dots & 1 & 0 & 0
 \end{bmatrix}
 \begin{bmatrix}
 \lambda_{X_1} \\
 \cdot \\
 \cdot \\
 \cdot \\
 \lambda_{X_n} \\
 \lambda_{Y_1} \\
 \cdot \\
 \cdot \\
 \cdot \\
 \lambda_{Y_m} \\
 \mu_X \\
 \mu_Y
 \end{bmatrix}
 =
 \begin{bmatrix}
 C_X(\bar{u}_o, \bar{u}_{X_1}) \\
 \cdot \\
 \cdot \\
 \cdot \\
 C_X(\bar{u}_o, \bar{u}_{X_n}) \\
 C_C(\bar{u}_o, \bar{u}_{Y_1}) \\
 \cdot \\
 \cdot \\
 \cdot \\
 C_C(\bar{u}_o, \bar{u}_{Y_m}) \\
 1 \\
 0
 \end{bmatrix}$$

Del lado izquierdo, se tiene una matriz de tamaño $(n + m + 2)$ que debe resolverse para obtener los valores λ_{X_i} y λ_{Y_k} . En comparación con el Kriging ordinario, el tamaño de la matriz es significativamente mayor.

También puede obtenerse la expresión para la varianza de error de estimación, la cual puede escribirse como

$$\sigma^2_E = C(\bar{u}_o, \bar{u}_o) - \sum_{i=1}^n \lambda_{X_i} C_X(\bar{u}_o, \bar{u}_{X_i}) - \sum_{k=1}^m \lambda_{Y_k} C_C(\bar{u}_o, \bar{u}_{Y_k}) - \mu_X \dots \dots \dots (4.27)$$

Como en el caso del Kriging ordinario, la varianza de error es una indicación de la configuración relativa alrededor de los datos no muestreados.

4.3. MODELADO POR SIMULACIÓN GAUSSIANA SECUENCIAL.

En un modelado por simulación, se emplean técnicas cuyos procedimientos permiten, como su nombre lo indica, simular en lugar de estimar, varios atributos en localidades donde no es posible obtener muestras. La meta principal que se persigue, es la de simular la realidad en lugar de obtener una “radiografía” del yacimiento.

Existe una gran cantidad de metodologías reportadas en la literatura, para realizar un modelado por simulación, como lo son la simulación secuencial por indicadores^{7, 8}, y la simulación Gaussiana secuencial^{9, 10, 11}. Ambos métodos se utilizan ampliamente para la simulación de variables discretas y continuas. Para desarrollar un modelado por simulación Gaussiana secuencial, se requieren cinco pasos. Estos cinco pasos involucran la transformación de los datos originales en un nuevo dominio, el modelado de variogramas en el dominio transformado, seleccionar un patrón de manera aleatoria que permita encontrar secuencialmente valores en las localidades donde no es posible obtener muestras con base en técnicas empleadas en el modelado por Kriging, y finalmente retransformar los valores al dominio original.

4.3.1 SIMULACIÓN DE UNA VARIABLE SIMPLE.

La simulación secuencial Gaussiana de una variable simple se realiza como se enuncia en los siguientes tópicos.

Transformación de datos dentro de un nuevo dominio.

El propósito principal de transformar los datos dentro de un nuevo dominio, es el de cuantificar apropiadamente la incertidumbre en las localidades donde no es posible obtener muestras. Esto se logra al realizar una transformación Gaussiana (normal), lo cual es posible tanto para variables continuas como para variables discretas.

En lo que refiere a variables continuas, todos los valores de las muestras se ordenan de forma ascendente, construyendo una función de distribución acumulativa, la cual debe variar entre cero y uno. Una vez que se construye la función de distribución acumulativa para la muestra, la transformación dentro de un dominio Gaussiano es directa, y se ilustra en la figura 4.4.

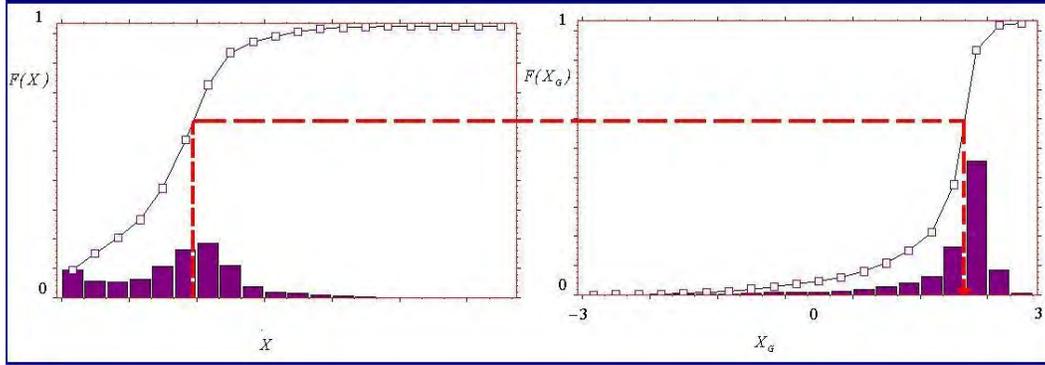


Figura 4.4. Transformación Gaussiana de una variable continua.

Para una variable discreta, la transformación dentro de un espacio Gaussiano requiere una transformación de la función densidad de probabilidad, dentro de una función de distribución acumulativa. Una vez que la función de distribución acumulativa se construye, la transformación dentro de un espacio Gaussiano es similar al procedimiento para una variable continua. Esquemáticamente, este procedimiento se muestra en la figura 4.5.

Para explicar la figura 4.5, se asume que la variable discreta esta representada por A, B, C y D (facies geológicas por ejemplo). Cada categoría tiene una función densidad (masa) de probabilidad, $p(K)$. Cuando se determina el orden de las categorías -arbitrariamente- se calcula la función de distribución acumulativa, $F(K)$. En el ejemplo el orden es A, B, C y D .

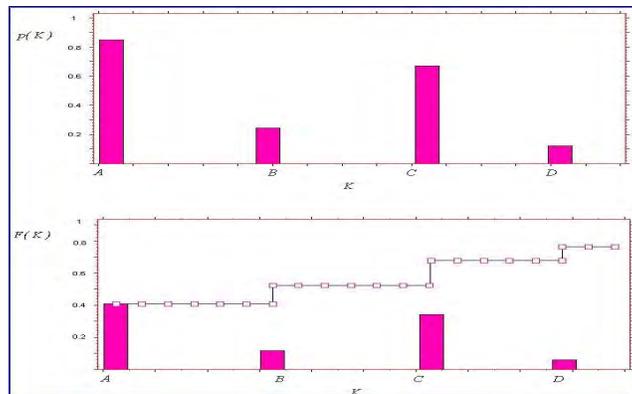


Figura 4.5. Transformación Gaussiana de una variable discreta.

En la transformación Gaussiana no existen pérdidas de información, cuando se transforman los datos discretos o continuos.

Modelado espacial en el dominio transformado.

Una vez que se transforman los datos dentro de un nuevo dominio, el segundo paso involucra el modelado de relaciones espaciales en dicho dominio. Este paso es similar a cualquier análisis convencional de variogramas, en el que el variograma en el dominio transformado, se estima y modela mediante una apropiada relación.

Para variables discretas, el número de valores límites es igual al número de categorías. Para variables continuas, el número podría variar dependiendo del número de puntos muestreados, además de otro tipo de razones físicas externas, que requieran de una revisión particular. Cuando se realiza la transformación Gaussiana solamente debe modelarse un variograma.

Selección de un patrón aleatorio.

Cuando se ha estimado y modelado el variograma, el tercer paso en una simulación Gaussiana secuencial consiste en seleccionar un patrón el cual permita ubicar aquella localidad donde no es posible obtener muestras. Normalmente con la ayuda de un número generado aleatorio, se crea una secuencia de números aleatorios que corresponden al número total de celdas de malla, y con base en el orden, se selecciona un patrón con el que se ubican todas las localidades donde no es posible obtener muestras. El motivo para seleccionar un patrón aleatorio, cuando se ubican las localizaciones donde no pueden obtenerse muestras, en lugar de ubicar las mismas en orden secuencial, se aclara en el siguiente paso.

Estimación en la localización no muestreable.

Cuando se selecciona aleatoriamente una localidad donde no es posible obtener muestras, y se ubica, se define la vecindad determinada, como en el modelado por Kriging, recolectando los puntos muestreados de dicha vecindad.

Una distinción importante entre las técnicas de Kriging y simulación Gaussiana secuencial, es la selección de muestras dentro de la vecindad determinada. Adicionalmente a lo anterior, en la simulación Gaussiana secuencial se seleccionan valores simulados previamente. La figura 4.6, ilustra la diferencia descrita. En el modelado por Kriging solo se seleccionan los puntos muestreados originales dentro de la vecindad determinada, en contraste, en la simulación Gaussiana secuencial se seleccionan las muestras originales además de valores previamente simulados dentro de la vecindad determinada. Dicha diferencia entre ambas técnicas produce un par de consecuencias. Primero, al seleccionar los valores simulados previamente dentro de la vecindad determinada, la relación espacial entre los valores simulados es explícitamente tomada en cuenta durante la simulación Gaussiana secuencial, lo cual no es posible en las técnicas para modelar por Kriging. Segundo la selección de valores previamente simulados ocasiona que la estimación en las localidades donde no es posible obtener muestras sea dependiente del orden en que se ubica a dichas muestras.

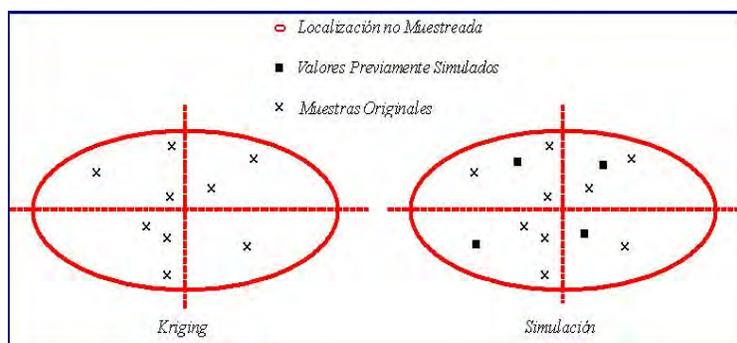


Figura 4.6. Selección de muestras.

Manteniendo el balance entre las muestras originales y los valores previamente simulados dentro de una vecindad determinada, se asegura que no se exagera en la estimación debido a los valores simulados previamente conforme avanza el proceso de simulación. Una vez que se seleccionan las muestras dentro de la vecindad determinada, la estimación en las localidades donde no es posible obtener muestras, requiere de las técnicas para modelar por Kriging.

Cuando se ha realizado la transformación Gaussiana, solo se requiere una estimación en las localidades donde no es posible obtener muestras. Utilizando las formulas para el Kriging ordinario se tiene lo siguiente

$$X_G^*(\bar{u}_o) = \sum_{i=1}^n \lambda_i X_G(\bar{u}_i) \dots\dots\dots(4.28)$$

Para una variable continua, y

$$K_G^*(\bar{u}_o) = \sum_{i=1}^n \lambda_i K_G(\bar{u}_i) \dots\dots\dots(4.29)$$

Para una variable discreta.

Con el variograma modelado y las muestras seleccionadas (originales además de los valores previamente simulados), se estima el valor de la variable transformada en el espacio Gaussiano así como la varianza de error con ecuaciones empleadas para modelar por Kriging.

Retransformación de datos.

Ya que se ubicaron todas las localidades donde no es posible obtener muestras, y se estiman los valores e incertidumbres locales en el dominio transformado, el siguiente paso consiste en simular uno de los valores

estimados así como su incertidumbre local, para retransformar posteriormente dicho valor.

El proceso de muestreo de valores para variables continuas y discretas es el mismo. Como se muestra en la figura 4.7. Mediante el número generador aleatorio, se toma un valor de la distribución Gaussiana. En la figura 4.7, con base en el número aleatorio se toma la realización $x_G(\bar{u}_o)$. Se generan distintas realizaciones de manera similar. El valor de cada realización, con un 99% de probabilidad, caerá entre $X_G^*(\bar{u}_o) \pm 3\sigma_E(\bar{u}_o)$.

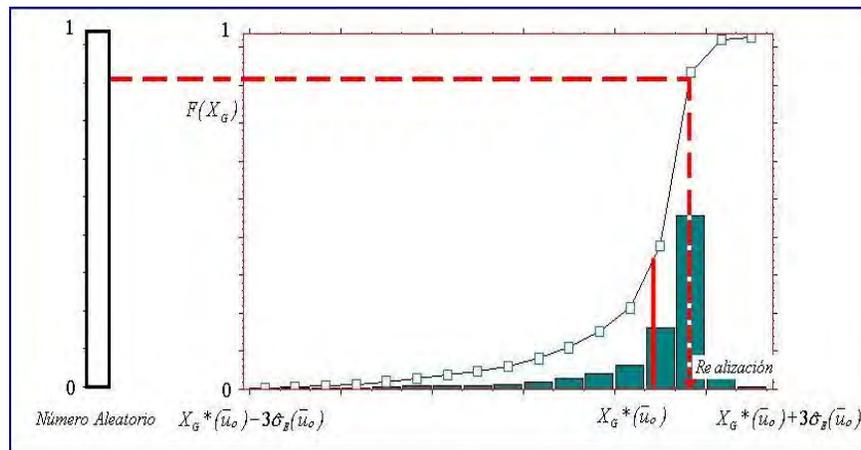


Figura 4.7. Selección de una muestra en la transformación Gaussiana.

Ya que se muestrea el valor $x_G(\bar{u}_o)$, este puede retransformarse a su dominio original. Para una variable continua, como se ilustra en la figura 4.8, el proceso es discreto, es decir, es lo inverso a lo que se expresa en la figura 4.4. Con la distribución original de la muestra, se grafica un valor en el espacio Gaussiano, y para el mismo valor de la función de distribución acumulativa, se obtiene un valor retransformado. Lo que indica una realización en la localización (\bar{u}_o) .

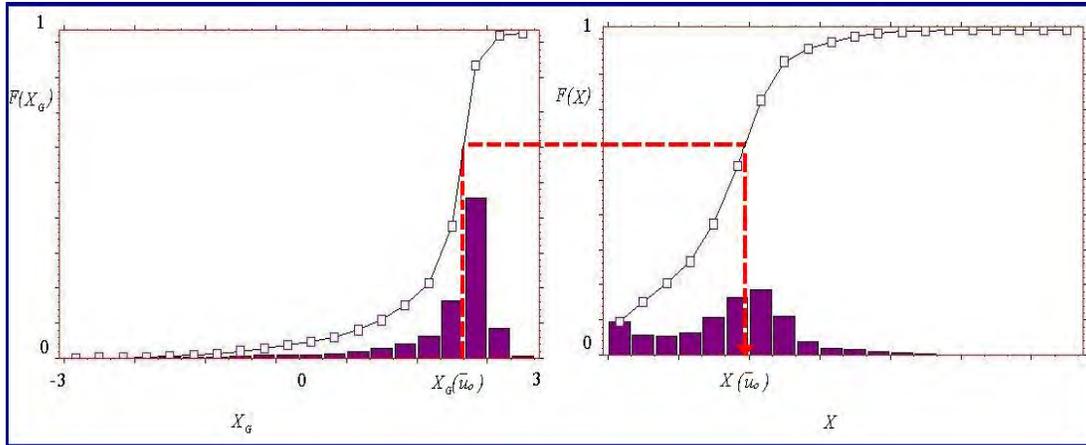


Figura 4.8. Retransformación de un espacio Gaussiano para una variable continua.

Para una variable discreta, la transformación puede hacerse de dos maneras. Una simple aproximación utiliza la distribución original (ver figura 4.5) así como el proceso inverso para obtener el valor para una categoría particular. Esto se muestra en la figura 4.9, en este caso, la categoría *A* se asigna a la localidad donde no es posible obtener muestras, de acuerdo al valor en el dominio transformado.

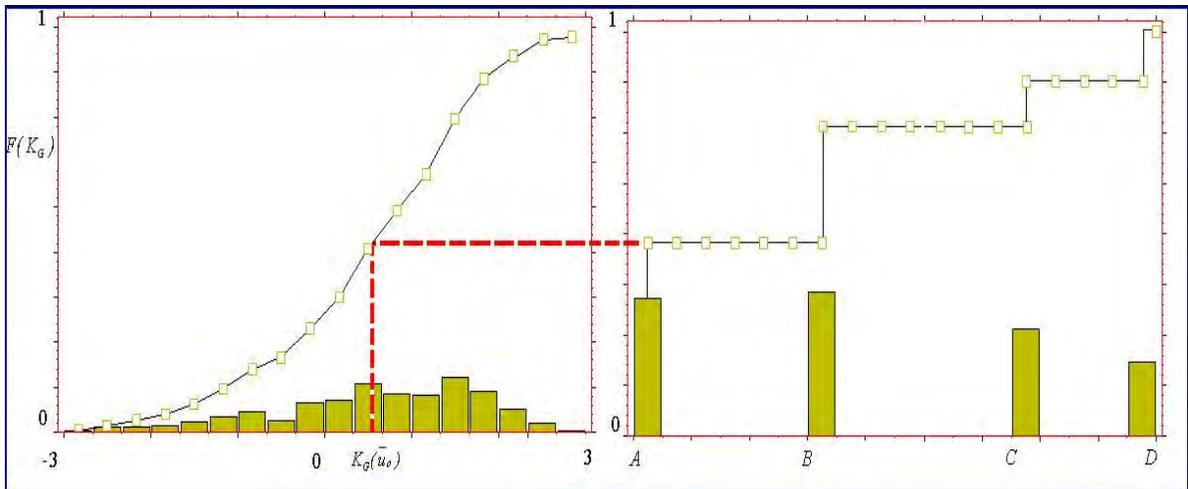


Figura 4.9. Retransformación de un espacio Gaussiano para una variable discreta.

La otra aproximación controla la retransformación, dependiendo en la localización vertical. Esto es importante desde una perspectiva geológica.

Al considerar la figura 4.10, en la que se muestra una sección transversal entre dos pozos. El geólogo identifica las capas marcadoras de tiempo. En un ejemplo simplificado, la cima contiene a las facies *A* y *B*, y la base contiene a las facies *C* y *D*.

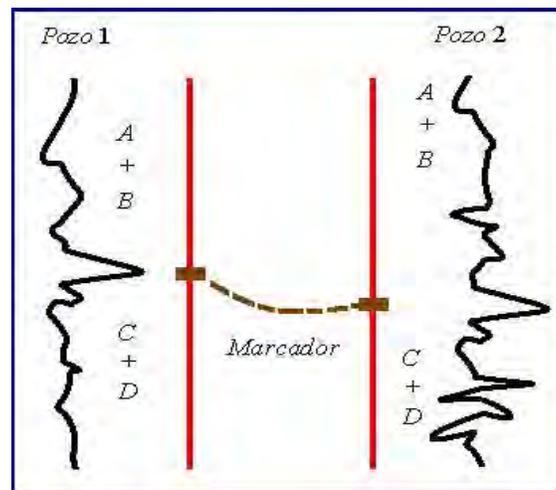


Figura 4.10. Intervalos geológicos isocronales.

4.3.2 COSIMULACIÓN GAUSSIANA SECUENCIAL.

La cosimulación Gaussiana secuencial es una extensión lógica de la simulación Gaussiana secuencial. El proceso de cosimulación consiste en la simulación de múltiples atributos. La ventaja principal de la cosimulación Gaussiana secuencial es la habilidad de tomar en cuenta las relaciones locales (de manera lineal) que existen entre varios atributos o variables. Una vez que se determina el orden de los atributos a simular, dicho orden se traslada a cada localidad donde no es posible obtener muestras para asegurar que las interrelaciones entre los distintos atributos, se toman en cuenta. Un ejemplo del orden que se menciona se ilustra en la figura 4.11.

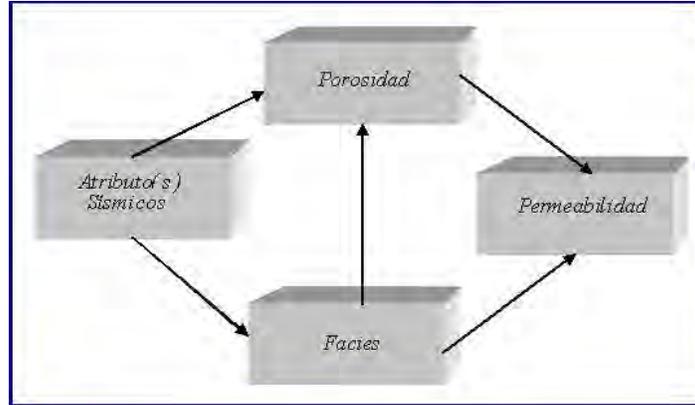


Figura 4.11. Atributos interrelacionados en celdas de malla individuales.

Para estimar un valor, se utiliza el procedimiento comentado para modelar por Cokriging. En términos generales, la estimación del atributo $X_k(\bar{u}_o)$, se consigue mediante la ecuación

$$X_k^*(\bar{u}_o) = \sum_{i=1}^{n_k} \lambda_{i_k} X_k(\bar{u}_i) + \sum_{l=1}^L \sum_{j=1}^{n_l} \lambda_{j_l} X_l(\bar{u}_j) \dots\dots\dots(4.30)$$

Donde λ_{i_k} es el peso asignado a la muestra $X_k(\bar{u}_i)$ (valor original o previamente estimado) dentro de la vecindad determinada. Estas son las muestras de la misma variable. λ_{j_l} es el peso asignado a las muestras covariables $X_l(\bar{u}_j)$ en la vecindad determinada y L cambia con la covariable simulada en la localidad donde no es posible obtener muestras. Con la figura 4.11, como ejemplo, para la estimación de facies, puede tenerse $L=1$, lo que representa la medición de un atributo sísmico. Para la estimación de la porosidad, $L=2$, lo cual incluye la medición de atributos sísmicos y valores para facies previamente simulados. Con la misma lógica, para la estimación de la permeabilidad, se tiene $L=3$.

Este procedimiento es muy eficiente desde el punto de vista computacional. La desventaja de dicho proceso es la limitación en cuanto a la relación lineal entre los distintos atributos¹².

4.4. APLICACIÓN PRÁCTICA.

El siguiente ejercicio tiene como objetivo mostrar algunas diferencias que se encontraron al realizar visualizaciones 2D de impedancia acústica al utilizar las técnicas para modelar por Kriging simple y simulación Gaussiana secuencial, mediante la utilización de PETREL 2004^{MR}. Este tipo de visualizaciones permite conocer como se comporta la distribución de una propiedad a lo largo de dos dimensiones, en este caso, la impedancia acústica. Por último se ejemplifican seis visualizaciones equiprobables obtenidas con una simulación Gaussiana secuencial, de acuerdo al manejo de números aleatorios.

