

3. Marco teórico

En esta sección se hizo una revisión de los principales conceptos de geoestadística y los diferentes algoritmos que se trabajaran para poblar la malla tridimensional construida con facies y propiedades petrofísicas

3.1 Marco teórico

En cualquier campo de petróleo donde se desee implementar un proyecto de recuperación secundaria por inyección de agua es necesario realizar una caracterización detallada de las zonas productoras de aceite con el fin de entender las heterogeneidades del yacimiento y como es el movimiento de los fluidos dentro de la roca. Los modelos tridimensionales poblados usando métodos geoestadísticos dan un modelo más aproximado de lo que está sucediendo en el subsuelo comparados con la información que suministran los modelos de dos dimensiones con los que usualmente se trabajan en esta industria (Jensen *et al.*, 2003). Por esta razón la implementación y uso de esta técnica ha ido incrementándose en los últimos años.

Los propósitos fundamentales al aplicar la geoestadística son los de predecir (Valores de la variable en sitios no muestreados) y simular (identificar el comportamiento espacial de la variable de interés bajo algunos supuestos fijados a priori) variables regionalizadas (Davis, 2003).

El trabajo de modelamiento geoestadístico se divide en dos actividades: la primera de estas es poblar con facies la malla 3D y la segunda es poblar con propiedades petrofísicas (porosidad y permeabilidad) la misma. Una vez finalizado el modelamiento de facies y propiedades se pueden realizar los cálculos de volúmenes de aceite, conociendo su distribución tanto areal como vertical en el yacimiento. Esto permitirá tomar decisiones más acordes a la realidad de los proyectos que se deseen implementar en el área.

3.2 Geoestadística

La geoestadística es una rama de la estadística aplicada que se especializa en el análisis, modelación y predicción de la variabilidad espacial de fenómenos en Ciencias de la Tierra. En geoestadística se trabaja con variables distribuidas espacialmente, estas variables no se consideran independientes como en la estadística clásica sino que están correlacionadas unas con otras en una dependencia espacial (variables regionalizadas). Esto indica que mientras más cercanos estén situados dos puntos hay más correlación y mientras más separados hay menos correlación (Caers, 2005).

Este comportamiento de variables distribuidas espacialmente se puede observar en los cuerpos de arenas asociados a depósitos fluviales como los de la Zona C de la Formación Mugrosa (Oligoceno Temprano). En este caso las variables serán: tipo de roca, porosidad y permeabilidad. Los valores de estas variables se conocen en cada uno de los pozos, pero se desconoce su valor en las áreas del yacimiento donde no se tienen pozos. Este proyecto pretende predecir el valor de estas variables y su correlación en cualquier punto del yacimiento mediante el uso de las diferentes herramientas de la geoestadística, lo cual permitirá conocer la geometría de los canales (roca reservorio de aceite) en cualquier punto del yacimiento.

Los estudios de Deutsch (2002), Jensen (2003) y Caers (2005) explican como el trabajo geoestadístico se desarrolla mediante tres actividades: análisis exploratorio de datos, análisis de la relación espacial (variografía) y predicción de valores (Simulación).

3.2.1 Análisis exploratorios de datos

Se basa en técnicas estadísticas y gráficas que permiten establecer el entendimiento básico del comportamiento de los datos y las relaciones existentes entre las variables que se estudian. Adicionalmente permite detectar fallos en el diseño y adquisición de datos, así como también la identificación de valores atípicos. Las herramientas más usadas en esta etapa son la estadística univariada, bivariada y regresiones. A continuación se hace una definición de cada una de estas herramientas.

- **Estadística univariada.** La estadística univariada es el análisis básico y primario de las características de una variable. En geoestadística se trabaja con variables aleatorias. Una variable aleatoria X es aquella que toma una serie de valores o realizaciones (X_i) con una probabilidad de ocurrencia (P_i) (ver Ecuación (3.1)).

$$P_i \geq 0 \text{ y } \sum P_i = 1 \tag{3.1}$$

Existen dos tipos de variables aleatorias:

- Discreta: está formada por un número finito de elementos
- Continua: está formada por un número infinito de valores