Para la construcción de estas visualizaciones, se utilizan como parámetros de entrada los datos del variograma ajustado obtenidos en el ejercicio del capítulo tres, empleando solo la orientación de mayor anisotropía, debido a que es la de interés

$$Meseta = 0.54$$

$$Nugget = 0.14$$

$$Alcance \text{ Mximo (Mayor anisotropa)} = 12,965 \text{ m}$$

$$Alcance \text{ Mnimo (Menor anisotropa)} = 7,820 \text{ m}$$

$$Orientacin \text{ de la zona de mayor anisotropa} = -51^\circ$$

Los resultados obtenidos se muestran en las siguientes figuras

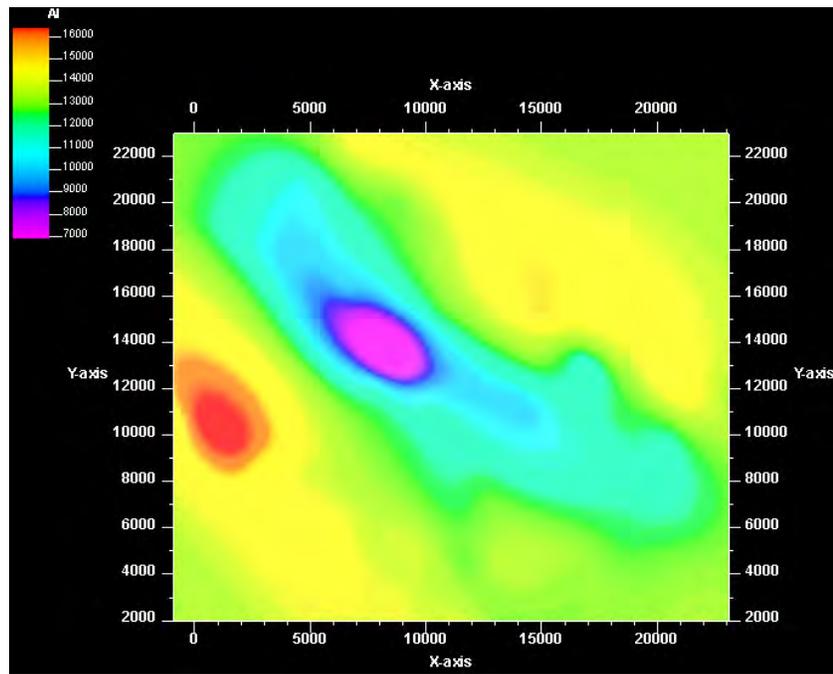


Figura 4.12. Visualización en 2D de impedancia acústica: estimación por Kriging simple.

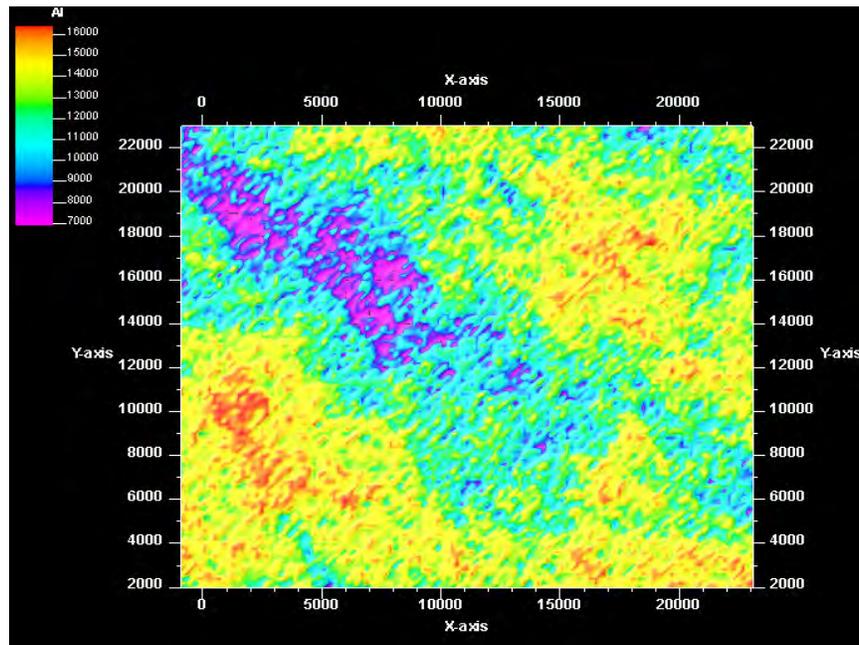


Figura 4.13. Visualización en 2D de impedancia acústica: simulación Gaussiana secuencial.

Al observar ambas visualizaciones, en la correspondiente, al Kriging simple se denota la “suavidad” de la imagen. Otra forma para diferenciar los modelos ya descritos, es el realizar una comparación entre los histogramas obtenidos por Kriging simple y por simulación Gaussiana secuencial, como se muestra en las figuras 4.14 y 4.15.

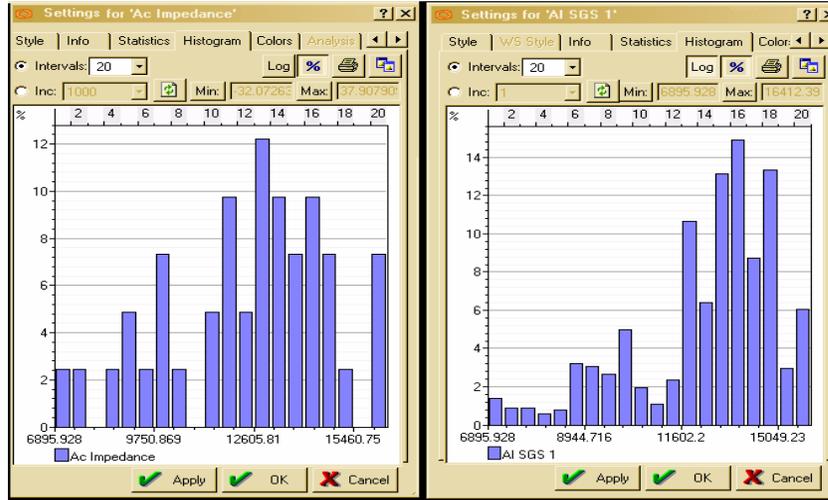


Figura 4.14. Comparación de histogramas: datos originales de impedancia acústica vs datos obtenidos mediante Kriging simple.

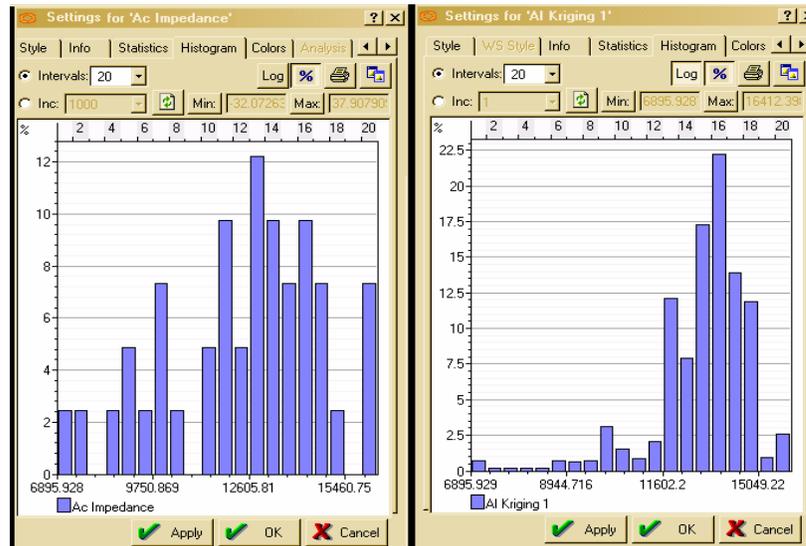


Figura 4.15. Comparación de histogramas: datos originales de impedancia acústica vs datos obtenidos mediante simulación Gaussiana secuencial.

Al comparar ambos métodos, respecto a los datos de entrada originales, se observa que en general se preserva la distribución de manera uniforme en el modelado por Kriging, pero los valores máximos y mínimos de las muestras de entrada no se capturan adecuadamente, lo cual puede ser en ocasiones poco apropiado si se intenta preservar información clave del yacimiento. En lo que respecta con el modelado por simulación Gaussiana secuencial se respeta a distribución inicial de los datos de impedancia acústica, lo cual es una ventaja.

A continuación se presentan seis versiones equiprobables para la impedancia acústica obtenidas con la aplicación de una simulación Gaussiana secuencial. Para realizar dichas versiones se configura una malla con las siguientes dimensiones:

Malla de tipo rectangular

Número de bloques en la dirección - i- = 121

Número de bloques en la dirección - y- = 106

Número de bloques en la dirección - z- = 1

Total número de bloques = 12826

PETREL 2004^{MR} toma un patrón de manera aleatoria de acuerdo al número de celdas, es decir, que dicho número marcará la celda de inicio para realizar la simulación, obteniendo así, en cada proceso, una visualización distinta.

Las siguientes figuras ilustran con mayor detalle la descripción anterior.

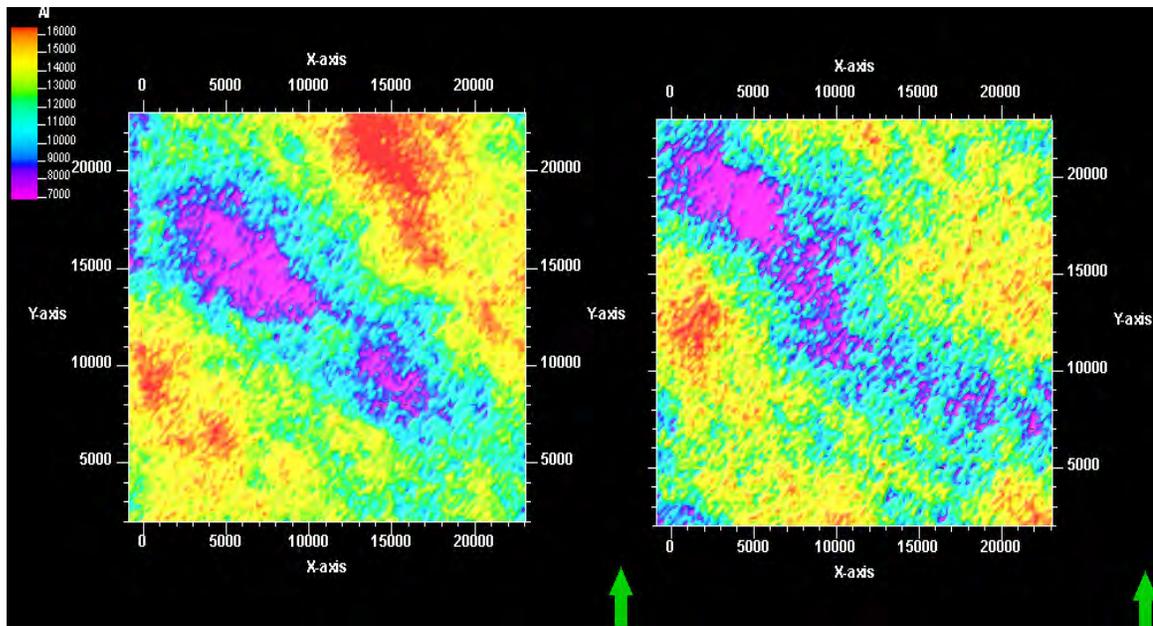


Figura 4.16a. Visualizaciones equiprobables de simulación Gaussiana. Izquierda-número aleatorio= 567, derecha-número aleatorio= 1009.

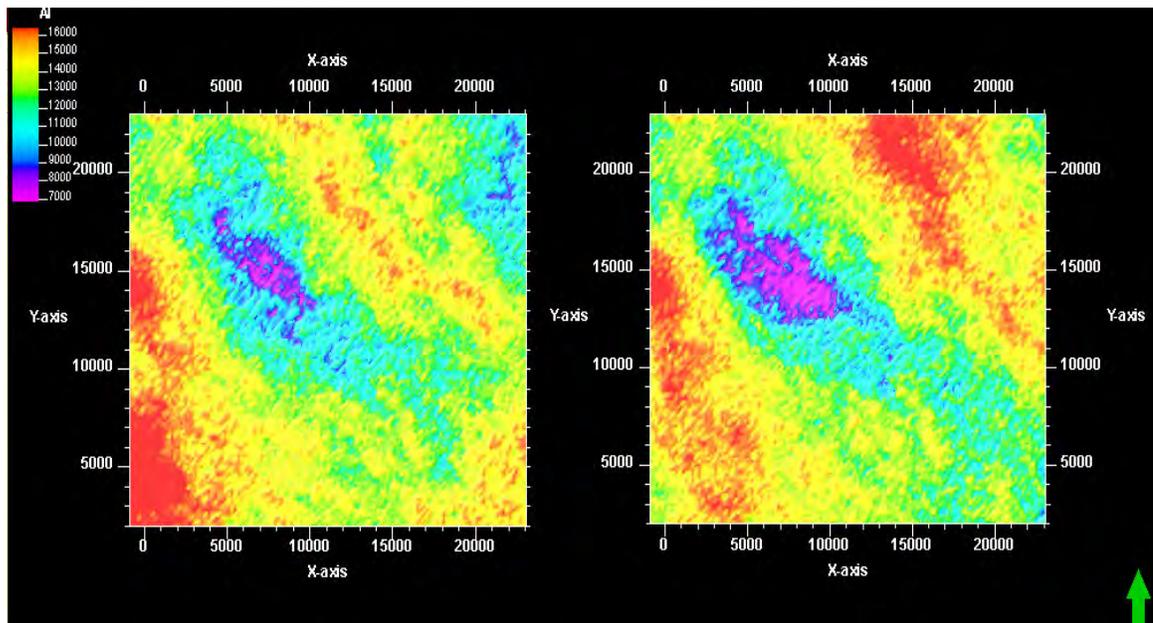


Figura 4.16b. Visualizaciones equiprobables de simulación Gaussiana. Izquierda-número aleatorio= 12337, derecha-número aleatorio= 1789.

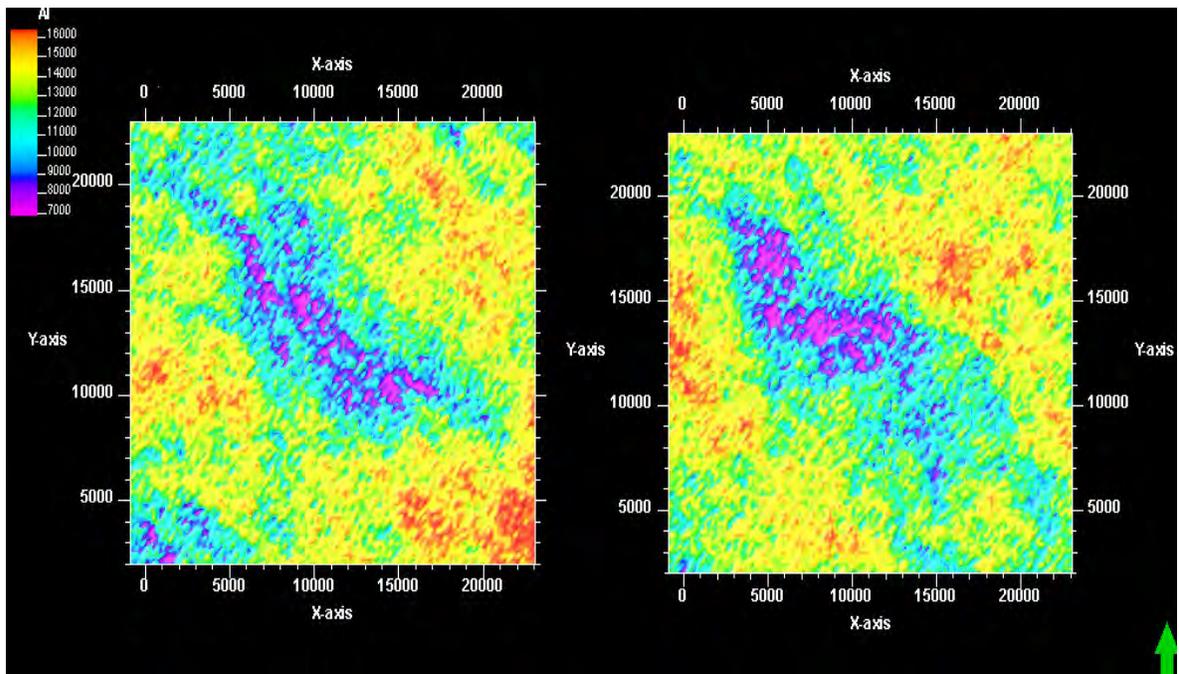


Figura 4.16c. Visualizaciones equiprobables de simulación Gaussiana. Izquierda-número aleatorio= 478, derecha-número aleatorio= 9520.

CAPITULO 5

ESCALAMIENTO DE PROPIEDADES

5.1. INTRODUCCIÓN.

Este capítulo presenta algunas técnicas de escalamiento utilizadas en el cálculo de propiedades del yacimiento para la construcción de mallas a escala gruesa que usualmente poseen celdas más grandes, en comparación con las mallas a escala fina, en las cuales son generadas las descripciones de propiedades geológicas y petrofísicas.

Asimismo se realiza la descripción de técnicas de escalamiento para propiedades estáticas del yacimiento incluyendo espesor neto, porosidad y saturación. También se describen algunas técnicas de escalamiento para propiedades dinámicas como permeabilidad y permeabilidad relativa. Las técnicas de escalamiento descritas se basan en formulaciones matemáticas, por lo que los resultados obtenidos dependerán en gran magnitud de las condiciones bajo las que sean aplicadas las técnicas de escalamiento.

5.2. DEL MODELO ESTÁTICO AL MODELO DINÁMICO.

En ocasiones, las propiedades del yacimiento se describen en escalas pequeñas con la intención de capturar las variaciones observadas en registros, núcleos, modelos geológicos y datos sísmicos. Sin embargo la resolución de los modelos necesarios para otras aplicaciones, como la simulación de flujo, debe ser más gruesa debido a que el máximo número de celdas de malla utilizable, se limita a los recursos computacionales. La figura 5.1 muestra el ejemplo de una malla a escalas fina y gruesa. La figura 5.2 muestra el ejemplo del refinamiento de una celda alrededor de un pozo. Las técnicas de escalamiento de propiedades permiten estimar algunas propiedades petrofísicas para celdas gruesas, como los que se muestran en las figuras 5.1 y 5.2, con base en una malla a escala fina.

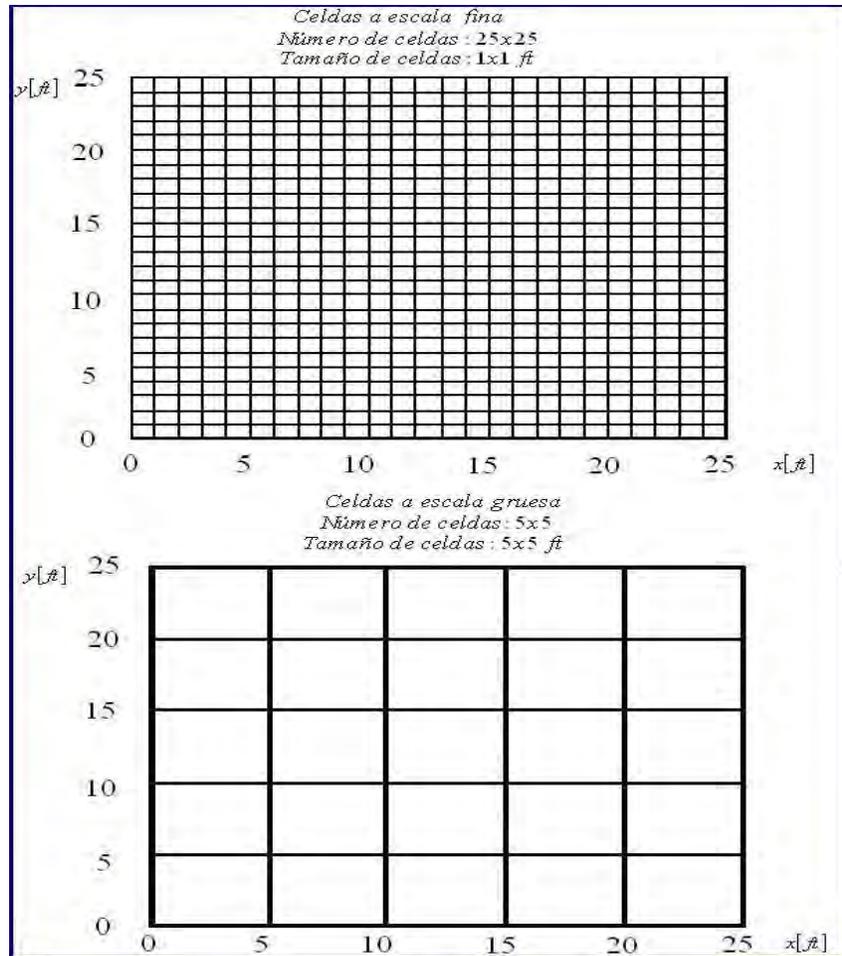


Figura 5.1. Celdas de malla a escalas gruesa y fina.

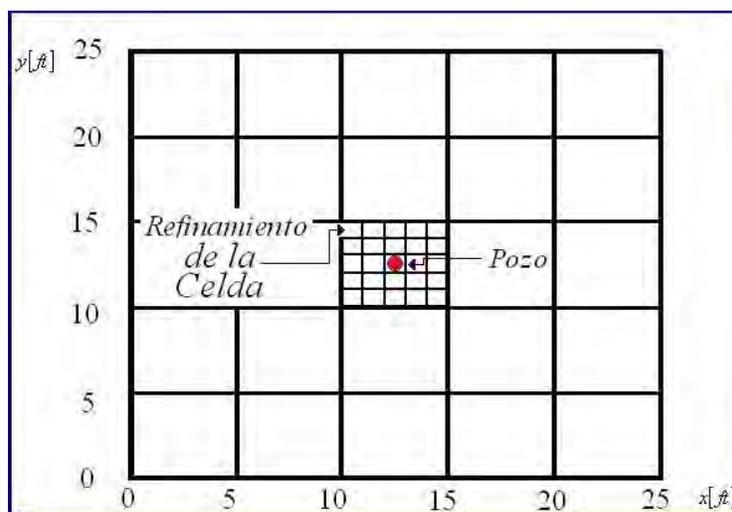


Figura 5.2. Refinamiento de una celda a escala gruesa alrededor de un pozo.

5.3. MALLA DEL MODELO DINÁMICO.

Es necesario realizar una apropiada integración de los datos sísmicos, petrofísicos y geológicos para mejorar la confiabilidad en un modelo final del yacimiento. Un modelo completo del yacimiento en 3D incorpora la estructura, cálculos volumétricos y propiedades que representen el modelo dinámico del yacimiento de acuerdo a un modelo detallado de litofacies. Para obtener este objetivo, se generan múltiples modelos en los que es posible involucrar técnicas geoestadísticas para capturar la incertidumbre en dichos modelos.

Para simular el flujo de fluidos en el yacimiento es necesario partir de un modelo geológico. Este modelo se describe con la intención de utilizar los datos disponibles en cálculos numéricos, dividiendo al yacimiento en un conjunto finito de celdas homogéneas, donde cada una de estas contendrá datos para cada parámetro geológico-petrofísico de acuerdo a la celda de interés o estudio.

Posteriormente se realiza la descripción a detalle de los fluidos en el yacimiento.

5.4. ESCALAMIENTO DE PROPIEDADES PETROFÍSICAS ESTÁTICAS.

En esta sección se describen técnicas de escalamiento de propiedades estáticas. Dichas propiedades incluyen el espesor neto, porosidad y saturación. El escalamiento de estas propiedades es relativamente directo y solo requiere obtener simples promedios. Los detalles se discuten en las siguientes secciones.

5.4.1. ESPESOR NETO.

La técnica de escalamiento para el espesor neto, calcula el espesor de una celda gruesa, con base en los espesores de celdas a escala pequeña que se alojan dentro de esta. Los cálculos de escalamiento consideran que el volumen total de la celda a escala gruesa debe ser igual a la suma del total de volúmenes de las celdas a escala fina anidadas dentro de la celda gruesa. El volumen total V_T de la celda gruesa esta dado por

$$V_T = A_T \tilde{h} \dots\dots\dots(5.1)$$

Donde A_T es la superficie de la celda y \tilde{h} es el espesor neto de la celda gruesa. Usando las propiedades de las celdas a escala fina anidadas en la celda gruesa, puede calcularse el volumen total mediante

$$V_T = \sum_{i=1}^n A_i h_i \dots\dots\dots(5.2)$$

El área total esta dada por

$$A_T = \sum_{i=1}^n A_i \dots\dots\dots(5.3)$$

Donde n es el número de celdas a escala fina que se anidan en la celda a escala gruesa; A_i es el área total, h_i es el espesor neto de la celda a escala fina i . El espesor neto de la celda gruesa se obtiene al combinar las ecuaciones 5.1 – 5.3 y esta dado por

$$\tilde{h} = \frac{\sum_{i=1}^n A_i h_i}{\sum_{i=1}^n A_i} \dots\dots\dots(5.4)$$

En muchos casos, la superficie total de las celdas a escala fina es la misma; como ejemplo se muestra la figura 5.1. Para dichos casos, el espesor neto dado en la ecuación 5.4 se vuelve

$$\tilde{h} = \frac{I}{n} \sum_{i=1}^n h_i \dots\dots\dots(5.5)$$

5.4.2. POROSIDAD.

El cálculo de escalamiento para la porosidad, se basa en el principio de que el volumen de poro para una celda a escala gruesa es igual a la suma del volumen de poro de las celdas a escala fina alojadas dentro de la celda gruesa. El volumen de poro de una celda a escala gruesa, V_{p_T} , esta dado por

$$V_{p_T} = V_T \tilde{\phi} \dots\dots\dots(5.6)$$

Donde V_T es el volumen total y $\tilde{\phi}$ es la porosidad de la celda a escala gruesa. El volumen de poro puede calcularse adicionando los volúmenes de poro de las celdas a escala fina anidadas en la celda gruesa. El V_{p_T} esta dado por

$$V_{p_T} = \sum_{i=1}^n A_i h_i \phi_i \dots\dots\dots(5.7)$$

ϕ_i es la porosidad de la celda a escala fina i . Substituyendo las ecuaciones 5.2 y 5.7 en la ecuación 5.6, se obtiene la ecuación para la porosidad, la cual se convierte en

$$\tilde{\phi} = \frac{\sum_{i=1}^n A_i h_i \phi_i}{\sum_{i=1}^n A_i h_i} \dots\dots\dots(5.8)$$

Hay casos en los que el área total y el espesor neto de las celdas a escala fina son comunes. Para dichos casos, la ecuación 5.8 puede simplificarse. Para el caso en el que el área total de las celdas a escala fina sea constante, la porosidad esta dada por el espesor neto promedio

$$\tilde{\phi} = \frac{\sum_{i=1}^n h_i \phi_i}{\sum_{i=1}^n h_i} \dots\dots\dots(5.9)$$

Para el caso en el que el espesor neto y el área total de las celdas a escala fina son constantes, la porosidad esta dada por

$$\tilde{\phi} = \frac{1}{n} \sum_{i=1}^n \phi_i \dots\dots\dots(5.10)$$

Lo cual es el promedio aritmético de las porosidades a escala fina.

5.4.3. SATURACIÓN DE AGUA.

La técnica de escalamiento para la saturación de agua inicial se basa en un principio de conservación, el cual enuncia que el volumen a escalas fina y gruesa debe ser el mismo. El volumen total de agua, V_{w_T} , esta dado por

$$V_{w_T} = V p_T \tilde{S} w_i \dots\dots\dots(5.11)$$

Donde $\tilde{S} w_i$ es la saturación de agua inicial de la celda a escala gruesa. El volumen de agua esta dado por la suma de todos los volúmenes de las celdas a escala fina anidadas en la celda gruesa.

$$V_{w_T} = \sum_{j=1}^n A_j h_j \phi_j S w_{i_j} \dots\dots\dots(5.12)$$

Donde Sw_{i_j} es la saturación de agua inicial de las celdas a escala fina. La saturación de agua inicial se obtiene al sustituir las ecuaciones 5.7 y 5.11 en la ecuación 5.12, y esta dada por

$$\tilde{S}w_i = \frac{\sum_{j=1}^n A_j h_j \phi_j Sw_{i_j}}{\sum_{j=1}^n A_j h_j \phi_j} \dots\dots\dots(5.13)$$

El lado izquierdo de esta ecuación se refiere en ocasiones, al volumen de poro promedio. Para algunos casos frecuentes, la ecuación 5.13 puede simplificarse para calcular la saturación de agua. Para el caso en el que el área total de las celdas a escala fina sea constante, la saturación de agua inicial es

$$\tilde{S}w_i = \frac{\sum_{j=1}^n h_j \phi_j Sw_{i_j}}{\sum_{j=1}^n h_j \phi_j} \dots\dots\dots(5.14)$$

Para en el caso en el que el área total y el espesor neto de las celdas a escala fina son constantes, la saturación de agua inicial es

$$\tilde{S}w_i = \frac{\sum_{j=1}^n \phi_j Sw_{i_j}}{\sum_{j=1}^n \phi_j} \dots\dots\dots(5.15)$$

Para el caso en el que el área total, espesor neto y la porosidad de las celdas a escala fina sean constantes, la saturación de agua inicial es

$$\tilde{S}w_i = \frac{1}{n} \sum_{j=1}^n Sw_{i_j} \dots\dots\dots(5.16)$$

Lo cual corresponde a un promedio aritmético.

5.5. ESCALAMIENTO DE PROPIEDADES PETROFÍSICAS DINÁMICAS - FLUJO MONOFÁSICO.

En esta sección se discuten las técnicas de escalamiento para propiedades dinámicas considerando flujo monofásico. La permeabilidad es una propiedad dinámica del yacimiento. El cálculo de la permeabilidad para una celda gruesa con base en permeabilidades a escala fina, es más desafiante que el cálculo de propiedades estáticas. Debido al hecho de que la permeabilidad es una propiedad dinámica, el escalamiento depende del relativo arreglo espacial de las permeabilidades a escala fina, condiciones de frontera y anisotropía. Para algunos arreglos espaciales ideales de permeabilidades y condiciones de frontera, se dispone de ecuaciones analíticas para realizar el escalamiento. Para casos más generales, el escalamiento de permeabilidades se basa en soluciones numéricas de ecuaciones diferenciales parciales que gobiernan en el flujo monofásico.

5.5.1. VALORES DE PERMEABILIDAD PARA CASOS IDEALES.

Las permeabilidades para tres casos ideales, están dadas por promedios aritméticos, harmónicos y geométricos¹. Sin embargo los casos ideales, rara vez representan la heterogeneidad de las rocas, aunque proporcionan mediciones útiles para determinar las fronteras de la permeabilidad.

El promedio aritmético proporciona valores de permeabilidad, para el caso en el que existe flujo a lo largo de los estratos, con permeabilidad constante (ver figura 5.3). La permeabilidad de estratos individuales probablemente puede cambiar, pero permanece constante a lo largo de la dirección de flujo. El promedio aritmético de la permeabilidad para la dirección x , $\bar{k}_{x,a}$ esta dado por

ESCALAMIENTO DE PROPIEDADES

$$\tilde{k}_{x,a} = \frac{1}{n} \sum_{i=1}^n k_{x,i} \dots\dots\dots(5.17)$$

Donde n es el número de celdas a escala fina anidadas en la celda a escala gruesa y $k_{x,i}$ es la permeabilidad a lo largo de la dirección x para las celdas a escala fina. Esta ecuación asume que las celdas tienen dimensiones constantes en las direcciones x, y, z . Si el espesor de los estratos varia, la permeabilidad esta dada por

$$\tilde{k}_{x,a} = \frac{\sum_{i=1}^n k_{x,i} h_i}{\sum_{i=1}^n h_i} \dots\dots\dots(5.18)$$

De manera similar, pueden obtenerse las permeabilidades para las direcciones y y z , $\tilde{k}_{y,a}$, $\tilde{k}_{z,a}$.

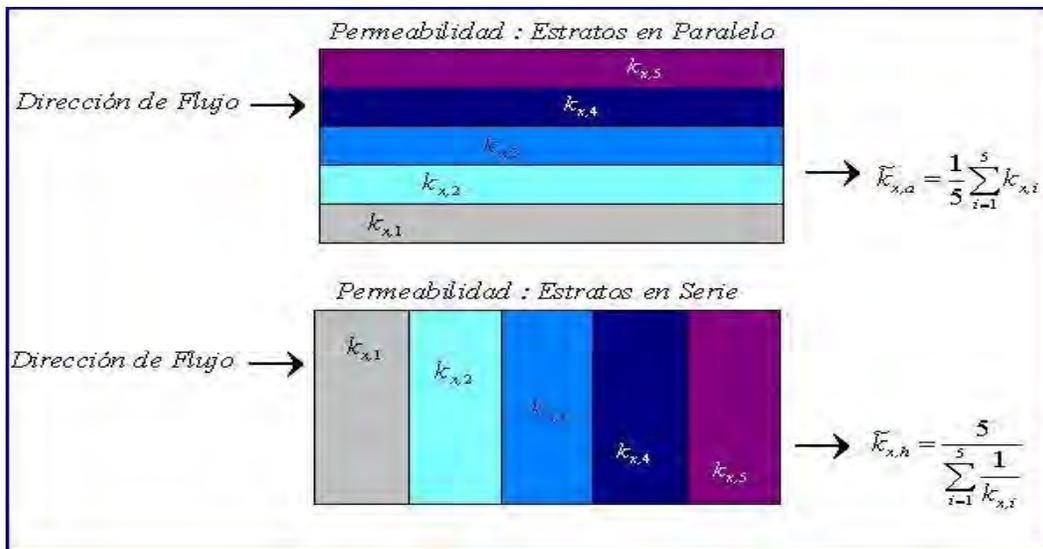


Figura 5.3. Permeabilidades para casos ideales donde las capas están ordenadas de manera paralela y en serie, con respecto a la dirección de flujo.

El promedio armónico proporciona valores de permeabilidad, para el caso en el que existen arreglos de permeabilidad en serie, lo cual cambia solo en la dirección perpendicular a la dirección de flujo (ver figura 5.3). La permeabilidad a lo largo de la dirección x , dada por el promedio armónico es

$$\tilde{k}_{x,h} = \frac{n}{\sum_{i=1}^n \frac{1}{k_{x,i}}} \dots\dots\dots(5.19)$$

De manera similar, pueden derivarse ecuaciones para $\tilde{k}_{y,h}$, $\tilde{k}_{z,h}$.

Para el caso en el que la longitud de la celda varia sobre la dirección de flujo, la permeabilidad esta dada por

$$\tilde{k}_{x,h} = \frac{\sum_{i=1}^n L_i}{\sum_{i=1}^n \frac{L_i}{k_{x,i}}}$$

El promedio geométrico proporciona valores de permeabilidad para el caso en el que las distribuciones de permeabilidad son completamente incorrelacionables (estratos heterogéneos aleatorios), o cuando el rango de correlación del variograma es cero. La media geométrica de la permeabilidad, $\tilde{k}_{x,g}$, para celdas con la misma dirección, esta dada por

$$\tilde{k}_{x,g} = \left[\prod_{i=1}^n k_{x,i} \right]^{n^{-1}} \dots\dots\dots(5.20)$$

Pueden derivarse ecuaciones similares para $\tilde{k}_{y,g}$, $\tilde{k}_{z,g}$.

El promedio aritmético aporta los valores más altos de permeabilidad, para los mismos valores a escala fina, mientras que el promedio armónico aporta los valores más bajos de permeabilidad, y el promedio geométrico proporciona los valores de permeabilidad que caen entre estos dos.

El promedio en potencia es una técnica más general para promediar, la cual incluye las técnicas de promedio antes mencionadas, como casos limitados. Para calcular los valores de permeabilidad $\tilde{k}_{x,p}$, con el promedio en potencia se utiliza la ecuación²

$$\tilde{k}_{x,p} = \left[\frac{1}{n} \sum_{i=1}^n (k_{x,i})^p \right]^{p^{-1}} \dots\dots\dots(5.21)$$

Donde p es la potencia cuyos valores se alojan en un rango de -1 a +1, exceptuando para cero. $p=1$ representa el promedio aritmético, $p=-1$ representa el promedio armónico, y $p=0$, es un caso limitado, que representa el promedio geométrico.

5.5.2. MÉTODO DEL ESTRATO INCOMPLETO.

Este método proporciona una aproximación de escalamiento cuando los valores de permeabilidad varían en más de una dirección. Cardwell y Parsons³ mostraron que la permeabilidad cae entre las fronteras superior e inferior, las cuales están dadas por combinaciones de promedios aritméticos y armónicos. La figura 5.4 ilustra el cálculo de permeabilidades en las fronteras superior e inferior. Para detalles de la derivación de ecuaciones, puede consultarse Cardwell y Parsons³.

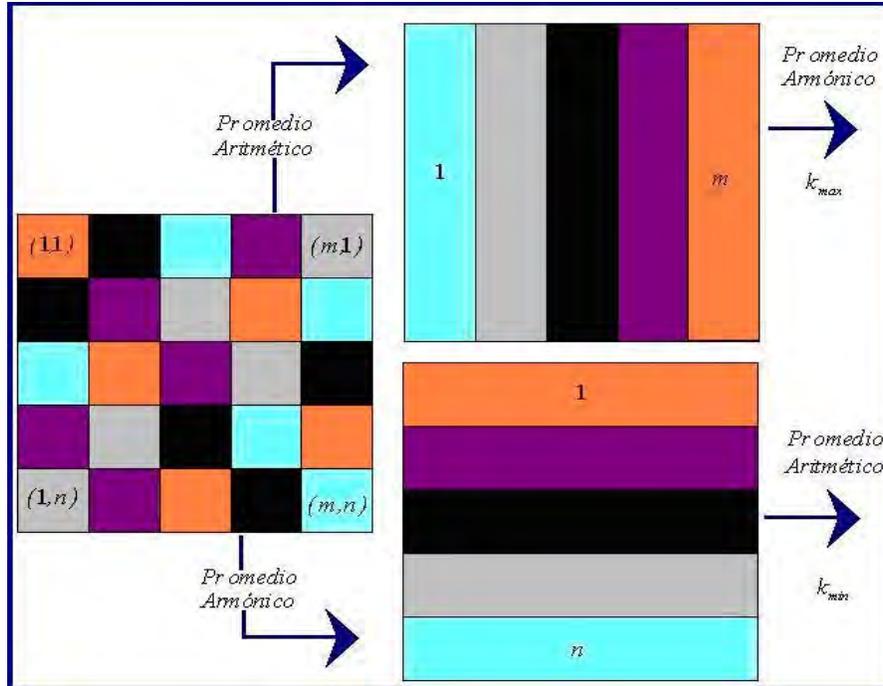


Figura 5.4. Fronteras superior e inferior para el método del estrato incompleto.

La frontera superior, k_{max} , consiste en el promedio armónico de los valores del promedio aritmético.

$$\tilde{k}_{max} = \frac{m}{\sum_{i=1}^m \left[\frac{1}{n} \sum_{j=1}^n k_{i,j} \right]^{-1}} \dots\dots\dots(5.22)$$

La frontera inferior, k_{min} , consiste en el promedio aritmético de los valores del promedio armónico.

$$\tilde{k}_{min} = \frac{1}{n} \sum_{j=1}^n \left[\frac{m}{\sum_{i=1}^m \frac{1}{k_{i,j}}} \right] \dots\dots\dots(5.23)$$

Una aproximación al valor de permeabilidad para una celda gruesa, está dado por el promedio geométrico de las fronteras⁴

$$\tilde{k} = \sqrt{k_{min} k_{max}} \dots\dots\dots(5.24)$$

5.5.3. MÉTODO DE RENORMALIZACIÓN.

Este método utiliza la analogía entre el flujo en medios porosos y el flujo en circuitos eléctricos para calcular la permeabilidad de una celda gruesa. King⁵ introdujo este método. Los valores de permeabilidad se calculan repetidamente en varios pasos. Primero, la malla a escala fina se divide dentro de grupos de cuatro celdas bidimensionales u ocho celdas tridimensionales. La permeabilidad se calcula para cada grupo con las ecuaciones derivadas de circuitos análogos. En los pasos subsecuentes, se crean nuevos grupos de cuatro u ocho celdas considerando las permeabilidades calculadas en el paso anterior. El procedimiento continua hasta que las celdas alcancen la escala requerida. Las ecuaciones 5.25, 5.26, y 5.27 se utilizan, por ejemplo⁵, para calcular la permeabilidad en una configuración 2D, que se muestra en la figura 5.5. Para detalles de la derivación de ecuaciones, puede consultarse King⁵.

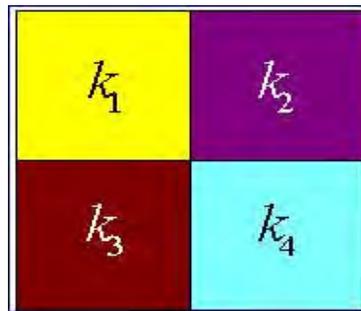


Figura 5.5. Arreglo de permeabilidades para los métodos de renormalización y tensor.

$$\bar{k}_x = \frac{4(k_1 + k_3)(k_2 + k_4)C_1}{(k_1 + k_2 + k_3 + k_4)C_1 + 3C_2} \dots\dots\dots(5.25)$$

$$C_1 = k_2k_4(k_1 + k_3) + k_1k_3(k_2 + k_4) \dots\dots\dots(5.26)$$

$$C_2 = (k_1 + k_2)(k_3 + k_4)(k_1 + k_3)(k_2 + k_4) \dots\dots\dots(5.27)$$

La permeabilidad, \bar{k} , debe calcularse para la dirección y y z . La ecuación 5.25 asume que las fronteras permanecen cerradas en las direcciones

perpendiculares al flujo. King noto, que la derivación para 3D, no alcanza la convención cerrada, como la ecuación 5.25, por lo tanto dicha derivación tridimensional debe resolverse numéricamente.

5.5.4. TENSOR DE PERMEABILIDAD.

Los tensores de permeabilidad permiten representar de manera más rigurosa los valores de permeabilidad. Estos contabilizan correctamente el flujo cruzado creado por la heterogeneidad dentro de una celda de malla. En ocasiones se requiere un tensor de permeabilidad detallado para describir la permeabilidad de las celdas de malla^{6, 7} cuando la estructura de los estratos no esta alineada con ejes coordenados de las celdas, y la heterogeneidad no esta concentrada dentro de los celdas de malla.

En 2D, los tensores de permeabilidad tienen cuatro componentes. Dos componentes contabilizan el flujo a lo largo de las direcciones principales debido a los gradientes de presión en dichas direcciones (k_{xx} y k_{yy}), y las otras dos componentes contabilizan el flujo cruzado debido a los gradientes de presión en direcciones perpendiculares (k_{xy} y k_{yx}). Las ecuaciones que enuncian la ley de Darcy, y describen el campo de velocidad para el tensor de permeabilidad son

$$(v_x)_x = -\frac{k_{xx}}{\mu} \left(\frac{\partial p}{\partial x}\right)_x - \frac{k_{xy}}{\mu} \left(\frac{\partial p}{\partial y}\right)_x \dots\dots\dots(5.28)$$

$$(v_y)_x = -\frac{k_{yx}}{\mu} \left(\frac{\partial p}{\partial x}\right)_x - \frac{k_{yy}}{\mu} \left(\frac{\partial p}{\partial y}\right)_x \dots\dots\dots(5.29)$$

$$(v_x)_y = -\frac{k_{yx}}{\mu} \left(\frac{\partial p}{\partial x}\right)_y - \frac{k_{xy}}{\mu} \left(\frac{\partial p}{\partial y}\right)_y \dots\dots\dots(5.30)$$

y

$$(v_y)_y = -\frac{k_{yx}}{\mu} \left(\frac{\partial p}{\partial x}\right)_y - \frac{k_{yy}}{\mu} \left(\frac{\partial p}{\partial y}\right)_y \dots\dots\dots(5.31)$$

ESCALAMIENTO DE PROPIEDADES

$(v_x)_x$ y $(v_x)_y$ son los componentes de velocidad en la dirección x para el flujo a lo largo de las direcciones x e y respectivamente. $(v_y)_x$ y $(v_y)_y$ son los correspondientes vectores de velocidad en la dirección y . $(\frac{\partial p}{\partial x})_x$ y $(\frac{\partial p}{\partial x})_y$ son los gradientes de presión en la dirección x para el flujo a lo largo de las direcciones x e y respectivamente. $(\frac{\partial p}{\partial y})_x$ y $(\frac{\partial p}{\partial y})_y$ son los correspondientes gradientes de presión en la dirección y .

El método propuesto por Aasum et al.⁸, sirve para calcular el tensor de permeabilidad de una celda gruesa, recurriendo a arreglos de dos por dos, con las celdas a escala fina, como se muestra en la figura 5.5.

Las ecuaciones que calculan el tensor de permeabilidad para arreglos de dos por dos en las celdas son

$$\tilde{k}_{xx} = \frac{\tilde{k}_{xapp}}{1 - \left\{ \begin{array}{c} \left[\begin{array}{c} (\frac{\partial p}{\partial x})_y \\ (\frac{\partial p}{\partial y})_y \end{array} \right] * \left[\begin{array}{c} (\frac{\partial p}{\partial y})_x \\ (\frac{\partial p}{\partial x})_x \end{array} \right] \end{array} \right\}} \dots\dots\dots(5.32)$$

$$\tilde{k}_{yy} = \frac{\tilde{k}_{yapp}}{1 - \left\{ \begin{array}{c} \left[\begin{array}{c} (\frac{\partial p}{\partial x})_y \\ (\frac{\partial p}{\partial y})_y \end{array} \right] * \left[\begin{array}{c} (\frac{\partial p}{\partial y})_x \\ (\frac{\partial p}{\partial x})_x \end{array} \right] \end{array} \right\}} \dots\dots\dots(5.33)$$

$$\tilde{k}_{xy} = \tilde{k}_{xx} \left[\begin{array}{c} (\frac{\partial p}{\partial x})_y \\ (\frac{\partial p}{\partial y})_y \end{array} \right] \dots\dots\dots(5.34)$$

$$\tilde{k}_{yx} = \tilde{k}_{yy} \left[\begin{array}{c} (\frac{\partial p}{\partial y})_x \\ (\frac{\partial p}{\partial x})_x \end{array} \right] \dots\dots\dots(5.35)$$

\tilde{k}_{xapp} y \tilde{k}_{yapp} son las permeabilidades aparentes x e y , y se utilizan como permeabilidades, si el simulador no esta equipado para manejar la formulación del tensor de permeabilidad. \tilde{k}_{xapp} , se reduce a \tilde{k}_{xx} , y \tilde{k}_{yapp} , se reduce a \tilde{k}_{yy} , si es que la razón de gradientes de potencial de las ecuaciones 5.32 y 5.33 es cero. La razón de gradiente de potencial se define como la relación de los gradientes

de presión transversal y longitudinal. $\left[\frac{(\frac{\partial p}{\partial x})_y}{(\frac{\partial p}{\partial y})_y} \right]$, es la razón de gradiente de

potencial inducida por el flujo a lo largo del eje y ; mientras $\left[\frac{(\frac{\partial p}{\partial y})_x}{(\frac{\partial p}{\partial x})_x} \right]$ es la razón

de gradiente de potencial inducida por el flujo a lo largo del eje x . Para detalles de la derivación de ecuaciones, puede consultarse Aasum et al.⁸.

El método de tensor de permeabilidad puede extenderse a flujo en 3D⁹, sin embargo las ecuaciones se vuelven mucho más complejas. Adicionalmente a los métodos antes discutidos, se han desarrollado otro tipo de métodos analíticos^{10, 11}.