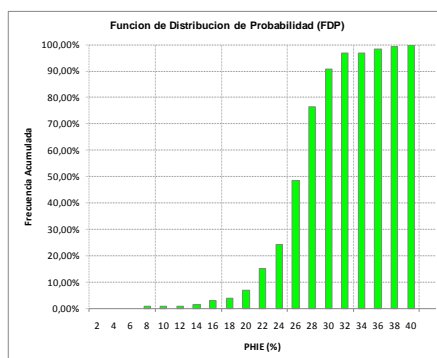
Las variables aleatorias (X) se caracterizan mediante:

- **Función de distribución de probabilidad (FDP).** Función que asigna a cada suceso definido sobre la variable aleatoria la probabilidad de que dicho suceso ocurra. Se usa para identificar el tipo de distribución y a determinar si están presentes poblaciones mixtas. Si la función de probabilidad solo toma valores positivos en un conjunto de valores finito se denomina distribución de variable aleatoria discreta (Jensen *et al.*, 2003) (ver Ecuación (3.2)).

$$F_X(x) = P(X \leq x) \in [0,1] \tag{3.2}$$

Su gráfica es el Histograma acumulativo. La figura 3-1, muestra un ejemplo de Función de distribución de probabilidad.

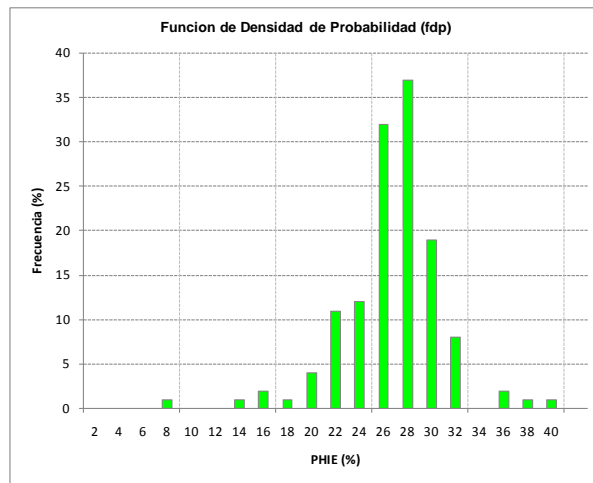
Figura 3-1: Función de distribución de probabilidad (FDP).



- **Función de densidad de probabilidad (fdp).** Función matemática que caracteriza el comportamiento probable de una población. La función de densidad define la concentración de probabilidad alrededor de los valores de una variable aleatoria continua (Jensen *et al.*, 2003) (ver Ecuación (3.3)). Su gráfica es el Histograma (Ver figura 3-2).

$$f(z) = \frac{dF(z)}{dz} \quad (3.3)$$

Figura 3-2: Función de densidad de probabilidad (fdp).



Parámetros una Variable aleatoria. La función de densidad o distribución de probabilidad de una variable aleatoria contienen detalladamente toda la información sobre la variable. El parámetro resume en valores numéricos todas sus características principales. Estos parámetros son:

- **Esperanza-momento de primer orden.** Es el valor más probable que puede tomar una variable aleatoria. También se conoce como valor esperado, media poblacional o media. Se define como la suma del producto de la probabilidad de cada suceso por el valor de dicho suceso (Armstrong, 1998). La esperanza de una variable aleatoria discreta se define como (ver Ecuación (3.4)):

$$m = E[X] = \sum_{i=1}^n x_i p(x_i) \quad (3.4)$$

La esperanza de una variable aleatoria continua se define como (ver Ecuación (3.5)):

$$m = E[X] = \int_{-\infty}^{\infty} xf(x)dx \quad (3.5)$$

• **Momento de orden r de una FDP.** Se define como (ver Ecuación (3.6)):

$$m_r = E[Z^r] = \int_{-\infty}^{+\infty} z^r dF(z) = \int_{-\infty}^{+\infty} z^r f(z)dz \quad (3.6)$$

• **Momento central de orden r de una FDP.** Se define como (ver Ecuación (3.7)):

$$\mu_r = E[(Z - m)^r] = \int_{-\infty}^{+\infty} (z - m)^r dF(z) = \int_{-\infty}^{+\infty} (z - m)^r f(z)dz \quad (3.7)$$

• **Varianza (σ^2)-momento central de segundo orden.** Es una medida de la dispersión de una variable aleatoria respecto a su esperanza. Se define como la esperanza del cuadrado de la desviación de dicha variable respecto a su media (Armstrong, 1998) (ver Ecuación (3.8)):

$$\sigma^2 = V(X) = E[(X - E[X])^2] \geq 0 \quad (3.8)$$

La varianza puede verse influenciada por los valores atípicos y está relacionada con la desviación estándar, la cual es la raíz cuadrada de la varianza (ver Ecuación (3.9)).