5.5.5. FLUJO EN ESTADO ESTACIONARIO - SOLUCIÓN NUMÉRICA.

Este método utiliza la solución numérica de ecuaciones diferenciales parciales que gobiernan las condiciones del flujo monofásico en estado estacionario para calcular la permeabilidad de una celda e escala gruesa. El desarrollo de la ecuación para este sistema, contempla el empleo de celdas a escala fina anidadas en una celda gruesa. La ecuación empleada es

$$\frac{\delta}{\delta x} \left(k_x \frac{\delta p}{\delta x} \right) + \frac{\delta}{\delta y} \left(k_y \frac{\delta p}{\delta y} \right) + \frac{\delta}{\delta z} \left(k_z \frac{\delta p}{\delta z} \right) = 0 \dots\dots\dots(5.36)$$

ESCALAMIENTO DE PROPIEDADES

Esta ecuación se resuelve para condiciones de presión constante, a la entrada y a la salida de las caras de la celda, lo cual es perpendicular a la dirección en la que se efectuará el cálculo de la permeabilidad, con las fronteras cerradas en las otras direcciones. Las condiciones de frontera que permiten calcular la permeabilidad para la dirección x son

$$p = (x = 0) = p_0 \dots\dots\dots(5.37)$$

$$p = (x = L) = p_L \dots\dots\dots(5.38)$$

$$\left. \frac{\partial p}{\partial y} \right|_{y=0, L_y} = 0 \dots\dots\dots(5.39)$$

$$\left. \frac{\partial p}{\partial z} \right|_{z=0, L_z} = 0 \dots\dots\dots(5.40)$$

La solución numérica para las ecuaciones 5.36 - 5.40 permite obtener las distribuciones de presión y gastos de flujo en el sistema a escala fina. La suma de los gastos de flujo a la salida de la celda, sirven para calcular el gasto total de flujo, Q_x . Posteriormente, la permeabilidad para la celda gruesa en la dirección x , se calcula mediante la ecuación de Darcy, escrita como

$$\tilde{k}_x = \frac{Q_x \mu L_x}{A(p_0 - p_L)} \dots\dots\dots(5.41)$$

Para las direcciones y y z , se derivan ecuaciones similares utilizando condiciones de frontera apropiadas en las ecuaciones 5.37 – 5.40.

5.6. ESCALAMIENTO DE PROPIEDADES PETROFÍSICAS DINÁMICAS - FLUJO MULTIFÁSICO.

Bajo condiciones de flujo multifásico, se deben considerar dos propiedades adicionales: la permeabilidad relativa y la presión capilar.

El escalamiento de la permeabilidad relativa recibe mucho más atención conforme mejora la habilidad para generar una gran cantidad de celdas de malla que permiten obtener descripciones del yacimiento. En general, los procedimientos de escalamiento pueden dividirse en dos grupos: analíticos y numéricos. Los procedimientos analíticos se aplican bajo casos limitados y bajo reducidas descripciones de heterogeneidades. Sin embargo, dichos procedimientos, son utilizados para realizar aplicaciones en muchas situaciones prácticas. Los métodos numéricos, por otro lado son más generales y demandan mayor poder computacional.

5.6.1. MÉTODOS ANALÍTICOS.

Cuando se considera flujo multifásico en el medio poroso, puede pensarse en tres tipos de fuerzas importantes que afectan el comportamiento de flujo, las cuales son viscosidad, gravedad y capilaridad. La magnitud relativa de cada fuerza determina la manera en la cual se afecta el comportamiento del flujo. En casos limitados, una de las fuerzas antes mencionadas, domina el comportamiento del flujo sobre las demás; en esta situación se pueden aplicar procedimientos de escalamiento analítico. En general, los dos casos limitados principales que pueden considerarse son el flujo no cruzado y el equilibrio vertical. Bajo una condición de flujo no cruzado, no existe comunicación en la dirección vertical, y la única comunicación existen es con el pozo. Esto es posible si la permeabilidad vertical es muy pequeña, o si existe una barrera de permeabilidad más baja entre dos estratos. Bajo condiciones de equilibrio vertical, el gradiente de potencial en la dirección vertical se considera como cero. El flujo en la dirección vertical es tan rápido que la presión en dicha dirección alcanza el equilibrio rápidamente, Zapata y Lake¹², introducen un número adimensional para predecir la presencia de un caso limitado particular.

$$R_L = \frac{L}{H} \sqrt{\frac{k_v}{k_h}} \dots\dots\dots(5.42)$$

Donde k_v es la permeabilidad vertical que se obtiene con un promedio armónico, k_h es la permeabilidad horizontal obtenida con un promedio aritmético, L es la longitud en la dirección horizontal y H es el espesor. Zapata y Lake¹², observaron que si $R_L > 10$, permanece la suposición de que existe equilibrio vertical, y si $R_L < 0.1$, permanece la suposición de que no existe flujo cruzado. Para valores entre 0.1 y 10, no permanece la suposición de los casos limitados.

Bajo la suposición de equilibrio vertical, hay cuatro posibilidades a considerar¹³:

- **Equilibrio viscoso en flujo cruzado.** Bajo esta condición, las fuerzas gravitacional y capilar son insignificantes, y solo la fuerza debido a la viscosidad establece el equilibrio vertical.
- **Equilibrio capilar-gravitacional.** Ambas fuerzas, capilar y gravitacional, dominan para establecer la distribución de saturaciones, por lo tanto, dominan el comportamiento del flujo.
- **Equilibrio gravitacional.** La fuerza gravitacional es la fuerza dominante.
- **Equilibrio capilar.** Solo la fuerza capilar determina la distribución de saturaciones en el yacimiento.

No existe una guía específica para determinar las fuerzas dominantes bajo ciertas condiciones. Sin embargo, pueden examinarse dos números adimensionales que pueden auxiliar para encontrar relaciones entre las fuerzas

$$N_c = \frac{1.127 \times 10^{-3} k_h A p_c}{\mu q L} \dots\dots\dots(5.43)$$

Donde k_h es la permeabilidad horizontal en md , p_c es la presión capilar en $psia$, μ es la viscosidad en cp , q es el gasto de flujo en la dirección horizontal en $\frac{B}{D}$, L es la longitud horizontal en ft , A es el área de la sección cruzada en ft^2 . N_c representa la razón entre las fuerzas debido a la capilaridad y la viscosidad.

El otro número es

$$N_g = \frac{1.127 \times 10^{-3} \sqrt{k_h k_v} (0.4335) \Delta \gamma A}{q \mu} \dots\dots\dots(5.44)$$

Donde $\Delta \gamma$ es la diferencia de gravedades específicas, A es el área de la sección cruzada en ft^2 , q es el gasto de flujo en la dirección horizontal en $\frac{B}{D}$, μ es la viscosidad en cp , y k_h, k_v , son las permeabilidades horizontal y vertical promedio respecto al espesor, respectivamente. N_g representa la razón entre las fuerzas debido al gravedad y a la viscosidad. Existe evidencia, la cual indica, que para $N_g > 0.1$ el flujo puede considerarse como dominado por la fuerza gravitacional¹⁴. No existe un criterio para determinar dominación debido a la fuerza capilar.

Existen soluciones analíticas para casos limitados. Dichas soluciones se restringen para yacimientos que se representan por estratos con diferentes propiedades para cada uno. No obstante, se asume que cada capa es homogénea. También se restringe el desarrollo de ecuaciones para procesos de inyección de agua, aunque es posible generar ecuaciones para otro tipo de desplazamientos.

Flujo no cruzado. Al definir al yacimiento como estratificado, (ver figura 5.6), los estratos se ordenan de acuerdo a la velocidad a la cual el fluido fluye entre estos. El estrato en el que el flujo es más rápido se encuentra en la cima y su rompimiento será primero, el estrato en el que el flujo es más lento se encuentra en el fondo y su rompimiento será el último. La velocidad in-situ para un estrato determinado se relaciona con la cantidad $\frac{k}{\phi\Delta Sw}$, donde k es la permeabilidad; ϕ es la porosidad, y ΔSw es $(1 - Sw_i - So_r)$. Sw_i es la saturación de agua inicial y So_r es la saturación de aceite residual. Se asume que el desplazamiento en cada estrato es de tipo pistón, y aun más, se supone que la relación de movilidad es la misma para todos lo estratos.

$$M = \frac{\frac{k'_{rw}}{\mu_w}}{\frac{k'_{ro}}{\mu_o}} \dots\dots\dots(5.45)$$

En la ecuación 5.45 k'_{rw} y k'_{ro} son el punto final de las permeabilidades relativas del agua y aceite, respectivamente, μ_o y μ_w son las viscosidades del aceite y agua respectivamente.

Al suponer que el rompimiento ocurre a través del estrato l , se tiene la siguiente expresión

$$R_{jl} = \frac{\left(\frac{\phi\Delta Sw}{kk'_{rw}}\right)_l}{\left(\frac{\phi\Delta Sw}{kk'_{rw}}\right)_j} \dots\dots\dots(5.46)$$

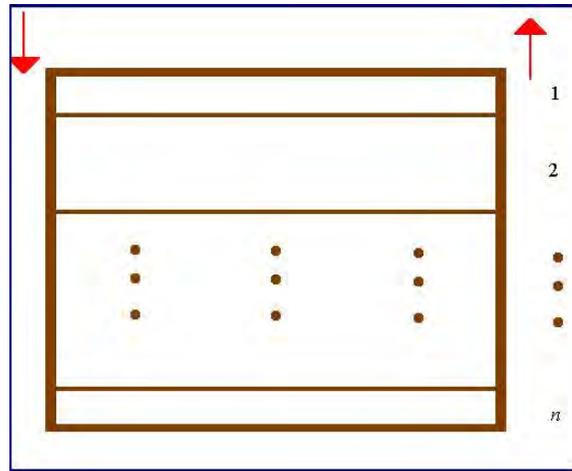


Figura 5.6. Representación de un yacimiento en estratos con n capas.

Donde j es cualquier capa por debajo de l . La distancia que recorre el frente de inyección, en la capa j , en el tiempo l -ésimo en que se presenta el rompimiento, y esta se calcula como

$$x_{Dj} = \frac{-M + \sqrt{M^2 + R_{jl}(1 - M^2)}}{1 - M} , si \ M \neq 1 \dots\dots\dots(5.47a)$$

$$x_{Dj} = R_{jl} , si \ M = 1 \dots\dots\dots(5.47b)$$

Donde x_D es la distancia adimensional en la dirección x y es igual a $\frac{x}{L}$.

Definiendo el peso aritmético promedio para la permeabilidad como

$$\tilde{k}H = \frac{\sum_{i=1}^n k_i h_i}{n} \dots\dots\dots(5.48)$$

Donde H es el espesor total. La permeabilidad relativa se calcula mediante

$$\tilde{k}_{rw} = \frac{\sum_{i=1}^n (kk'_{rw} h)_i}{\tilde{k}H} \dots\dots\dots(5.49)$$

Para el agua.

Y

$$\tilde{k}_{ro_i} = \frac{\sum_{i=l+1}^n \frac{(kk'_{ro} h)_i}{1 + x_{Di}}}{M \tilde{k}H} \dots\dots\dots(5.50)$$

Para el aceite.

Es de notar que la ecuación 5.50 se puede simplificar cuando $M = 1$. La saturación promedio es

$$\tilde{S}_{w_l} = \frac{\sum_{i=1}^l [h\phi(1 - S_{or})]_i \sum_{i=l+1}^n \{[h\phi(1 - S_{or})]_i x_{Di} + (h\phi S_{wi})(1 - x_{Di})\}}{\sum_{i=1}^l [h\phi]_i} \dots\dots\dots(5.51)$$

Al repetir los cálculos para l , variando entre 1 y n , puede calcularse el rango para las permeabilidades relativas.

Equilibrio vertical. Bajo condiciones de equilibrio vertical, existen cuatro posibles situaciones:

- Flujo cruzado vertical debido a fuerzas viscosas.
- Desplazamiento dominado por fuerzas gravitacionales.
- Desplazamiento dominado por fuerzas gravitacionales y fuerzas capilares.
- Desplazamiento dominado por fuerzas capilares.

Las expresiones analíticas más practicas respecto a la permeabilidad relativa se refieren al caso en el que domina el desplazamiento debido a las fuerzas gravitacionales¹⁵.

Flujo cruzado dominado por fuerzas gravitacionales. Cuando el gasto de inyección es bajo, y los efectos de capilaridad son insignificantes, puede

suponerse flujo cruzado dominado por fuerzas gravitacionales, En muchos casos, es posible suponer que el flujo es dominado por fuerzas gravitacionales en presencia de fuerzas capilares, si

$$Capilaridad \text{ en la zona de Transición } \gg H \dots\dots\dots(5.52)$$

Lo anterior significa que la zona de transición es tan larga, comparada con el espesor total, que los cambios en la saturación respecto al espesor, pueden ser despreciables.

En el desplazamiento dominado por fuerzas gravitacionales, se asume que el agua se segrega hacia el fondo, y el rompimiento se presenta en las capas más pequeñas. La permeabilidad relativa, empezando por la capa más pequeña puede calcularse como

$$\tilde{k}_{rw_i} = \frac{\sum_{i=1}^l (kk'_{rw})_i h_i}{\tilde{k}H} \dots\dots\dots(5.53)$$

Para el agua.

Y

$$\tilde{k}_{ro_i} = \frac{\sum_{i=1}^n (kk'_{ro})_i h_i}{\tilde{k}H} \dots\dots\dots(5.54)$$

Para el aceite.

Además

$$\tilde{S}_{w_i} = \frac{\sum_{i=1}^l [h\phi(1 - S_{w_i})_i] + \sum_{i=l+1}^n [h\phi S_{w_i}]_i}{\sum_{i=1}^n [h\phi]_i} \dots\dots\dots(5.55)$$

5.6.2. MÉTODOS NUMÉRICOS.

Los métodos analíticos presentan la ventaja de ser simples de utilizar además de que su implementación es benéfica. Sin embargo, dichos métodos no pueden aplicarse en todos los tipos de condiciones necesarias.

- La aplicación de los métodos analíticos se restringe para yacimientos estratificados. En un típico proceso de escalamiento, interesa escalar la permeabilidad relativa de múltiples mallas, ya sea en 2D o en 3D, a una malla equivalente. Las propiedades del yacimiento como permeabilidad, porosidad y saturación de fluidos, pueden variar en las tres direcciones.
- Los métodos analíticos son de gran utilidad bajo condiciones limitadas. Si la combinación de fuerzas es tal, que estas desempeñan un papel importante en la determinación de la distribución de fluidos, no es posible aplicar una simple solución analítica.
- Los métodos analíticos suponen que las permeabilidades relativas son independientes de las condiciones de frontera. Por lo tanto, las propiedades no dependerán del gasto de flujo o de aquellos fluidos que se inyectan bajo condiciones de presión variables o constantes. Estas suposiciones se vuelven verdaderas, siempre y cuando las saturaciones de los fluidos desplazantes se muevan a velocidades constantes. En otras palabras, puede aplicarse la solución analítica del método de Buckley-Leverett para desplazamiento en una dimensión. La curva de flujo fraccional normalmente puede exhibir un frente de inyección, así como un cambio gradual con la saturación después de que se presente el rompimiento. La velocidad para una saturación dada, se muestra mediante

$$\frac{dx}{dt} \propto \left(\frac{df_w}{dS_w} \right)_{S_w} \dots\dots\dots(5.56)$$

Donde $\left(\frac{df_w}{dS_w} \right)$ es la tangente para una saturación dada en la curva de flujo fraccional. De acuerdo a dicha curva, la pendiente después del rompimiento, se incrementa conforme la saturación se incrementa. Desafortunadamente, en la presencia de heterogeneidades del yacimiento, este comportamiento ideal del fenómeno, no podría aplicarse. Las saturaciones se mueven a velocidades constantes, por lo tanto el principio básico de la teoría de Buckley-Leverett^{16, 17}, no se respeta. Bajo estas condiciones, las permeabilidades relativas deben definirse entonces como funciones de las condiciones iniciales de frontera.

Cuando aumentan las condiciones previamente mencionadas, los procedimientos numéricos proporcionan un camino en el cual las permeabilidades relativas pueden estimarse. En muchos artículos técnicos se hace referencia a las permeabilidades relativas como pseudofunciones; es importante la distinción entre este par de terminologías. Las permeabilidades relativas suponen ser propiedades del sistema roca-fluidos independientes a las condiciones de frontera, mientras que las pseudofunciones no solo dependen del sistema roca-fluidos, también dependen de las condiciones de frontera.

De esta manera, se supone que las pseudofunciones permiten realizar cálculos para el comportamiento de flujo no ideal, adecuándose a condiciones apropiadas del flujo, bajo las que se conduzca la simulación actual del mismo.

Stone¹⁸ propuso que algunas pseudofunciones pueden obtenerse de acuerdo a relaciones de movilidad. El flujo fraccional medible durante una simulación puede expresarse como

$$f_w = \frac{I}{I + \frac{\tilde{k}_{ro} \mu_o}{\tilde{k}_{rw} \mu_w}} \dots\dots\dots(5.57)$$

La ecuación anterior esta escrita en ausencia de efectos capilares y gravitacionales.

La movilidad total puede calcularse si se conoce el gasto del flujo proveniente de las celdas de malla a escala fina, esto es

$$M_t = \left(\frac{\tilde{k}_{rw}}{\mu_w} + \frac{\tilde{k}_{ro}}{\mu_o} \right) = \frac{\sum_{i=1}^n k_i h_i \left(\frac{k_{ro_i}}{\mu_{oi}} + \frac{k_{rw_i}}{\mu_{wi}} \right)}{\sum_{i=1}^n k_i h_i} \dots\dots\dots(5.58)$$

Al conocer el flujo fraccional y la movilidad total, las permeabilidades pseudorelativas pueden calcularse mediante

$$\tilde{k}_{rw} = M_t f_w \mu_w \dots\dots\dots(5.59)$$

$$\tilde{k}_{ro} = M_t (1 - f_w) \mu_o \dots\dots\dots(5.60)$$

El método descrito es riguroso; sin embargo, en presencia de significantes fuerzas gravitacionales y capilares, el método arroja pobres resultados. Aunado a esto, el método resulta ineficiente si existen importantes variaciones en la movilidad¹⁹.

El método de Stone presenta mejores resultados cuando se sugiere otro tipo de alternativas para calcular la movilidad total (ecuación 5.58)²⁰. Sin embargo, en ocasiones se desprecia la fuerza gravitacional y puede no reproducir exactamente la solución a escala fina en una celda gruesa.

En realidad aun no existe un método único que permita calcular adecuadamente pseudofuncines dinámicas. Idealmente, cada celda gruesa depende de las distribuciones de permeabilidad a escala fina, y debe de asignarse un arreglo de permeabilidades pseudorelativas, en el cual la variación dependa de las condiciones de frontera. Esto no es muy práctico.

Una posible alternativa, seria dividir al yacimiento en muchas celdas gruesas pequeñas, de tal manera que a cada celda pudiera asignársele un solo tipo de roca, y por lo tanto, un solo conjunto de permeabilidades relativas a escala fina, generando un número suficiente de descripciones representativas, así como pseudofunciones para cada tipo de roca, las cuales serán dependientes del gasto de flujo u otro tipo de condiciones de frontera. Por lo tanto es necesario verificar que las celdas gruesas permitan generar resultados de simulación a escala fina, mediante un modelo para las condiciones de frontera más representativas. Una vez que se realiza la verificación, las pseudofunciones se utilizan para obtener los ejercicios de simulación¹⁹.

5.7. PRESIÓN CAPILAR.

Si las fuerzas capilares son importantes, la generación del escalamiento de presiones capilares es relativamente directa. A escala fina cuando se conoce la distribución de saturaciones, para cada celda de malla a dicha escala, se asignan valores de presión capilar. Este valor depende de la saturación además de la permeabilidad del tipo de roca. Es posible calcular la presión capilar promedio mediante la expresión

$$\bar{p}_c = \frac{\sum_{i=1}^n \phi_i h_i p_{c_i}}{\sum_{i=1}^n \phi_i h_i} \dots\dots\dots(5.61)$$

La presión capilar puede determinarse si se conoce la saturación de agua promedio.

5.8. APLICACIÓN PRÁCTICA.

En este último ejercicio se presenta e ilustra el concepto de escalamiento de propiedades de dos maneras. La primera se realiza para un conjunto de registro geofísicos, cuyo resultado permite realizar el modelado petrofísico y construir así la malla fina. En la segunda se ilustra, como se puede llevar a cabo el escalamiento de propiedades de dicha malla fina a una malla gruesa.

Se eligieron la permeabilidad y la porosidad como propiedades petrofísicas representativas, para realizar este ejercicio. La malla fina de construye con muy poca información, se utilizan como datos de entrada catorce registros geofísicos de porosidad y permeabilidad respectivamente. El modelado petrofísico se realiza para un yacimiento dividido en cinco subzonas en el que existe un previo estudio geológico, estructural y estratigráfico.

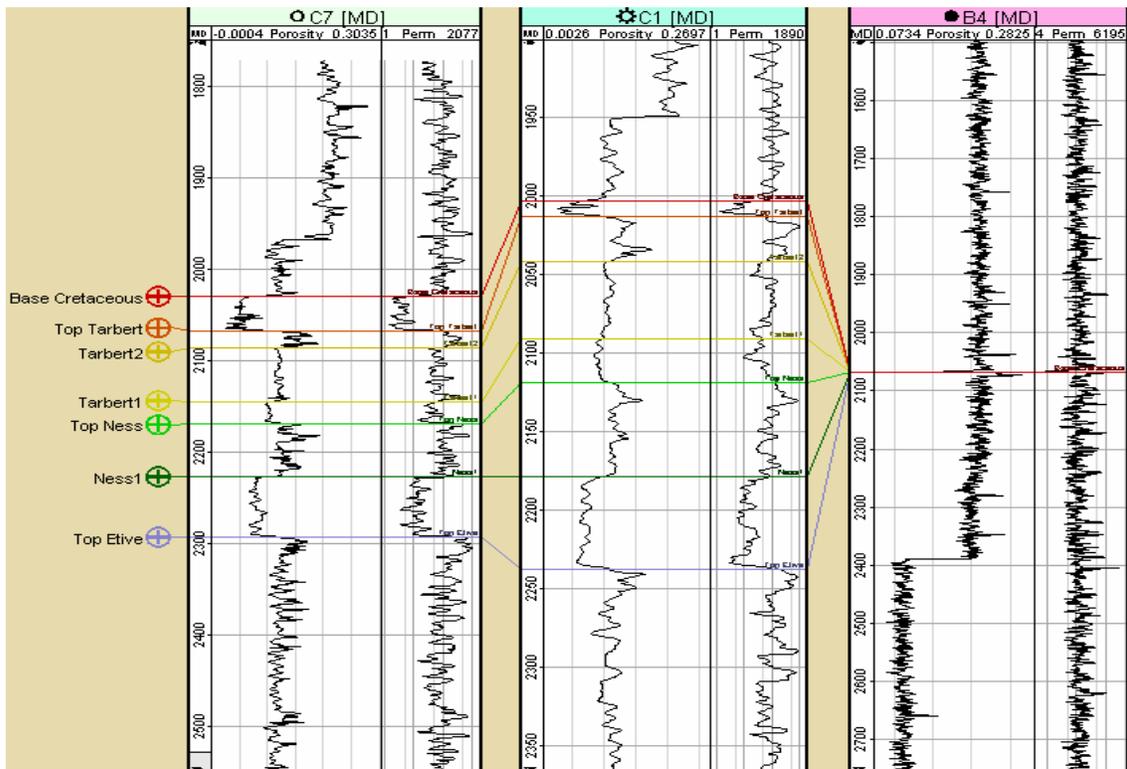


Figura 5.7a. Conjunto de registros de porosidad y permeabilidad originales.

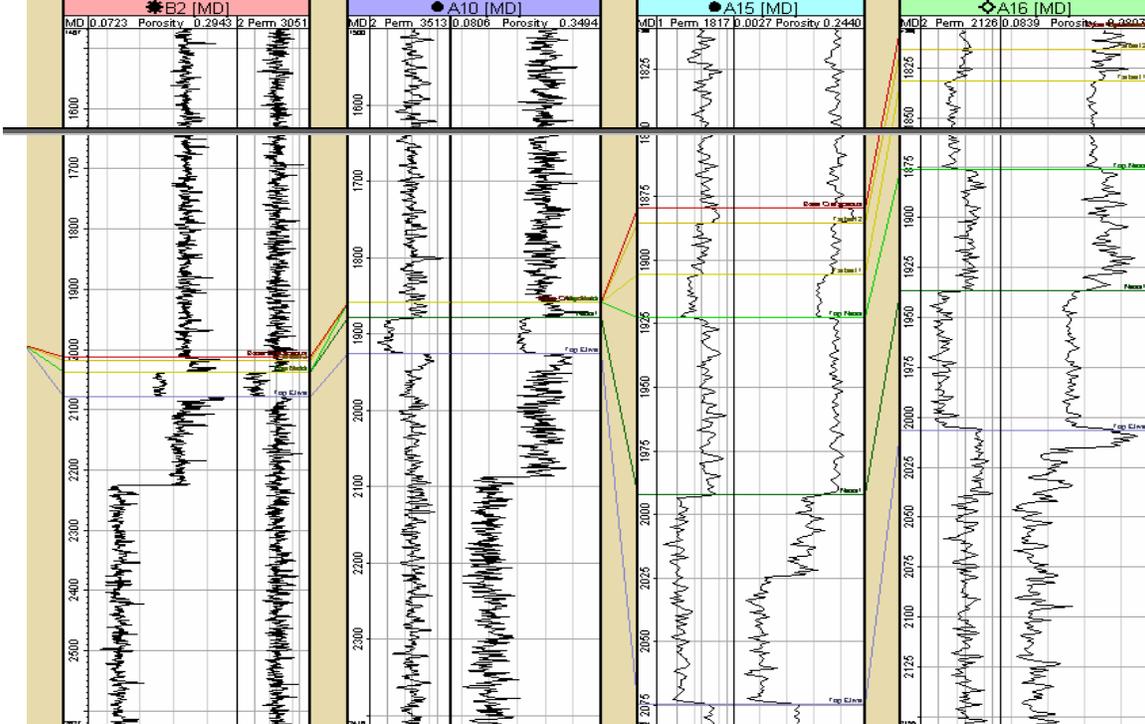


Figura 5.7b. Conjunto de registros de porosidad y permeabilidad originales.

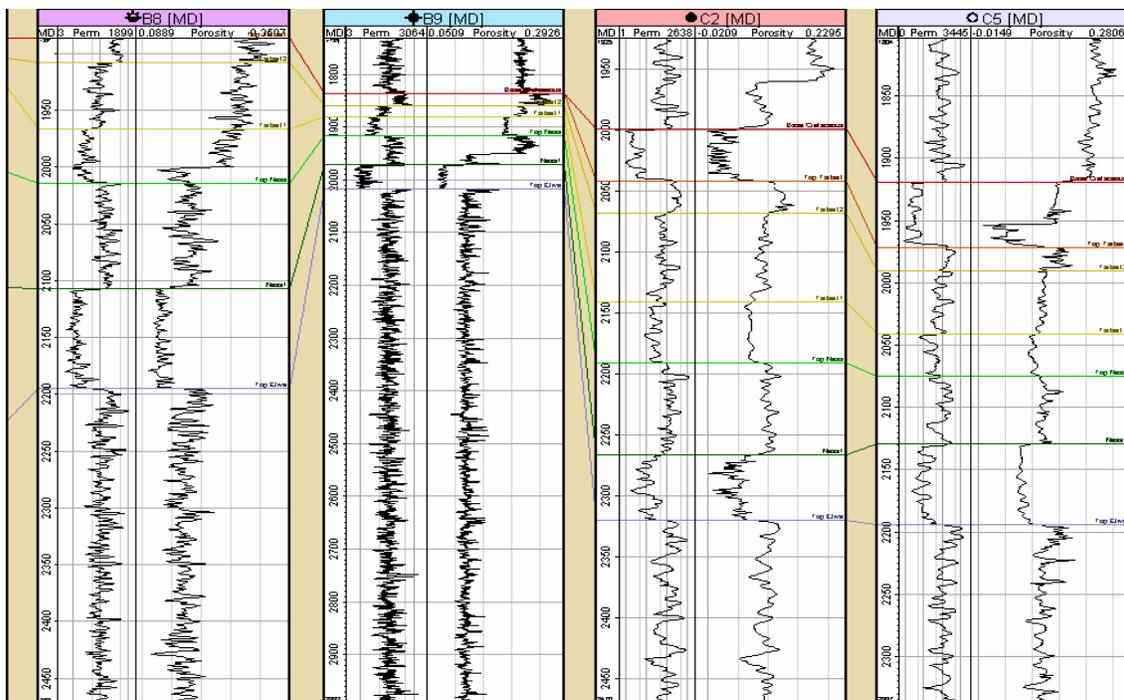


Figura 5.7c. Conjunto de registros de porosidad y permeabilidad originales.

ESCALAMIENTO DE PROPIEDADES

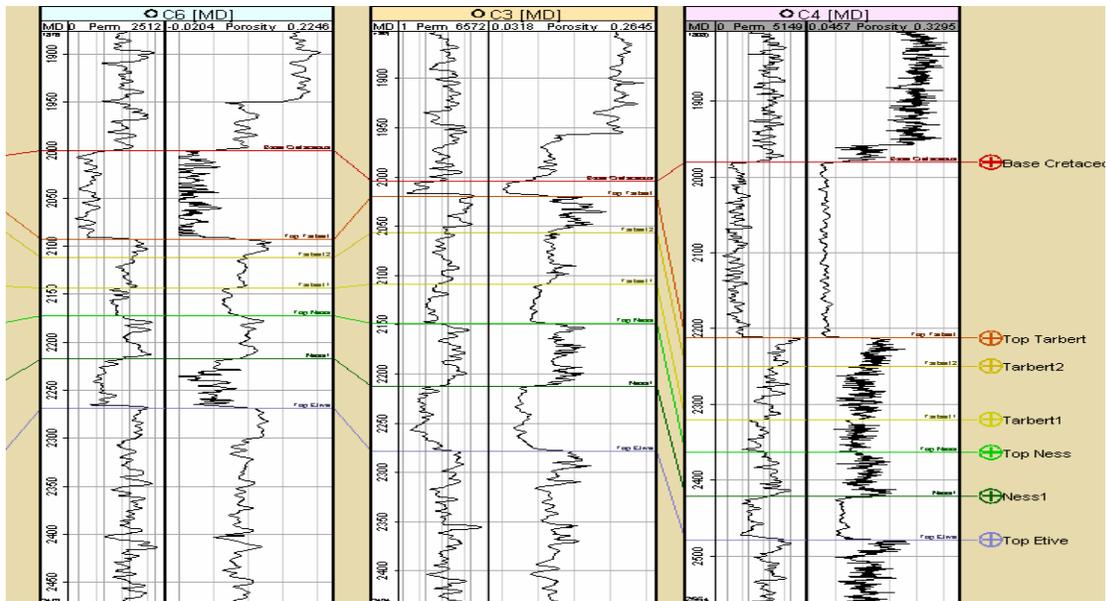


Figura 5.7d. Conjunto de registros de porosidad y permeabilidad originales.

Con los datos del conjunto de registros mostrados, se obtiene el escalamiento vertical de los mismos; para la porosidad se emplea la técnica de promedio aritmético y para la permeabilidad la de promedio geométrico. En las siguientes figuras se muestra el resultado que se mediante una visualización 3D realizada en PETREL 2004^{MR}.

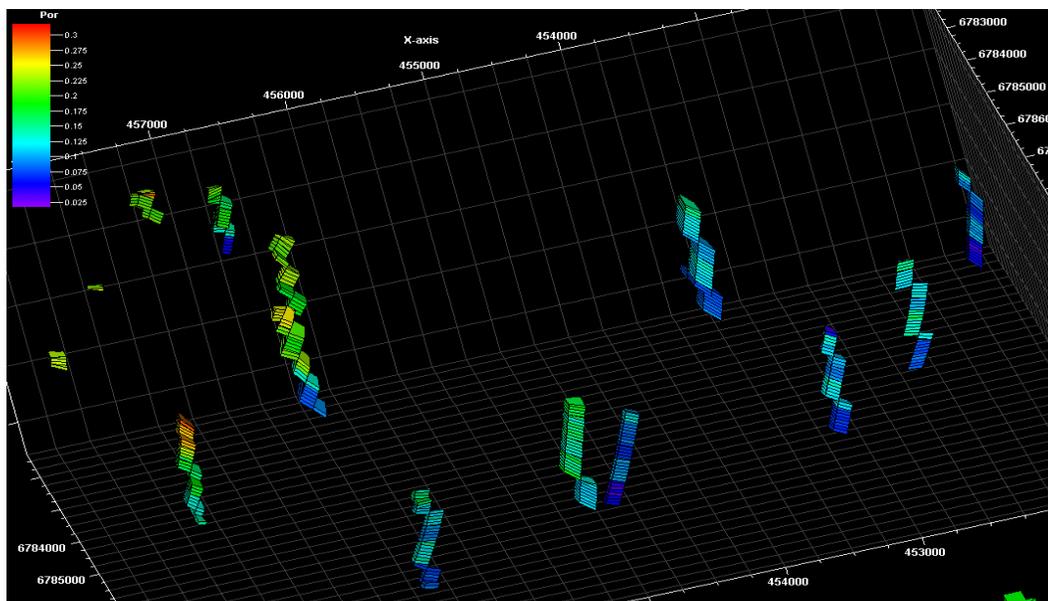


Figura 5.8. Escalamiento de registros de porosidad.

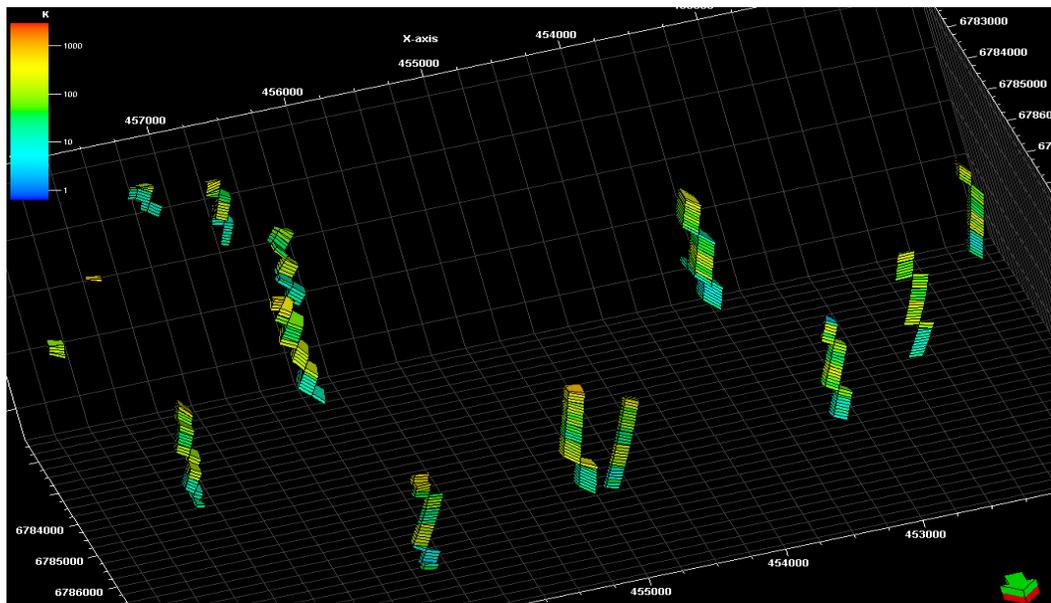


Figura 5.9. Escalamiento de registros de permeabilidad.

Posteriormente al escalamiento vertical se prosigue a realizar un análisis de datos para obtener los parámetros óptimos que permiten realizar la construcción de un modelado petrofísico.

Para la porosidad, en el análisis de datos se crea un variograma experimental el cual posteriormente se ajusta de acuerdo a el análisis practico que se trata en el capítulo tres.

Dichos parámetros así como el variograma ajustado se muestran a continuación.

Orientación de la zona de mayor anisotropía = -74°

Meseta = 1.52

Nugget = 0.045

Alcance Mínimo (Menor anisotropía) = 2,860 m

Alcance Máximo (Mayor anisotropía) = 4,299 m

ESCALAMIENTO DE PROPIEDADES

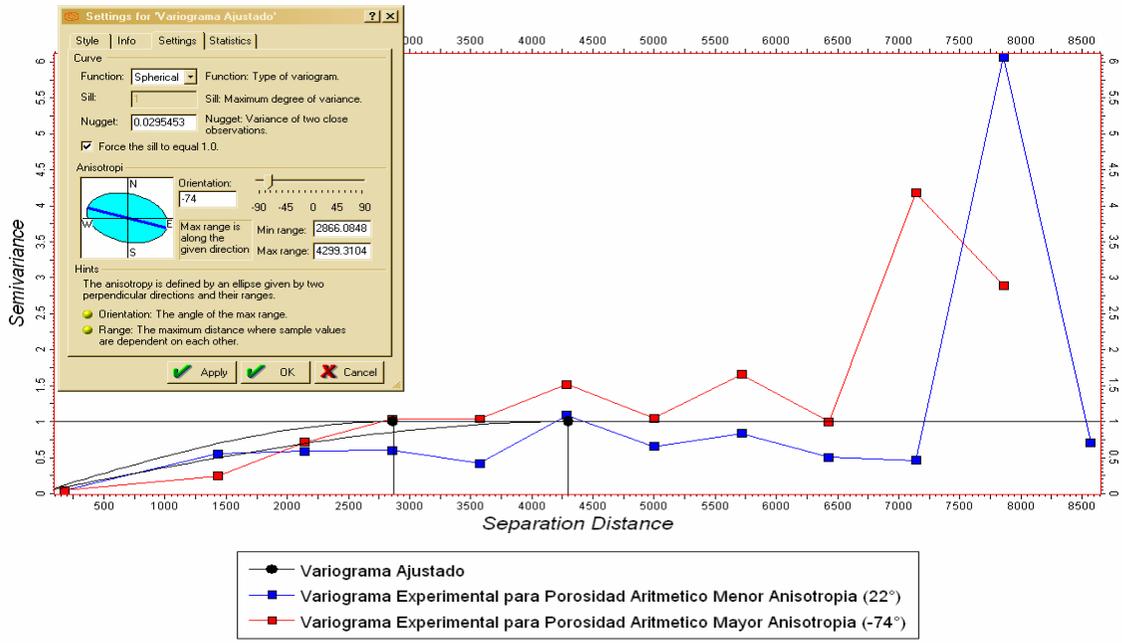


Figura 5.10. Porosidad: Ajuste de variogramas para la obtención de parámetros óptimos.

Para realizar el modelado petrofísico referente a la porosidad se elige el algoritmo por simulación Gaussiana secuencial el cual se construye con PETREL 2004^{MR}. La siguiente figura ilustra el resultado que se obtiene.

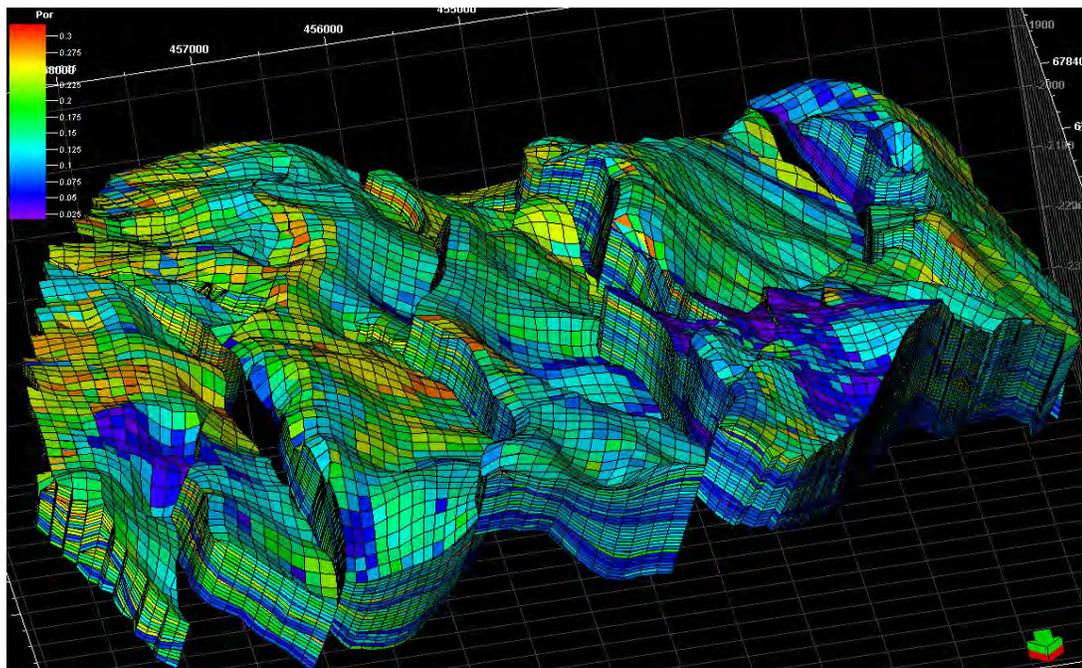


Figura 5.11. Modelado de porosidad mediante una simulación Gaussiana secuencial.

Para la permeabilidad, en el análisis de datos se crea un variograma experimental de igual manera a la porosidad. Los parámetros óptimos así como el variograma ajustado se muestran a continuación.

Orientación de la zona de mayor anisotropía = 86°

Meseta = 1

Nugget = 0.129

Alcance Mínimo (Menor anisotropía) = 2,789 m

Alcance Máximo (Mayor anisotropía) = 3,184 m

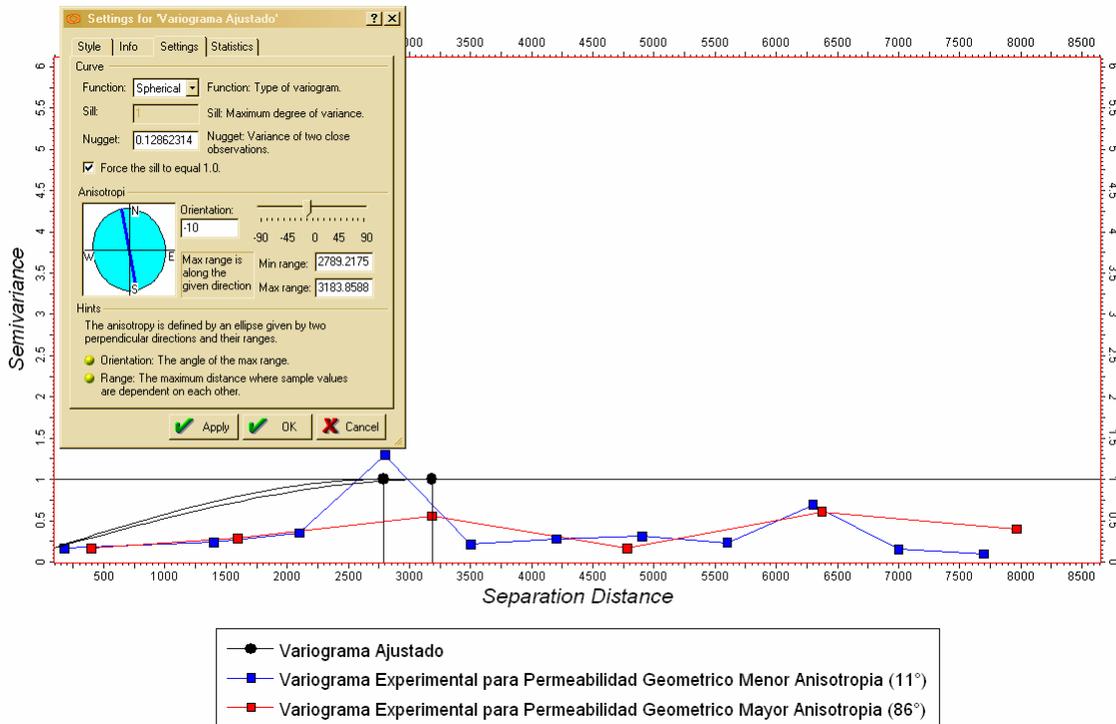


Figura 5.12. Permeabilidad: Ajuste de variogramas para la obtención de parámetros óptimos.

Para realizar el modelado petrofísico referente a la permeabilidad se elige el algoritmo por simulación Gaussiana secuencial el cual se construye con PETREL 2004^{MR}. La siguiente figura ilustra el resultado que se obtiene.

ESCALAMIENTO DE PROPIEDADES

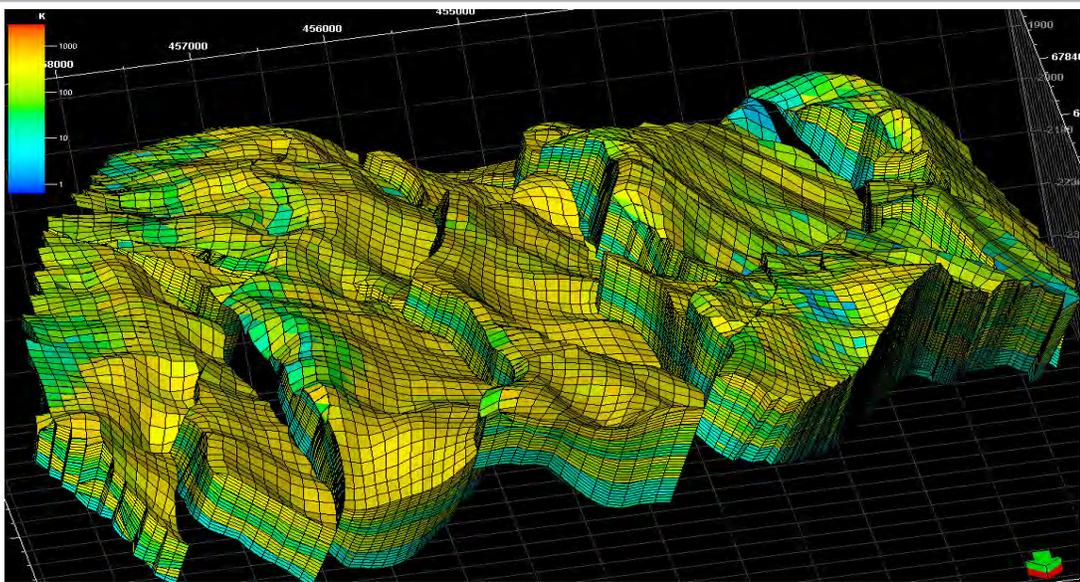


Figura 5.13. Modelado de permeabilidad mediante una simulación Gaussiana secuencial.

Finalmente se realiza el escalamiento para ambas propiedades de una malla fina a una malla gruesa.

Para el escalamiento de porosidad también se utiliza el promedio aritmético, ya que es una propiedad estática. La siguiente figura muestra el resultado.

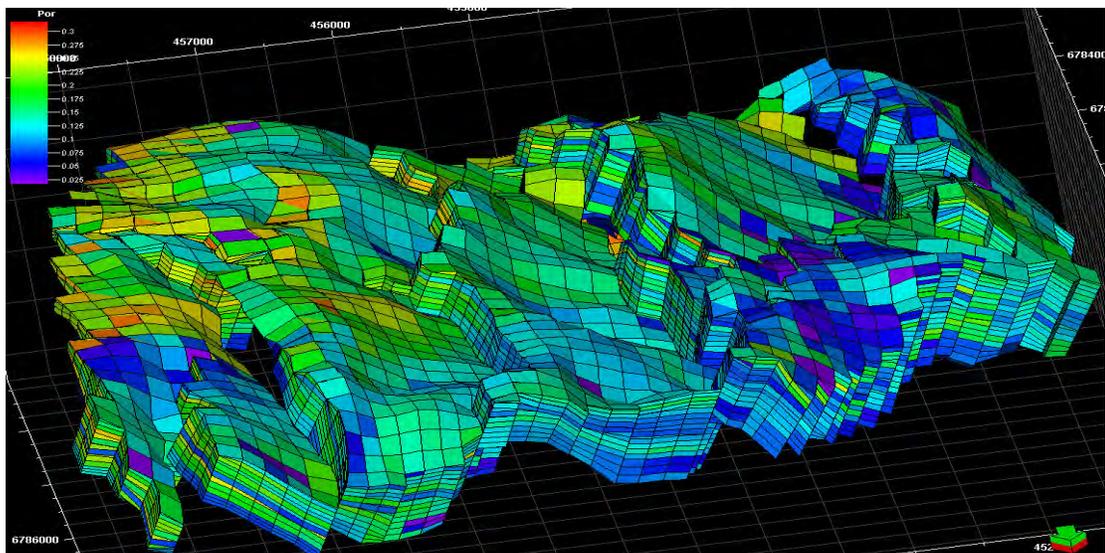


Figura 5.14. Escalamiento de porosidad con un promedio aritmético.