$$\sigma = \sqrt{V(X)} \quad \text{o} \quad \sigma^2 = V(X) \quad (3.9)$$

La varianza para una variable aleatoria continua es (ver Ecuación (3.10)):

$$V(X) = \int (x - \mu)^2 f(x)dx, \text{ donde } \mu = \int xf(x)dx \text{ y las integrales est\u00e1n definidas sobre el rango de } x. \quad (3.10)$$

La varianza de una variable aleatoria discreta es (ver Ecuación (3.11)):

$$V(X) = \sum_{i=1}^n (x_i - \mu)^2 p_i \quad \text{o} \quad (\sigma^2) = \frac{1}{n-1} \sum_{i=1}^n (x_i - m)^2 \quad (3.11)$$

• **Desviación estándar (σ).** La desviación estándar o desviación típica (es una medida de centralización o dispersión. Se define como la raíz cuadrada de la varianza.

Junto con este valor, la desviación típica es una medida (cuadrática) que informa de la media de distancias que tienen los datos respecto de su media aritmética, expresada en las mismas unidades que la variable. Se define como (ver Ecuación (3.12)):

$$\sigma = \sqrt{Var [Z]} \quad (3.12)$$

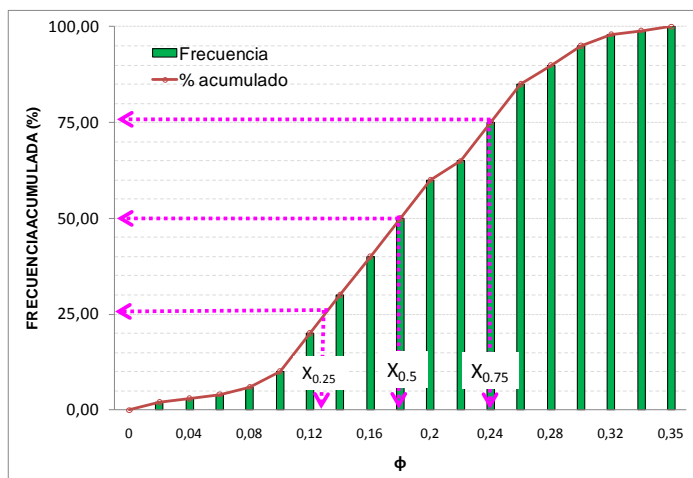
Cuántiles. Es una medida de posición no central que permite conocer otros puntos característicos de la distribución. Se definen como aquellos valores de la variable que ordenados de menor a mayor dividen la distribución en partes de tal manera que cada una de ellas contiene el mismo número de frecuencias (Armstrong, 1998). Algunos cuantiles importantes son:

- Percentiles. Dividen la distribución en cien partes. $F(X_p) = p$

Cuartiles. Son los tres valores que dividen al conjunto de datos ordenados en cuatro partes porcentualmente iguales (ver figura 3-3):

- Primer cuartil $p = 0.25$; $X_{0.25} = F^{-1}(0.25)$
- Mediana: $p=0.5$; $M = F^{-1}(0.5)$
- Tercer cuartil $p = 0.75$; $X_{0.75} = F^{-1}(0.75)$
- Rango intercuartil: diferencia entre el tercer y primer cuartil $[X_{0.25}, X_{0.75}]$

Figura 3-3: Ejemplo de Cuartiles: $X_{0.25}$, $X_{0.5}$ y $X_{0.75}$



Otras medidas descriptivas son:

- **Media aritmética.** La media o promedio aritmético de un conjunto finito de números es igual a la suma de todos sus valores dividida entre el número de sumandos. Dados los n números $\{a_1, a_2, \dots, a_n\}$, la media aritmética se define simplemente como (ver Ecuación (3.13)):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n a_i = \frac{a_1 + a_2 + \dots + a_n}{n} \quad (3.13)$$

- **Media geométrica.** La media o promedio geométrico de una cantidad arbitraria de números (n números) es la raíz n -ésima del producto de todos los números. Se define como (ver Ecuación (3.14)):

$$\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i} = \sqrt[n]{x_1 \cdot x_2 \cdot \dots \cdot x_n} \quad (3.14)$$

La media geométrica se usa por que considera todos los valores de la distribución y es menos sensible que la media aritmética a los valores extremos en distribuciones logarítmicas.