Para el escalamiento de permeabilidad se utiliza el promedio geométrico y un tensor de permeabilidad, ya que es una propiedad dinámica. Las siguientes figuras muestran las diferencias encontradas.

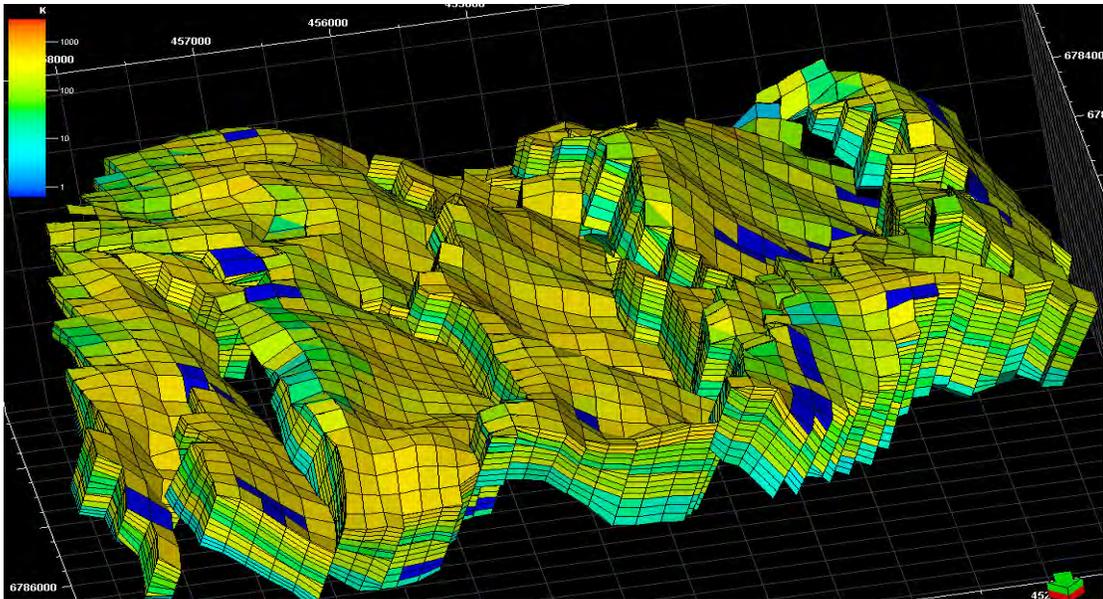


Figura 5.15. Escalamiento de permeabilidad con un promedio geométrico.

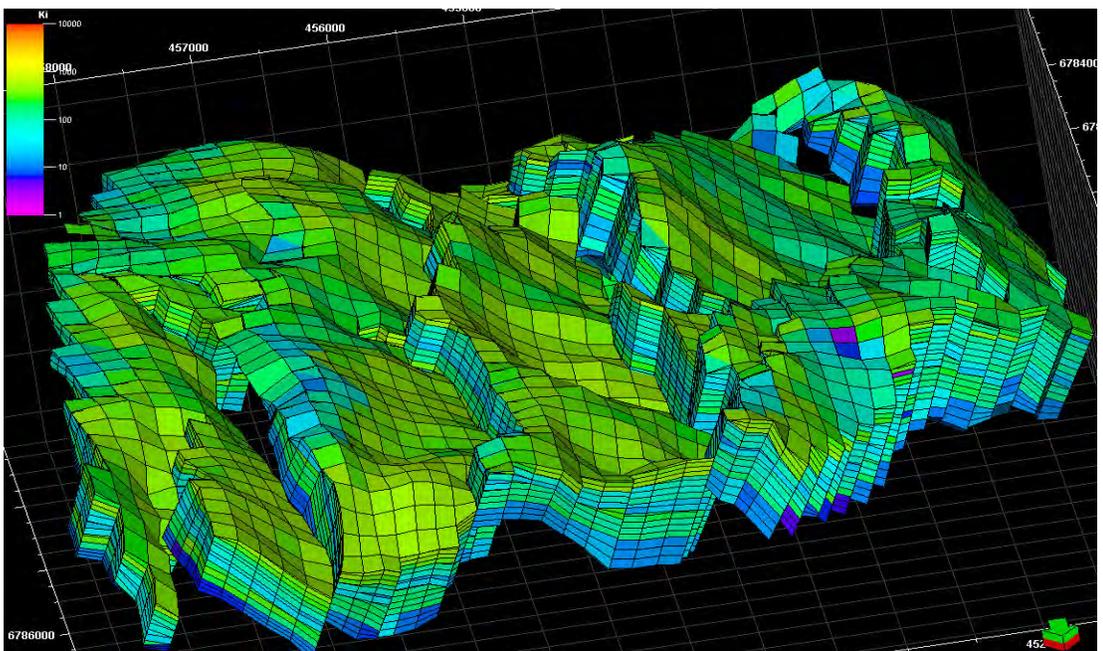


Figura 5.16a. Escalamiento de permeabilidad con un tensor de permeabilidad en la dirección i .

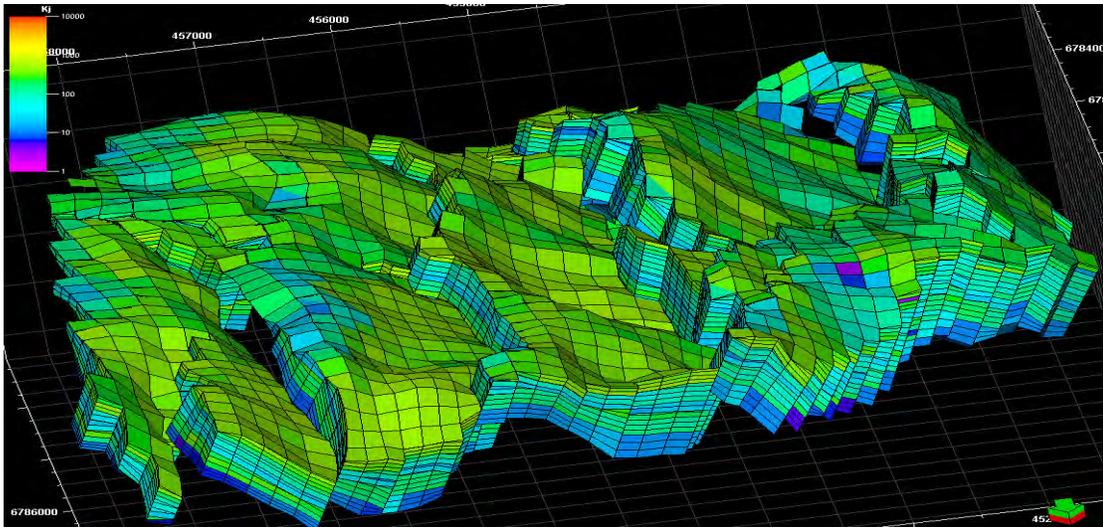


Figura 5.16b. Escalamiento de permeabilidad con un tensor de permeabilidad en la dirección j .

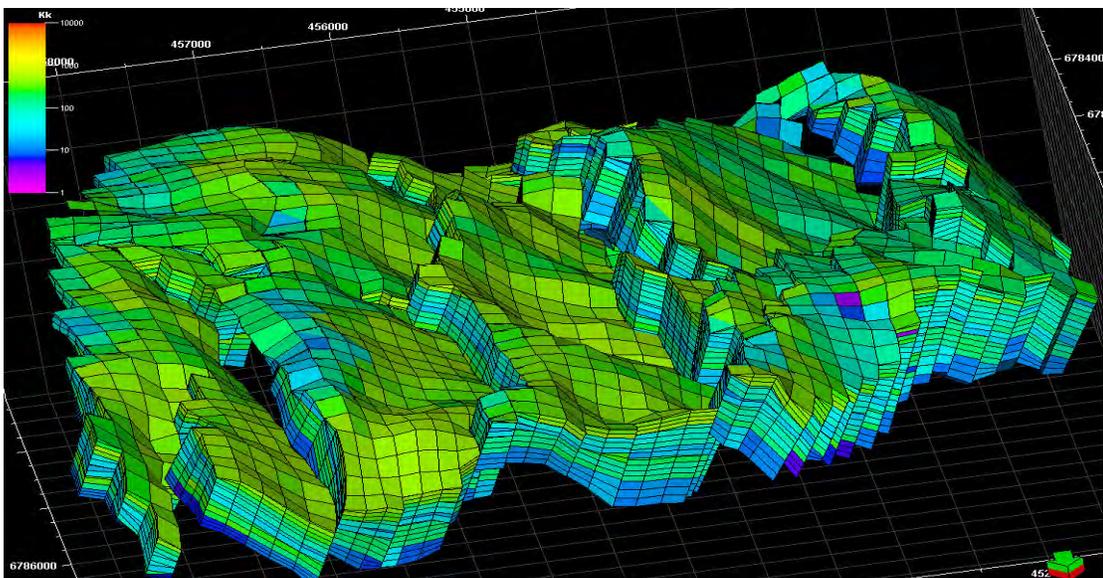


Figura 5.16c. Escalamiento de permeabilidad con un tensor de permeabilidad en la dirección k .

En lo que respecta al tensor de permeabilidad, no existen grandes diferencias entre las direcciones i,j,k . Sin embargo entre el tensor de permeabilidad y el promedio geométrico si se distinguen diferencias en el resultado obtenido respecto a los valores de permeabilidad en la malla.

CAPITULO 6

CONCLUSIONES Y RECOMENDACIONES

La Geoestadística puede proporcionar soluciones prometedoras en las áreas de ingeniería y geociencias, como la integración de información proveniente de ambas disciplinas y evitar por ejemplo, que los datos de entrada en un modelado sean los mismos de salida. Así mismo, las herramientas geoestadísticas son de gran utilidad para lograr una mejor descripción del yacimiento y minimizar incertidumbres en la interpretación final, lo cual permite prevenir costosos errores a lo largo de la vida de este.

La aplicación de métodos geoestadísticos permite solucionar el problema de información limitada y estimar o simular propiedades del yacimiento.

La geoestadística es una técnica estadística que permite obtener o contabilizar las relaciones espaciales existentes entre distintas variables al estimar sus valores en localidades donde no es posible obtener muestras (no observables).

Las relaciones espaciales incluyen información de cómo se da la dependencia entre los valores de una misma variable, o información de cómo es que dos variables diferentes tengan la posibilidad de común dependencia.

Es necesario asignar valores a cada celda de las mallas de simulación con propiedades petrofísicas (porosidad permeabilidad, saturación de agua), para lograr este objetivo es recomendable utilizar técnicas determinísticas y estocásticas de manera conjunta, sin olvidar que los resultados que se obtengan dependerán de las ventajas y desventajas de ambas así como la información disponible. A su vez los modelos de estimación y/o simulación son necesarios para obtener escenarios pesimistas y optimistas, y poder comparar posteriormente dichos modelos con el calculo de volúmenes por zona, facies y/o unidades estratigráficas, con la finalidad de determinar el modelo que mejor se ajuste a la realidad y a la necesidad del problema a resolver para integrar la información en un modelado de flujo.

La resolución de una malla será clave en la construcción del modelo. En una malla con alta resolución (muchas celdas) se presenta una gran complejidad en cuanto a espacio por lo que el tiempo de proceso entre celda y celda se incrementa. Una malla de menor resolución tendrá un objetivo menos complejo, y permitirá trabajar de manera más rápida.

La decisión que se tome dependerá del propósito del modelo, el detalle y la cantidad de datos disponibles. En ocasiones es recomendable comenzar con una malla de poca resolución, probando los efectos en los cambios y posteriormente incrementar la resolución conforme los parámetros se vuelven más precisos.

Con lo que refiere al modelado de yacimientos carbonatados, este debe realizarse con mucho más atención dedicación y detalle ya que es necesario recordar que las propiedades que se miden con registros geofísicos arrojan valores totales, por lo tanto deben de tomarse en cuenta tanto la porosidad primaria y secundaria para este tipo de yacimientos, es decir, es necesario emplear un tipo de medición especial, por ejemplo, los registros de imagen del subsuelo a lo largo de un pozo permiten conocer la orientación y tipo de fractura, que en combinación con el análisis de núcleos, puede complementar la información con la que se alimenta al modelo. Adicionalmente, la interpretación de facies no es suficiente en el modelado yacimientos carbonatados, por lo tanto, la misma, debe de ajustarse con estudios de interpretación estructural. El estudio geoestadístico de este tipo de yacimientos debe ser mucho más específico, ya que la relación que existe entre los datos puede ser poco significativa de acuerdo al resultado esperado.

APÉNDICE A

GLOSARIO

TÉCNICAS CUANTITATIVAS.

Medición de valores de variables que intervienen en un determinado proceso histórico, con las que se pueden hacer manipulaciones estadísticas, y siempre con el objetivo de medir relaciones o descubrir nuevas relaciones tanto dentro de una serie dada como en la comparación de varias de ellas¹.

VARIABLE.

Propiedad o rasgo de un hecho u objeto (no constante) por la que puede ser caracterizado o clasificado. Los caracteres de un elemento pueden ser de muy diversos tipos, por lo que es posible realizar una clasificación en dos grandes clases:

- a) En función de su naturaleza:
 - 1) Cualitativas: carácter no numérico
 - 2) Cuantitativas: carácter numérico
 - 3) Discretas: adoptan solo valores enteros
 - 4) Continuas: adoptan cualquier valor

- b) En función de su escala de medición:
 - 1) Nominales: categorías que no implican orden
 - 2) Dicotómicas: sólo dos categorías
 - 3) Ordinales: categorías que implican orden, pero sin una distancia medible entre ellas
 - 4) De intervalo: con orden y distancia entre sus elementos, y nunca con valor cero
 - 5) De proporción o de razón: con orden y distancia entre sus elementos, y posible valor cero

MEDICIÓN.

Es la acción de Medir, contar o comparar una unidad con otra, asignando una valoración numérica; esta característica es variable (y no constante), por lo que puede adoptar valores diversos, dos al menos.

INDICADOR.

Es una variable indirecta que permite lograr un acercamiento a la medición de un problema en el que se desconoce la evolución de la variable directa (de la producción en el antiguo régimen, por ejemplo, el diezmo), o en caso de que esta no pueda ser expresada en una medición única (del nivel de vida de un país, por ejemplo, la renta per cápita), o, simplemente, cuando la variable deseada no tenga expresión directa posible (del prestigio, por ejemplo, el modelo de automóvil).

DATOS.

Características o números recolectados por observación. Son los valores cualitativos o cuantitativos mediante los cuales se miden las características de los objetos, sucesos o fenómenos a estudiar.

MATRIZ DE DATOS.

Presentación organizada (tabulada) del conjunto de datos a analizar mediante técnicas estadísticas.

POBLACIÓN O UNIVERSO.

Es el total del conjunto de elementos, objetos u observaciones reales o potenciales, de los cuales se quiere obtener información. Desde el punto de vista estadístico, el término población tiene un significado mucho más amplio que el usual, ya que puede referirse a personas, cosas, actos, áreas geográficas e incluso al tiempo.

La población debe estar perfectamente definida en el tiempo y en el espacio, de modo que ante la presencia de un potencial integrante de la misma, se pueda decidir si forma parte o no de la población bajo estudio². Por lo tanto, al definir una población, se debe cuidar que el conjunto de elementos que la integran quede perfectamente delimitado. El tamaño de una población viene dado por la cantidad de elementos que la componen.

La población puede ser según su tamaño de dos tipos:

Población finita: cuando el número de elementos que la forman es finito, por ejemplo el número de alumnos de un centro de enseñanza, o grupo clase.

Población infinita: cuando el número de elementos que la forman es infinito, o tan grande que pudiesen considerarse infinitos.

La forma en que se distribuye la característica de interés en la población se denomina distribución poblacional, frecuentemente la distribución poblacional puede representarse a través de un modelo matemático, tal como la distribución normal (o Campana de Gauss), el modelo matemático que mejor represente a la distribución poblacional puede ser discreto o continuo y puede depender de los parámetros poblacionales. La distribución normal depende de los parámetros μ y σ^2 que en este caso representan respectivamente a la media y la varianza poblacionales.

ELEMENTO O UNIDAD DE ANÁLISIS.

En estadística, un elemento o unidad de análisis puede ser algo con existencia real, como un automóvil o una casa, o algo más abstracto como la temperatura o un intervalo de tiempo. Dada esta definición, puede redefinirse población como el conjunto de unidades de análisis.

CARACTERES.

Son series de características de cada elemento de la población que pueden ser objeto del estudio estadístico. Así por ejemplo si se considera como elemento a una persona, se puede distinguir en ella los siguientes caracteres: sexo, edad, nivel de estudios, profesión, peso, altura y color de pelo.

Por lo tanto, de cada elemento de la población es posible estudiar uno o más aspectos cualidades o caracteres.

FENÓMENO.

Un fenómeno es considerado como un sinónimo de un experimento y se define como toda aquella acción que se realiza con el fin de observar el resultado.

Por su naturaleza pueden clasificarse de varias formas:

Por la capacidad de predecir el resultado:

- **Determinísticos:** Es aquel fenómeno cuyo resultado se puede predecir con certeza.
- **Aleatorios:** Es aquel fenómeno en el cual no se puede predecir con certeza el resultado.

Por la capacidad de enumerar los resultados:

- **Continuos:** Se considera a un fenómeno como continuo cuando el número posible de observaciones o resultados no es un conjunto medible, es decir, los resultados pueden tomar cualquier valor.
- **Discretos:** Se considera un fenómeno como discreto cuando el número posible de observaciones o resultados es un conjunto, en el cual se

puede determinar con certeza su tamaño, es decir, los resultados toman valores establecidos.

Evento: Algo que ocurre al azar.

Acto: Algo que ocurre con la intervención humana

Hecho: Algo que ocurre al azar con la intención humana.

Evento Elemental: Un evento elemental o simple es aquel cuyo espacio muestral está compuesto por un sólo elemento.

Evento Compuesto: Es aquel que en dos o más cantidades compone un espacio muestral.

Evento Continuo: Es aquel cuyo resultado puede tomar cualquier valor.

Evento Discreto: Es aquel cuyo resultado sólo puede tomar determinados valores.

Eventos mutuamente excluyentes: Si se tienen dos eventos A y B , los eventos mutuamente excluyentes son aquellos en donde la ocurrencia de uno implica que el otro no puede ocurrir.

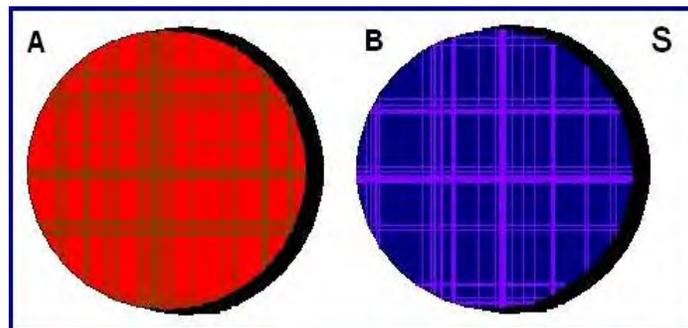


Figura A.1. Eventos mutuamente excluyentes: Diagrama de Venn.

Eventos colectivamente exhaustivos. Son aquellos cuya unión conforma totalmente el espacio muestral.

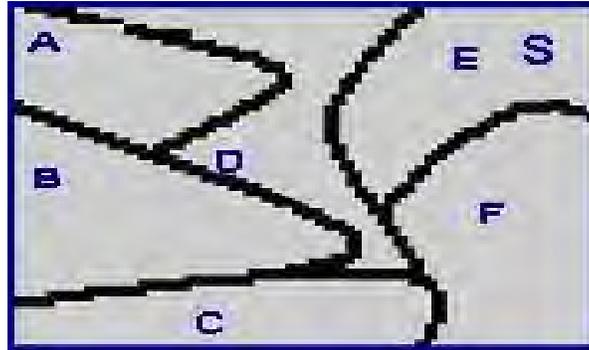


Figura A.2. Eventos colectivamente exhaustivos.

Donde: $A \cup B \cup C \dots Z = S$

APÉNDICE B

ESTADÍSTICA PARA

DISTRIBUCIONES

UNIDIMENSIONALES

B.1. MEDIDAS DE POSICIÓN¹.

Las medidas de posición son aquellos valores numéricos que permiten obtener o dar alguna medida de tendencia central, dividiendo el recorrido de la variable en dos, o bien fragmentar la cantidad de datos en partes iguales.

B.1.1. MEDIA.

La idea de media o promedio (también llamada media aritmética) formaliza el concepto intuitivo de punto de equilibrio de las observaciones. Es decir, es el punto medio del recorrido de la variable según la cantidad de valores obtenidos. Ese valor tiene varias propiedades importantes.

- a) Si se suma la distancia de todos los valores respecto de la media, esa suma da cero.
- b) Si se toma una cantidad de conjuntos de valores cualesquiera, cada uno con su respectiva media, la media del conjunto general es igual a la suma de cada una de las medias de los diferentes conjuntos.
- c) Es posible hallar la media de un conjunto de valores de una variable tomando la distancia de las observaciones a un valor cualquiera (pertenezca o no al recorrido de la variable).
- d) Si a un conjunto de observaciones de una variable se le realiza una operación matemática usando un valor constante, entonces la media del nuevo grupo de valores así obtenidos es igual a la aplicación de la misma operación matemática usando ese valor constante sobre la media original.

Cálculo para datos sin agrupar. Dado un conjunto de observaciones $x_1, x_2, \dots, x_{n-1}, x_n$ la media se representa mediante \bar{x} y se obtiene dividiendo la suma de todos los datos por el número de ellos, es decir:

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{i=1}^n x_i}{n} = \frac{1}{n} \sum_{i=1}^n x_i \dots\dots\dots(B.1)$$

Cálculo para datos agrupados: Considerando el caso en que se tiene una distribución de frecuencia con marcas de clase $C_1, C_2, \dots, C_{k-1}, C_k$ (en el caso de variables discretas no dispuestas en intervalos, se considerara que el valor de la variable es un intervalo de amplitud 1, en ese caso la marca de clase coincide con el valor de la variable) y frecuencias relativas $f_{R1}, f_{R2}, \dots, f_{R(k-1)}, f_{Rk}$ respectivamente. La ecuación B.1 puede interpretarse como la suma del producto de los datos por sus frecuencias relativas. Dado que en el caso de los datos agrupados (especialmente en el caso del uso de intervalos) no se cuenta con el valor concreto de los datos, se usa el punto medio del intervalo (que funcionaría como una especie de media del intervalo) y así la media se expresaría como la suma de los productos de las marcas de clase por sus frecuencias relativas:

$$\bar{x} = \sum_{i=1}^k C_i f_{Ri} \dots\dots\dots(B.2)$$

B.1.2. MEDIANA.

La mediana, a diferencia de la media no busca el valor central del recorrido de la variable según la cantidad de observaciones, sino que busca determinar el valor que tiene aquella observación que divide la cantidad de observaciones en dos mitades iguales. Por lo tanto es necesario atender a la ordenación de los datos, y debido a ello, este cálculo depende de la posición relativa de los valores obtenidos. Es necesario, antes que nada, ordenar los datos de menor a mayor o viceversa.

Hay que tener en cuenta que si $x_1, x_2, \dots, x_{n-1}, x_n$ se utiliza para denotar el conjunto de las observaciones, donde el subíndice indica el orden en el dato que fue obtenido o registrado, suele utilizarse $x_{(1)}, x_{(2)}, \dots, x_{(n-1)}, x_{(n)}$ para representar las mismas observaciones, pero ahora ordenadas de menor a mayor, por lo tanto aparecerá primero el dato más pequeño y al último el más grande.

Cálculo para datos sin agrupar es el valor del dato central si hay un número impar de observaciones o el valor de la semisuma de los valores de los dos datos centrales si se tiene un número par de datos. Así, la mediana puede expresarse como

$$\tilde{x} = x_{\left(\frac{n+1}{2}\right)} \dots\dots\dots(B.3a)$$

En caso de que n sea impar, y

$$\tilde{x} = \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2} \dots\dots\dots(B.3b)$$

En caso de que n sea par.

Cálculo para datos agrupados, en el caso de variables discretas donde cada categoría es el valor de la variable, se puede tomar como un caso de intervalo de amplitud 1 y en este caso el cálculo de la mediana funciona exactamente como lo visto para datos sin agrupar. Esto vale también para el resto de las medidas que se verán en esta sección (cuartiles, quintiles, etc.)

Cuando se emplean variables agrupadas por intervalos es imposible determinar con precisión los valores que toman los datos, ya que esa información se ha perdido en privilegio del agrupamiento por intervalos. Por lo tanto, en este caso, se debe buscar otro método para determinar el valor de la mediana.

El *cálculo a partir de la frecuencia relativa*, es sabiendo que la mediana divide al conjunto de las observaciones en dos grupos iguales, es claro que ella se ubicará en el intervalo donde las frecuencias relativas acumuladas alcanzan el valor 0.50.

El cálculo de la mediana se haría con la siguiente fórmula

$$\tilde{x} = \left(\frac{0.50 - F_{i-1}}{f_{Ri}} \right) a_i + l_{i-1} \dots \dots \dots (B.3c)$$

Es decir, la mediana sería el valor que resulta del siguiente conjunto de operaciones:

- a) Restarle a 0.50 el valor de la frecuencia relativa acumulada anterior al intervalo que contiene el valor 0.50
- b) Dividir ese resultado por la frecuencia relativa del intervalo que contiene el valor 0.50
- c) Multiplicar ese resultado por la amplitud del intervalo que contiene al valor 0.50 y
- d) Sumar a la cantidad así obtenida el valor que toma el límite inferior del intervalo que contiene el valor 0.50 de la frecuencia relativa.

Para el *cálculo a partir de la frecuencia absoluta*, la mediana se ubicará en el intervalo donde las frecuencias absolutas toman el valor $\frac{n}{2}$; que equivale a decir que la mediana de una distribución se ubicará en el intervalo en el cual la frecuencia absoluta acumulada alcanza el 50% de los datos.

Por lo tanto la fórmula para calcular la mediana sería

$$\tilde{x} = \left(\frac{\frac{n}{2} - n_{i-1}}{N_i} \right) a_i + l_{i-1} \dots \dots \dots (B.3d)$$

Es decir, la mediana sería el valor que resulta del siguiente conjunto de operaciones:

- a) Dividir el número de casos entre 2 y a ese resultado restarle la frecuencia absoluta acumulada hasta el intervalo anterior al que contiene al caso central
- b) dividir ese resultado por la frecuencia absoluta del intervalo que contiene al caso central
- c) multiplicar ese resultado por la amplitud del intervalo que contiene al dato central y
- d) sumar a la cantidad así obtenida el valor que toma el límite inferior del intervalo que contiene al caso central.

B.1.3. CUARTILES.

La mediana, como ya se mencionó, separa en dos mitades el conjunto ordenado de observaciones. Es posible aún dividir cada mitad en dos de tal manera que resulten cuatro partes iguales. Cada una de esas divisiones se conoce como Cuartil y se simbolizará mediante la letra Q agregando un subíndice según a cual de los cuatro cuartiles se hace referencia. Se llama primer cuartil Q_1 a la mediana de la mitad que contiene los datos más pequeños. Este cuartil, corresponde al menor valor que supera, o que deja por debajo de él, a la cuarta parte de los datos. Se llama tercer cuartil Q_3 a la mediana de la mitad formada por las observaciones más grandes. El tercer cuartil es el menor valor que supera, o que deja por debajo de él, a las tres cuartas partes de las observaciones. Con esta terminología, la mediana es el segundo cuartil Q_2 y el cuarto cuartil Q_4 coincide con el valor que toma el último dato, luego de ordenarlos. Dicho de otra manera, los cuartiles son valores que resultan de

dividir la población (el n de las observaciones) en cuatro partes iguales k (25% en c/u).

Cálculo para datos sin agrupar. El Q_k se obtiene identificando el valor que para la variable en cuestión tiene el individuo que ocupa la posición que corresponde al $k^*(25\%)$ de la población.

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$Q_k = \left(\frac{k(0.25) - F_{i-1}}{f_{Ri}} \right) a_i + l_{i-1} \dots \dots \dots (B.4a)$$

Por ejemplo:

$$Q_1 ; k=1; k^*(0.25)=0.25$$

$$Q_2 ; k=2; k^*(0.25)=0.5$$

$$Q_3 ; k=3; k^*(0.25)=0.75$$

$$Q_4 ; k=4; k^*(0.25)=1$$

Si el cálculo es a partir de la frecuencia absoluta.

$$Q_k = \left(\frac{k \frac{n}{4} - n_{i-1}}{N_i} \right) a_i + l_{i-1} \dots \dots \dots (B.4b)$$

B.1.4. QUINTILES.

Los quintiles son valores que resultan de dividir la población (el n de las observaciones) en cinco partes iguales g (20% en c/u).

Cálculo para datos sin agrupar. El Q^l_g se obtiene identificando el valor que para la variable en cuestión tiene el individuo que ocupa la posición que corresponde al $g^*(20\%)$ de la población.

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$Ql_g = \left(\frac{g(0.2) - F_{i-1}}{f_{Ri}} \right) a_i + l_{i-1} \dots \dots \dots (B.5a)$$

Si el cálculo es a partir de la frecuencia absoluta.

$$Ql_g = \left(\frac{g \frac{n}{5} - n_{i-1}}{N_i} \right) a_i + l_{i-1} \dots \dots \dots (B.5b)$$

B.1.5. DECILES.

Los deciles son valores que resultan de dividir la población (el *n* de las observaciones) en diez partes iguales *h* (10% en c/u).

Cálculo para datos sin agrupar. El *D_h* se obtiene identificando el valor que para la variable en cuestión tiene el individuo que ocupa la posición que corresponde al *h*(10 %)* de la población.

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$D_h = \left(\frac{h(0.1) - F_{i-1}}{f_{Ri}} \right) a_i + l_{i-1} \dots \dots \dots (B.6a)$$

Si el cálculo es a partir de la frecuencia absoluta.

$$D_h = \left(\frac{h \frac{n}{10} - n_{i-1}}{N_i} \right) a_i + l_{i-1} \dots \dots \dots (B.6b)$$

B.1.6. PERCENTILES.

Los percentiles son valores que resultan de dividir la población (el n de las observaciones) en cien partes iguales j (1% en cada una).

Cálculo para datos sin agrupar. El P_j se obtiene identificando el valor que para la variable en cuestión tiene el individuo que ocupa la posición $j*(1\%)$.

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$P_j = \left(\frac{h(0.01) - F_{i-1}}{f_{Ri}} \right) a_i + l_{i-1} \dots \dots \dots (B.7a)$$

Si el cálculo es a partir de la frecuencia absoluta.

$$P_j = \left(\frac{j \frac{n}{100} - n_{i-1}}{N_i} \right) a_i + l_{i-1} \dots \dots \dots (B.7b)$$

B.1.7. MODA.

La moda es aquel dato de la variable que más se repite; es decir, aquel valor de la variable (que puede no ser un único valor) con una frecuencia mayor.

Cálculo para datos sin agrupar. En los datos sin agrupar o en los datos agrupados para variables discretas donde cada clase es un valor diferente de la variable, basta una simple inspección ocular.

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$Moda = \left(\frac{f_{Ri} - f_{R(i-1)}}{(f_{Ri} - f_{R(i-1)}) + (f_{Ri} - f_{R(i+1)})} \right) a_i + l_{i-1} \dots \dots \dots (B.8a)$$

Si el cálculo es a partir de la frecuencia absoluta.

$$Moda = \left(\frac{N_i - N_{i-1}}{(N_i - N_{i-1}) + (N_i - N_{i+1})} \right) a_i + l_{i-1} \dots \dots \dots (B.8b)$$

B.2. MEDIDAS DE DISPERSIÓN¹.

Se llaman medidas de dispersión aquellas que permiten retratar la distancia de los valores de la variable a un cierto valor central, o que permiten identificar la concentración de los datos en un cierto sector del recorrido de la variable.

B.2.1. DESVIACIÓN ESTÁNDAR.

Es posible identificar conjuntos de datos que a pesar de ser muy distintos en términos de valores absolutos, poseen la misma media. Una medida diferencial para identificar esos conjuntos de datos es la concentración o dispersión alrededor de la media.

Cálculo para datos sin agrupar. Una manera natural para construir una medida de dispersión sería promediar las desviaciones de la media, pero:

$$x_1 - \bar{x} + x_2 - \bar{x} + \dots + x_{n-1} - \bar{x} + x_n - \bar{x} = 0$$

Para evitar que los distintos signos se compensen, se elevan las desviaciones de la media al cuadrado, de manera que todas las desviaciones sean positivas.

La raíz cuadrada del promedio de estas cantidades recibe el nombre de desviación estándar, o desviación típica y es representada por la siguiente fórmula

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}} \dots\dots\dots(B.9a)$$

La desviación estándar sólo puede utilizarse en el caso de que las observaciones se hayan medido con escalas de intervalos o razones.

A mayor valor del coeficiente del desvío estándar, mayor dispersión de los datos con respecto a su media. Es un valor que representa los promedios de todas las diferencias individuales de las observaciones respecto a un punto de referencia común, que es la media aritmética. Se entiende entonces que cuando este valor es más pequeño, las diferencias de los valores respecto a la media, es decir, los desvíos, son menores y, por lo tanto, el grupo de observaciones es más “homogéneo” que si el valor de la desviación estándar fuera más grande. O sea que a menor dispersión mayor homogeneidad y a mayor dispersión, menor homogeneidad.

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$s = \sqrt{\sum_{i=1}^k f_{Ri} (C_i - \bar{x})^2} \dots\dots\dots(B.9b)$$

Si el cálculo es a partir de la frecuencia absoluta.

$$s = \sqrt{\frac{\sum_{i=1}^k N_i (C_i - \bar{x})^2}{n}} \dots\dots\dots(B.9c)$$

B.2.2. VARIANZA.

El cuadrado de la desviación estándar recibe el nombre de varianza y se representa por s^2 . La suma de los cuadrados de los desvíos de la totalidad de las observaciones, respecto de la media aritmética de la distribución, es menor que la suma de los cuadrados de los desvíos respecto de cualquier otro valor que no sea la media aritmética.

La varianza no es más que el desvío estándar al cuadrado. Que es $s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$ (B.10)

B.2.3. COEFICIENTE DE VARIACIÓN.

Para comparar la dispersión de variables que aparecen en unidades diferentes (metros, kilos, etc.) o que corresponden a poblaciones extremadamente desiguales, es necesario disponer de una medida de variabilidad que no dependa de las unidades o del tamaño de los datos. Este coeficiente únicamente sirve para comparar las dispersiones de variables correspondientes a escalas de razón.

Una manera de construir una medida de variabilidad que cumpla los requisitos anteriores es el llamado coeficiente de variación muestral

$$CV = \frac{s}{|\bar{x}|} \dots\dots\dots(B.11)$$

Para muestras que presentan diferentes órdenes en la variación de la magnitud, típicamente es $CV > 1$, en ocasiones puede alojarse en rangos de 2 a 5. Un alto valor de CV indica que en las muestras existen valores extremos que

pueden afectar la estimación durante el muestreo. A menor CV se considerara que la distribución de la variable medida es más homogénea.

B.2.4. COEFICIENTE DE ASIMETRÍA.

Otro rasgo interesante en una distribución de frecuencias es si los datos aparecen ubicados simétricamente o no respecto de la media. Si es necesario cuantificar la simetría, es necesario conservar la información acerca tanto del signo como de la distancia de cada dato a la media, centro de simetría. Este razonamiento lleva a utilizar una potencia impar de las desviaciones.

En el caso en que el coeficiente valga cero, la distribución es simétrica alrededor de la media. Los valores positivos, indicarán distribuciones con mayor sesgo a la derecha y los valores negativos indicarán un mayor sesgo a la izquierda.

Cálculo para datos sin agrupar. El coeficiente de asimetría se define como

$$CA = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^3}{n s^3}} \dots\dots\dots(B.12a)$$

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$CA = \sqrt{\frac{\sum_{i=1}^k (C_i - \bar{x})^3 f_{Ri}}{s^3}} \dots\dots\dots(B.12b)$$

Si el cálculo es a partir de la frecuencia absoluta.

$$CA = \sqrt{\frac{\sum_{i=1}^k (C_i - \bar{x})^3 N_i}{ns^3}} \dots\dots\dots(B.12c)$$

B.2.5. COEFICIENTE DE APUNTAMIENTO.

Otra característica relevante en una distribución es su apuntamiento o kurtosis, que describe lo picuda o plana que puede ser la distribución, es decir si los datos se concentran demasiado o no, comparados con un modelo de distribución llamado distribución normal.

- La distribución se denominará **leptocúrtica** cuando el $CA_p > 0$ (más apuntada que la normal)
- La distribución se denominará **mesocúrtica** cuando el $CA_p = 0$ (normal)
- La distribución se denominará **platicúrtica** cuando $CA_p < 0$ (menos apuntada que la normal)

Cálculo para datos sin agrupar. El coeficiente de apuntamiento se define como

$$CA_p = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^4}{ns^4}} \dots\dots\dots(B.13a)$$

Cálculo para datos agrupados. Se calcula a través de la siguiente relación:

Si el cálculo es a partir de la frecuencia relativa

$$CA_p = \sqrt{\frac{\sum_{i=1}^k (C_i - \bar{x})^4 f_{Ri}}{s^4}} \dots\dots\dots(B.13b)$$

Si el cálculo es a partir de la frecuencia absoluta.

$$CAp = \sqrt{\frac{\sum_{i=1}^k (C_i - \bar{x})^4 N_i}{ns^4}} \dots\dots\dots(B.13c)$$

B.2.6. RANGO INTERCUARTÍLICO.

El rango intercuartílico *RI* es, sencillamente, la diferencia entre el tercer y el primer cuartil, es decir

$$RI = Q_3 - Q_1 \dots\dots\dots(B.14)$$

Esto representa en cuántas unidades de los valores que toma la variable, se concentra el cincuenta por ciento central de los casos.

NOMENCLATURA

CAPITULO 2.

a_1, b_1 = Mínimo y máximo de una distribución uniforme respectivamente.

A, B, C = Evento de un experimento aleatorio.

b = Es la ordenada al origen o intersección con el eje y .

c_1, d_1 = Constantes utilizadas para definir distribuciones bidimensionales.

c_i, c_j, c_k = Categorías de la variable (clase i, j, k).

$c(x, y)$ = Covarianza muestral para las variables x, y .

$c[\bar{L}]$ = Covarianza estimada para un intervalo \bar{L} .

$C[\bar{L}]$ = Covarianza para un intervalo \bar{L} .

Cv = Coeficiente de variación poblacional.

$C[X, Y]$ = Covarianza poblacional para dos variables aleatorias X e Y .

$C[X(\bar{u}_i), X(\bar{u}_j)]$ = Covarianza de las variables aleatorias regionalizadas $X(\bar{u}_i), X(\bar{u}_j)$.

E = Experimento aleatorio.

$E[X(\bar{u}_i)]$ = Valor esperado o media de la variable aleatoria regionalizada X en la localización \bar{u}_i .

$E[X]$ = Valor esperado o media de la variable aleatoria X .

f_i, f_j, f_k = Frecuencia de clase o absoluta i, j, k .

f_{Ri}, f_{Rj}, f_{Rk} = Frecuencia relativa i, j, k .

$f(x)$ = Función densidad de probabilidad.

$f[X(\bar{u}_i)]$ = Función de distribución de probabilidad de la variable aleatoria regionalizada X en la localización \bar{u}_i .

$f(x, y)$ = Función densidad de probabilidad para una distribución bidimensional de dos variables aleatorias continuas X, Y .

$f_x(x), f_y(y)$ = Distribuciones marginales de variables aleatorias conjuntas X, Y respectivamente.

$F[X(\bar{u}_i)]$ = Función de distribución acumulativa conjunta de la variable aleatoria regionalizada X en la localización \bar{u}_i .

F_i, F_j, F_k = Frecuencia relativa acumulativa i, j, k .

$F(x)$ = Función de distribución acumulativa de la variable aleatoria X .

$F(x, y)$ = Función de distribución acumulativa conjunta de las variables aleatorias X, Y .

$g(x)$ = Función de la variable x .

$g(X)$ = Función de la variable aleatoria X .

k = Permeabilidad.

K = Constante.

l_i, l_j, l_k = Límite superior de un intervalo i, j, k .

$l_{i-1}, l_{j-1}, l_{k-1}$ = Límite inferior de un intervalo i, j, k .

\bar{L} = Vector de translación o intervalo.

m = Pendiente de una línea recta.

NOMENCLATURA

n = Número total de muestras, n .

n_A = Número de veces que sucede un experimento, n .

n_C = Número de veces que se repite un experimento bajo condiciones controladas, n .

n_X = Número de posibles eventos de la variable aleatoria X , n .

n_Y = Número de posibles eventos de la variable aleatoria Y , n .

N_i, N_j, N_k = Frecuencia absoluta acumulativa i, j, k .

$Np(\bar{L})$ = Número de pares de datos a la distancia del vector \bar{L} .

o = Número de posibles eventos n .

$p(A)$ = Probabilidad de que el evento A ocurra.

$p(B)$ = Probabilidad de que el evento B ocurra.

$p(\Phi)$ = Probabilidad de que un evento imposible ocurra.

$p(A \text{ I } B)$ = Probabilidad de que el evento A ocurra con la condición de que previamente haya ocurrido el evento B .

$p(B \text{ I } A)$ = Probabilidad de que el evento B ocurra con la condición de que previamente haya ocurrido el evento A .

$p(x)$ = Función masa de probabilidad.

$p(x, y)$ = Función masa de probabilidad para dos variables.

P_X, P_Y = Distribuciones marginales para variables aleatorias discretas X, Y .

$r(x, y)$ = Coeficiente de correlación muestral.

$r(\bar{L})$ = Coeficiente de correlación para un intervalo \bar{L} .

R = Región donde se realiza un muestreo.

s = Desviación Estándar muestral.

s_x = Desviación Estándar de la variable x .

s_y = Desviación Estándar de la variable y .

s^2 = Varianza muestral.

S = Espacio muestral.

\bar{u} = Localización \bar{u} .

$v.a$ = Variable aleatoria.

$v.a.b$ = Variable aleatoria bidimensional

$v.a.b.c$ = Variable aleatoria bidimensional continua.

$v.a.b.d$ = Variable aleatoria bidimensional discreta.

$v.a.c$ = Variable aleatoria continua.

$v.a.d$ = Variable aleatoria discreta.

$v.a.r$ = Variable aleatoria regionalizada.

$V[X]$ = Varianza de la variable aleatoria X .

$V[X(\bar{u}_i)]$ = Varianza de la variable aleatoria regionalizada X en la localización \bar{u}_i .

x_i = Valor de la muestra i para la variable x .

\bar{x} = Media muestral o aritmética.

NOMENCLATURA

\bar{x} = Mediana muestral

$x(\bar{u}_i)$ = Valor muestral de la variable aleatoria regionalizada X en la localización \bar{u}_i .

$X(u)$ = Función aleatoria.

$X(\bar{u}_1), \dots, X(\bar{u}_n)$ = Variable aleatoria regionalizada X en la localización $\bar{u}_1, \dots, \bar{u}_n$.

$X^*(\bar{u}_i)$ = Valor estimado de una variable aleatoria regionalizada en un análisis geostadístico.

X, Y, Z = Variables aleatorias.

y_i = Valor de la muestra i para la variable y .

ε = Error estándar.

ϕ = Porosidad [%]

$\gamma[\bar{L}]$ = Variograma para un intervalo \bar{L} .

$\hat{\gamma}[\bar{L}]$ = Variograma estimado para un intervalo \bar{L} .

$\rho[\bar{L}]$ = Coeficiente de correlación para un intervalo \bar{L} .

$\rho[X, Y]$ = Coeficiente de correlación poblacional.

μ = Media poblacional.

μ_x = Media poblacional de la variable aleatoria X .

μ_y = Media poblacional de la variable aleatoria Y .

σ = Desviación Estándar poblacional.

σ_x = Desviación Estándar poblacional de la variable aleatoria X .

σ_y = Desviación Estándar poblacional de la variable aleatoria Y .

σ^2 = Varianza poblacional.

CAPITULO 3.

a = Alcance (intervalo \bar{L}).

a' = Alcance efectivo o alcance práctico (intervalo \bar{L}).

C_o = Efecto pepita (distancia $\mathcal{F}[\bar{L}]$).

C_T = Meseta (distancia $\mathcal{F}[\bar{L}]$).

$\hat{F}[\bar{L}]$ = Función variograma relativo.

$i(\bar{u}, x_L)$ = Valor del indicador en la localización \bar{u} , para un límite x_L .

$I(\bar{u}, x_L)$ = Variable indicadora en la localización \bar{u} , para un límite x_L .

L_{\max} = Distancia hasta la cual será calculado el variograma.

\bar{L} = Vector de translación o intervalo.

$d\bar{L}$ = Tolerancia lineal \bar{L} .

n = Número total de muestras, n .

$Np(\bar{L})$ = Número de pares de datos a la distancia del vector \bar{L} .

p = Valor de la potencia en una transformación potencial o en potencia.

R_i = Rango del i ésimo valor i en una secuencia de n muestras.

NOMENCLATURA

$x(\bar{u}_i)$ = Valor muestral de la variable aleatoria regionalizada X en la localización \bar{u}_i .

$x_T(\bar{u}_i)$ = Valor muestral transformado en la localización \bar{u}_i .

$X(\bar{u}_i + \bar{L})$ = Variable aleatoria regionalizada en la localización $\bar{u}_i + \bar{L}$.

α, β = Ángulos [°]

$d\alpha, d\beta$ = Tolerancias angulares [°]

$\hat{\gamma}[\bar{L}]$ = Variograma estimado para un intervalo \bar{L} .

CAPITULO 4.

$|c|$ = Vector covarianza.

$|C|$ = Matriz covarianza.

$C(\bar{u}_i, \bar{u}_j)$ = Covarianza entre los puntos de localización \bar{u}_i, \bar{u}_j , respectivamente.

$C(\bar{u}_i, \bar{u}_o)$ = Covarianza entre la muestra en la localización \bar{u}_i , y la muestra que no puede obtenerse en la localización \bar{u}_o .

$e(\bar{u}_i)$ = Error de estimación en la localización \bar{u}_i .

$E[]$ = Valor esperado del término entre corchetes.

$F(x)$ = Función de distribución acumulativa de la variable aleatoria X .

$m(\bar{u})$ = Media local dentro de una vecindad determinada.

n = Número total de muestras, n .

$V[] =$ Varianza del término entre corchetes.

$x(\bar{u}_i) =$ Valor muestral de la variable aleatoria regionalizada X en la localización \bar{u}_i .

$x^*(\bar{u}_i) =$ Valor estimado muestral de la variable aleatoria regionalizada X en la localización \bar{u}_i .

$X(\bar{u}_i) =$ Variable aleatoria regionalizada X en la localización \bar{u}_i .

$X_G(\bar{u}_i) =$ Variable aleatoria continua regionalizada X en la localización \bar{u}_i .

$K_G(\bar{u}_i) =$ Variable aleatoria discreta regionalizada K en la localización \bar{u}_i .

$X^*(\bar{u}_o) =$ Valor estimado de la variable aleatoria regionalizada X en la localidad donde no es posible obtener muestras \bar{u}_o .

$X_G^*(\bar{u}_o) =$ Valor estimado de la variable aleatoria continua regionalizada X en la localidad donde no es posible obtener muestras \bar{u}_o , mediante una transformación Gaussiana.

$K_G^*(\bar{u}_o) =$ Valor estimado de la variable aleatoria discreta regionalizada K en la localidad donde no es posible obtener muestras \bar{u}_o , mediante una transformación Gaussiana.

$Y(\bar{u}) =$ Variable aleatoria regionalizada Y en la localización \bar{u}_i .

$\lambda_i =$ Peso o importancia asignado (a) a el valor de una variable en la localidad (\bar{u}_i).

$\mu =$ Multiplicador de Lagrange.

$\sigma^2_E =$ Varianza de error.

$|\Lambda| =$ Vector de pesos.

CAPITULO 5.

A = Área.

C = Valores definidos en las ecuaciones 5.26 y 5.27.

f = Flujo fraccional.

g = Aceleración gravitacional.

h = Espesor del estrato.

H = Espesor total.

j = Variable falsa (apoyo).

k = Permeabilidad.

L = Longitud o distancia.

m, n = Número de bloques de malla.

M = Razón de movilidad.

N_c = Número de capilaridad.

N_g = Número de gravedad.

p = Presión.

p_c = Presión capilar.

q, Q = Gasto de flujo.

R_{jl} = Razón de velocidad entre los estratos j y l .

R_L = Número adimensional en la ecuación 5.48.

S = Saturación.

t = Tiempo.

V = Volumen.

V_p = Volumen de poro.

V_w = Volumen de agua.

x = Distancia del frente de inyección.

x, y, z = Ejes de coordenadas.

Y = Dirección de flujo en la ecuación 5.70.

Δ = Incremento en la distancia o la presión.

ϕ = Porosidad.

∂ = Derivada parcial.

γ = Gravedad específica.

μ = Viscosidad.

π = Constante 3.1416.

ρ = Densidad.

Σ = Sumatoria.

Π = Multiplicatoria.

NOMENCLATURA

Subíndices.

a = Promedio aritmético.

app = Aparente.

D = Adimensional.

g = Promedio geométrico.

h = Promedio armónico. Horizontal.

i = Bloque de Malla, i .

j = Variable falsa (apoyo) representando un estrato.

l = Estrato en el que se presenta el rompimiento.

o = Aceite.

or = Aceite residual.

r = Relativa (o).

T, t = Total.

v = Vertical.

w = Agua.

x, y, z = Ejes de coordenadas.

xx, yy = Direcciones principales.

xy, yx = Elementos en posición diagonal en el tensor de permeabilidad.

Superíndices.

$\bar{\sim}$ = Promedio.

p = Potencia.

APÉNDICE B.

a_i, a_j, a_k = Amplitud de un intervalo i, j, k .

c_i, c_j, c_k = Categorías de la variable (clase i, j, k).

C_i, C_j, C_k = Marcas de clase i, j, k .

CA = Coeficiente de asimetría.

CAp = Coeficiente de apuntamiento.

CV = Coeficiente de variación muestral.

D_h = Decil h ; " h " representa la décima parte del porcentaje de las muestras, n .

f_i, f_j, f_k = Frecuencia de clase o absoluta i, j, k .

f_{Ri}, f_{Rj}, f_{Rk} = Frecuencia relativa i, j, k .

F_i, F_j, F_k = Frecuencia relativa acumulativa i, j, k .

l_i, l_j, l_k = Límite superior de un intervalo i, j, k .

$l_{i-1}, l_{j-1}, l_{k-1}$ = Límite inferior de un intervalo i, j, k .

n = Número total de muestras, n .

N_i, N_j, N_k = Frecuencia absoluta acumulativa i, j, k .

P_j = Percentil j ; " j " representa la centésima parte del porcentaje de las muestras, n .

Q_k = Cuartil k ; " k " representa la cuarta parte del porcentaje de las muestras, n .

NOMENCLATURA

Ql_g = Quintil g ; “ g ” representa la quinta parte del porcentaje de las muestras, n .

RI = Rango intercuartílico.

s = Desviación Estándar muestral.

s^2 = Varianza muestral.

x_i = Valor de la muestra i para la variable x .

\bar{x} = Media muestral o aritmética.

\tilde{x} = Mediana muestral.

RELACIÓN DE FIGURAS Y TABLAS

FIGURAS

CAPITULO 1.

Figura 1.1. Proceso típico de la simulación de yacimientos.

Figura 1.2. Escalas de Heterogeneidad.

Figura 1.3. Comparación entre los datos de una prueba de pozo y núcleos.

Figura 1.4. Representación de una trampa de aceite.

Figura 1.5. Efecto de las heterogeneidades en la eficiencia de barrido.

Figura 1.6. Efecto de las heterogeneidades gigascópicas en el comportamiento de un yacimiento

Figura 1.7. Modelos de Yacimientos.

Figura 1.8. Descripción Integral de Yacimientos.

Figura 1.9. (a) Modelado convencional vs (b) Modelado en Paralelo.

Figura 1.10. Predicción de un comportamiento futuro.

CAPITULO 2.

Figura 2.1. Ejemplo de una variable cuantitativa, registro de porosidad.

Figura 2.2. Histograma con intervalos iguales, para los valores de porosidad de la figura 2.1.

Figura 2.3. Polígono de frecuencias para los valores de porosidad de la figura 2.1.

Figura 2.4. Ejemplo de una ojiva para los valores de porosidad de la figura 2.1.

Figura 2.5. Distribución de dos variables.

Figura 2.6. Regresión lineal de porosidad vs permeabilidad de la figura 2.5.

Figura 2.7. Ejemplificación de la operación $A \cap B$.

Figura 2.8. Diagrama de árbol.

Figura 2.9. Regla de correspondencia: proyección de los resultados de un evento o experimento en la recta de los números reales.

Figura 2.10. El proceso integral de creación de una variable aleatoria.

Figura 2.11. Función masa de probabilidad.

Figura 2.12. Función densidad de probabilidad: Probabilidad de que un valor se aloje dentro de un intervalo (a_1, b_1) .

Figura 2.13. Carácter mixto de una función aleatoria.

Figura 2.14. Fenómeno natural caracterizado por la distribución espacial de una variable aleatoria regionalizada. (a) Concepto tradicional. (b) Concepto geoestadístico.

CAPITULO 3.

Figura 3.1. Sección transversal que muestra la correlación detallada con un despliegue de registros para tres pozos.

Figura 3.2. Transformación de datos originales en una distribución normal.

Figura 3.3. Direcciones para el cálculo de los variogramas. (a) Comenzando en 0° . (b) Comenzando en 45° .

Figura 3.4. Mapa de variogramas. Las flechas muestran las direcciones de anisotropía mayor (roja) y menor (azul).

Figura 3.5. Espacio definido por la tolerancia lineal.

Figura 3.6. Espacio definido por la tolerancia lineal y angular.

Figura 3.7. Espacio definido por las tolerancias y el ancho de banda.

Figura 3.8. Forma típica del variograma.

Figura 3.9. Comportamiento del variograma en el origen, (a) Parabólico, (b) Lineal, (c) Efecto Pepita, (d) Discontinuo puro.

Figura 3.10. Parámetros del variograma.

Figura 3.11. Modelo Efecto Pepita.

Figura 3.12. Modelo Esférico.

Figura 3.13. Modelo Exponencial.

Figura 3.14. Modelo Gaussiano.

Figura 3.15. Modelo con función Potencia.

Figura 3.16. Modelo Lineal.

Figura 3.17. (a) Anisotropía Geométrica. (b) Anisotropía Zonal.

Figura 3.18. Efecto Proporcional.

Figura 3.19. Datos de impedancia acústica distribuidos en 3D.

Figura 3.20. Datos de impedancia acústica: coordenadas x,y,z.

Figura 3.21a. Mapas de Variograma.

Figura 3.21b. Mapas de Variograma.

Figura 3.21c. Mapas de Variograma.

Figura 3.21d. Mapas de Variograma.

Figura 3.21e. Mapa de Variograma.

Figura 3.22. Determinación de zonas de anisotropía.

Figura 3.23. Variogramas experimentales de los tres mapas de variograma con mayor definición en cuanto a visualización.

Figura 3.24. Variograma experimental para una distancia máxima x,y de 30,000 m y un intervalo x,y de 10 m.

Figura 3.25. Variograma experimental para una distancia máxima x,y de 26,000 m y un intervalo x,y de 10 m.

Figura 3.26. Variograma experimental ajustado a un modelo esférico para una distancia máxima x,y de 26,000 m y un intervalo x,y de 20 m.

CAPITULO 4.

Figura 4.1. Determinación de tipos de vecindad.

Figura 4.2. Valor real vs valor estimado durante el proceso de validación cruzada.

Figura 4.3. Error (real-estimado) vs valor real durante el proceso de validación cruzada.

Figura 4.4. Transformación Gaussiana de una variable continua.

Figura 4.5. Transformación Gaussiana de una variable discreta.

Figura 4.6. Selección de muestras.

Figura 4.7. Selección de una muestra en la transformación Gaussiana.

Figura 4.8. Retransformación de un espacio Gaussiano para una variable continua.

Figura 4.9. Retransformación de un espacio Gaussiano para una variable discreta.

Figura 4.10. Intervalos geológicos isocronales.

Figura 4.11. Atributos interrelacionados en bloques de malla individuales.

Figura 4.12. Visualización en 2D de impedancia acústica: estimación por Kriging simple.

Figura 4.13. Visualización en 2D de impedancia acústica: simulación Gaussiana secuencial.

Figura 4.14. Comparación de histogramas: datos originales de impedancia acústica vs datos obtenidos mediante Kriging simple.

Figura 4.15. Comparación de histogramas: datos originales de impedancia acústica vs datos obtenidos mediante simulación Gaussiana secuencial.

Figura 4.16a. Visualizaciones equiprobables de simulación Gaussiana. Izquierda-número aleatorio= 567, derecha-número aleatorio= 1009.

Figura 4.16b. Visualizaciones equiprobables de simulación Gaussiana. Izquierda-número aleatorio= 12337, derecha-número aleatorio= 1789.

Figura 4.16c. Visualizaciones equiprobables de simulación Gaussiana. Izquierda-número aleatorio= 478, derecha-número aleatorio= 9520.

CAPITULO 5.

Figura 5.1. Bloques de malla a escalas gruesa y fina.

Figura 5.2. Refinamiento de un bloque a escala gruesa alrededor de un pozo.

Figura 5.3. Permeabilidades para casos ideales donde las capas están ordenadas de manera paralela y en serie, con respecto a la dirección de flujo.

Figura 5.4. Fronteras superior e inferior para el método del estrato incompleto.

Figura 5.6. Representación de un yacimiento en estratos con n capas.

Figura 5.5. Arreglo de permeabilidades para los métodos de renormalización y tensor.

Figura 5.6. Representación de un yacimiento en estratos con n capas.

Figura 5.7a. Conjunto de registros de porosidad y permeabilidad originales.

Figura 5.7b. Conjunto de registros de porosidad y permeabilidad originales.

Figura 5.7c. Conjunto de registros de porosidad y permeabilidad originales.

Figura 5.7d. Conjunto de registros de porosidad y permeabilidad originales.

Figura 5.8. Escalamiento de registros de porosidad.

Figura 5.9. Escalamiento de registros de permeabilidad.

Figura 5.10. Porosidad: Ajuste de variogramas para la obtención de parámetros óptimos.

Figura 5.11. Modelado de porosidad mediante una simulación Gaussiana secuencial.

Figura 5.12. Permeabilidad: Ajuste de variogramas para la obtención de parámetros óptimos.

Figura 5.13. Modelado de permeabilidad mediante una simulación Gaussiana secuencial.

Figura 5.14. Escalamiento de porosidad con un promedio aritmético.

Figura 5.15. Escalamiento de permeabilidad con un promedio geométrico.

Figura 5.16a. Escalamiento de permeabilidad con un tensor de permeabilidad en la dirección i.

Figura 5.16b. Escalamiento de permeabilidad con un tensor de permeabilidad en la dirección j.

Figura 5.16b. Escalamiento de permeabilidad con un tensor de permeabilidad en la dirección k.

APÉNDICE A.

Figura A.1. Eventos mutuamente excluyentes: Diagrama de Venn.

Figura A.2. Eventos colectivamente exhaustivos.

TABLAS

Tabla 1.1. Escalas de Heterogeneidad de Yacimientos.

Tabla 2.1. Frecuencias para variables cuantitativas discretas.

Tabla 2.2. Frecuencias para variables cuantitativas continuas.

Tabla 2.3. Medidas de posición para los datos del registro de porosidad de la figura 2.1.

Tabla 2.4. Medidas de dispersión para los datos del registro de porosidad de la figura 2.1.

Tabla 3.1a. Registros geofísicos básicos utilizados para correlación.

Tabla 3.1b. Registros geofísicos básicos utilizados para correlación (continuación)

REFERENCIAS BIBLIOGRÁFICAS

CAPITULO 1.

- 1) "Trends in Reservoir Management", Oilfield Review (January 1992).
- 2) Ahmed, U., Crary, S.F., and Coates, G.R.: "Permeability Estimation: The Various Sources and their Interrelationships", JPT (May 1991) 578; Trans., AIME.
- 3) Mattax, C.C. and Dalton, R.L.: "Reservoir Simulation", JPT (June 1990).
- 4) Webber, K.J. and van Geuns, L.C.: "Framework for Constructing Clastic Reservoir Simulation Models", JPT (October 1990) 1248; Trans., AIME.
- 5) "Reservoir Characterization Using Expert Knowledge and Statistics", www.oilfield.slb.com/media/services/resources/oilfieldreview.
- 6) Hazlett, W.G. et al.: "The integrated Approach to Performing Reservoir Simulation Studies", paper SPE 26224.

CAPITULO 2.

- 1) Weimer, Richard C.: "Estadística", Edit. Trillas, 2ª edición, 1991.
- 2) Kreyszig, Edwin.: "Introducción a la Estadística Matemática Principios y Metodos", Edit. Limusa, 1994.
- 3) Estadística.www.hrc.es.
- 4) Bonilla, Gildaberto.: "Estadística II: Métodos prácticos de Inferencia Estadística".
- 5) Borrás, et. al.: "Apuntes de Probabilidad y Estadística", Facultad de Ingeniería UNAM, México 1985.
- 6) "Conceptos de Estadística y Probabilidad", <http://dcb.fi-c.unam.mx>.
- 7) Canavos.: "Probabilidad y Estadística", Edit. Mc Graw Hill, México 1988.
- 8) Villarreal.: "Probabilidad y Modelos Probabilísticas", UAEM, México 1989.
- 9) Journel, A. G. and Huijbregts, C. J.: "Mining Geostatistics", Academic Press, New York, 1978.
- 10) Chauvet, P. : "Aide-Memoire de Géostatistique Minière", École des Mines de Paris, 1994.

- 11) Zhang, X.F., Van Eijkeren, J.C.H. and Heemink, A.W.: "On the Weighted Least Squares", 1995.
- 12) Matheron, G.: "The Theory of Regionalized Variables and its Applications", Paris School of Mines, Cah. Cent. Morphologic Math., Fontainebleau, France (1965).
- 13) Armstrong, M. et Carignan, J. : "Géostatistique Linéaire, Application au Domaine Minier.", École de Mines de Paris, 1997.
- 14) Abasov, M.T., Djafarov, I.S. and Askerov, G.I.: "Computer Based System for Exploration, Optimization, and Reserve Estimation at the Bakhar Field", South Apsheron, Azerbaijan, SSR, Computers & Geosciences, Vol. 16, No. 2, 1990.
- 15) Carr, J.R.: "Numerical Analysis for the Geological Sciences", Prentice Hall Inc, 1995.

CAPITULO 3.

- 1) Villamar, M: Apuntes del curso "Caracterización de Yacimientos", FI-UNAM, México, 2004.
- 2) Definer, P., Peyret, O., and Serra, O.: "Automatic Determination of Lithology from Well Logs," paper 13290 presented at the 1980 SPE Annual Technical Conference and Exhibition.
- 3) Raymer, L.L and Burgess, K.A.: "The Role of Well Logs in Reservoir Modelling", SPE 9342 presented at the 1980 SPE Annual Technical Conference and Exhibition.
- 4) Serra, O.: "Sedimentary Environments from Wireline Logs", Schlumberger. 1985.
- 5) Log Interpretation Charts, Schlumberger Well Services, Houston, 1986.
- 6) Serra, O., and Abbot, H.T.: "The contribution of Logging Data to Sedimentology and Stratigraphy", paper SPE 9270 presented at the 1982 SPE Annual Technical Conference and Exhibition.
- 7) Isaaks, E.: "Risk Qualified Mapping for Hazardous Waste Sites: A Case Study in Distribution-Free Geoestistics", MS thesis, Stanford U., Stanford, California (1984).
- 8) Albert, F.: "Stochastic Imaging of Spatial Distributions Using Hard and Soft Information", MS thesis, Stanford U., Stanford, California (1984).

REFERENCIAS BIBLIOGRÁFICAS

9) Armstrong, M. et Carignan, J. : “Géostatistique Linéaire, Application au Domaine Minier.”, École de Mines de Paris, 1997.

10) “*Geoestatistical Models and Methods*”,
http://zappa.nku.edu/~longa/geomed/modules/geostats_lite/lec.html#Variography

11) Lamorey, G. and Jacobsom, E.: “Estimation of Semivariogram Parameters and Evaluation of the Effects of Data Sparsity”, *Mathematical Geology*, Vol.27, No.3, 1995.

12) Sahin, A. and Hassan, H.M.: “Enhancement of Permeability Variograms Using Outcrop Data”, *The Arabian Journal for Sciences and Engineering*, Vol. 23, No. 1C, 1998.

13) Genton, M.G.: “Highly Robust Variograms Estimation”, *Mathematical Geology*, Vol. 30, No. 2, 1998a.

14) Frykman, P. and Rogon, T.A.: “Anisotropy in Pore Networks Analysed with 2-D Autocorrelation (Variomaps)”, *Computers & Geosciences*, Vol. 19, No. 7, 1993.

15) Homand-Etienne, F., Belen, T., Sabbadini, S., Shtuka, A. and Royer, J.J.: “Analysis of the evolution of Rock Joints Morphology with 2-D autocorrelation (Variomaps)”, *Application of Statistics and Probability*, Lemaire, Favre and Mébarki (eds.), Balkema Rotterdam, 1995.

16) Isaaks & Co.: “A Spatial and Geostatistical Environment for Variography”, San Mateo, California, <http://www.isaaks.com>, 1999, SAGE2001.

17) Deutsch, C.V. and Journel, A.G.: “GSLIB: Geostatistical Software Library and User’s Guide”, Second Edition, Oxford University Press, 1998.

18) Myers, D.E.: “Kriging, Cokriging, Radial Basic Functions and the Role of Positive Definiteness”, *Computers Mathematical Application*, Vol. 24, No.12.

19) Genton, M.G.: “Variograms Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure”, *Mathematica Geology*, Vol. 30., No. 4, 1998b.

20) Krajewski, S.A. and Gibbs, B.L. : “A Variaogram Primer”, Gibbs Associates, 1993.

21) Carr, J.R.: “Numerical Analysis for the Geological Sciences”, Prentice Hall Inc, 1995.

22) Journel, A. G. and Huijbregts, C. J.: “Mining Geostatistics”, Academic Press, New York, 1978.

CAPITULO 4.

- 1) Matheron, G.: "The Theory of Regionalized Variables and its Applications," Paris School of Mines, Cah. Cent. Morphologic Math., Fontainebleau, France (1965).
- 2) Deutsch, C.V. and Journel, A.G.: "GSLIB: Geostatistical Software Library and User's Guide", Second Edition, Oxford University Press, 1998.
- 3) Davis, B.B.: "Uses and Abuses of Cross-Validation in Geoestatísticas", Math Geology (1987) 19, N°3.
- 4) Kelkar, M., Perez, G., and Chopra, A.: "Applied Geostatistics for Reservoir Characterization", Second Edition, Society of Petroleum Engineers, 2002.
- 5) Myers, D.E.: "Matrix Formation of Co-Kriging", Math Geology (1982), 18.
- 6) Carr, J., Myers, D.E., and Glass, C.: "Cokriging: A Computing Program", Computers and Geosciences (1985), 11.

CAPITULO 5.

- 1) Amyx, J. W., Bass, D.M.Jr., and Whiting, R.L.: "Petroleum Reservoir Engineering Physical Properties", McGraw-Hill Book Co. Inc., New York City (1960).
- 2) Deutsch, C.: "Calculating Effective Absolute Permeability in Sandstone/Shale Sequences", SPEFE (September 1989).
- 3) Cardwell, W. T. Jr and Parsons, R.L.: "Average Permeabilities of Heterogeneous Oil Sands", Trans., AIME (March 1995).
- 4) Abbaszadeh, M. and Koide, N.: "Evaluation of Permeability Upscaling Techniques and a New Algorithm for Interblock Transmissibilities", paper SPE 36179 presented at the 1996 Abu Dhabi Intl. Petroleum Exhibition and Conference, Abu Dhabi, UAE, October.
- 5) King, P.R.: "Renormalization Calculations of Immiscible Flow", Transport in Porous Media (1993).
- 6) Kasap, E. and Lake, L.W.: "Calculating the Effective Permeability Tensor for a Gridblock", SPEFE (June 1990).

REFERENCIAS BIBLIOGRÁFICAS

- 7) Pickup, G.E. et al.: "Geology, Geometry and Effective Flow", Petroleum Geoscience (1995).
- 8) Aasum, Y., Kasap, E. and Kelkar, M.: "Analytical Up-Scaling of Small-Scale Permeability Using a Full Tensor", Petroleum Geoscience (1995).
- 9) Lee, J., Kasap, E., and Kelkar, M.G.: "Analytical Up-Scaling of Permeability for 3D Gridblocks", SPEJ (March 1996), 59.
- 10) Beggs, S.H. and King, P.R.: "Modeling the Effects of Shales on Reservoir Performance: Calculation of Effective Vertical Permeability" paper SPE 13529, presented at the 1985 SPE Symposium on Reservoir Simulation, Dallas, 10-13 February.
- 11) Chu, L., Schatzinger, R.A. and Tham, M.K.: "Application of Wavelet Analysis to Upscaling of Rock Properties", paper SPE 36517 presented at the 1996 SPE Annual Technical Conference and Exhibition, Denver, 6-9 October.
- 12) Zapata, V.J. and Lake, L.W.: "A Theoretical Analysis of Viscous Crossflow", paper SPE 10111 presented at the 1981 SPE Annual Technical Conference and Exhibition, San Antonio, 5-7 October.
- 13) Ahmadi, A., Labastie, A., and Quintard, M.: "Large-Scale Properties for Flow Through a Stratified Medium: Various Approaches", SPERE (August 1993).
- 14) Tompang, R.: "Relative Effects of Gravitational and Viscous Forces on Water Flooding Performance in a Linear Stratified Reservoir", MS thesis, U. of Tulsa, Tulsa (1987).

APÉNDICE A.

- 1) "Conceptos básicos de Estadística", www.uam.es.
- 2) Hines, William W. y Montgomery, Douglas C.: "Probabilidad y Estadística para Ingeniería y Administración", Edit. CECOSA 3ª edición, México 1993.

APÉNDICE B.

- 1) "Resumen de medidas para variables cuantitativas (I)", www.liccom.edu.uy.