- **Moda.** Es el valor con una mayor frecuencia en una distribución de datos o el dato más repetido (Journel *et al.*, 1978). Se usa para trabajar con variables aleatorias discretas como por ejemplo Facies. Se define como (ver Ecuación (3.15)):

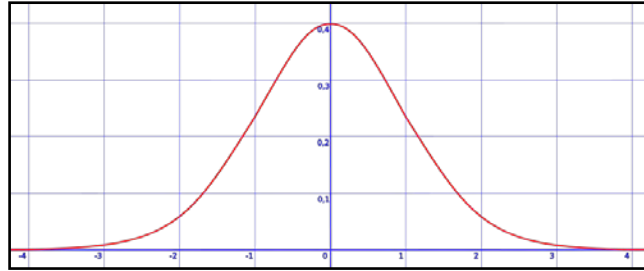
$$Md = x_i, \text{ Si } n_i = \max \{f_j, j \in \{1, 2, \dots, k\}\} \quad (3.15)$$

- **Distribución Normal** (Distribución de Gauss o Distribución Gaussiana). Es una distribución de probabilidad de una variable aleatoria continua. La grafica de su función de densidad tiene la forma de una campana (un solo pico: unimodal) y es simétrica respecto a un determinado parámetro (media) (Ver Figura 3-4). Los parámetros de media, mediana y moda coinciden. Esta distribución es importante porque permite modelar numerosos fenómenos naturales (Journel *et al.*, 1978). La distribución normal está caracterizada por la media y la varianza: $N(m, \sigma^2)$ (Díaz, 2002).

La Función de densidad de probabilidad de una distribución normal se define como (ver Ecuación (3.16)):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}, \quad x \in \mathbb{R} \quad (3.16)$$

Figura 3-4: Grafica de una distribución Normal.



- **Distribución Log-Normal.** Es una distribución de probabilidad de cualquier variable aleatoria con su logaritmo normalmente distribuido. La mayoría de distribuciones experimentales en ciencias de la tierra tienden a ser asimétricas (media y mediana diferente) y las variables en su mayoría toman valores NO negativos (Journel *et al.*, 1978) (Ver Figura 3-5). Una variable aleatoria positiva X tiene una distribución Log-Normal si su logaritmo $\ln(Y)$ esta normalmente distribuido. Si X (variable aleatoria) tiene una distribución normal entonces e^X tiene una distribución Log-Normal. Se define como (ver Ecuación (3.17)):

$$Y > 0 \rightarrow \log N(\mu, \sigma^2), \quad \text{si } X = \ln Y \rightarrow N(\alpha, \beta^2) \quad (3.17)$$

La distribución Log-Normal tiene una función densidad de probabilidad (ver Ecuación (3.18)):

$$f(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}} \quad (3.18)$$

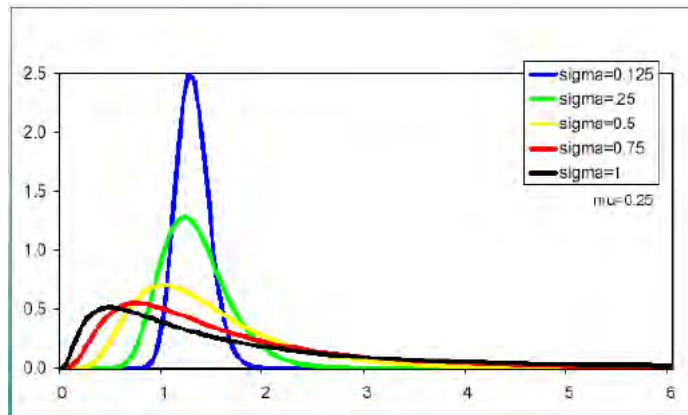
Para $x > 0$, donde μ (media) y σ (desviación estándar) del logaritmo de variable. La esperanza (E) y la varianza (V) en una distribución Log-Normal se definen como (ver Ecuación (3.19)) y (ver Ecuación (3.20)):

$$E(X) = e^{\frac{\mu + \sigma^2}{2}} \tag{3.19}$$

$$V(X) = (e^{\sigma^2} - 1)e^{2\mu + \sigma^2} \tag{3.20}$$

La distribución Log-Normal, la media geométrica y la desviación estándar geométrica están relacionadas. La Media geométrica es e^μ y la desviación estándar geométrica es e^σ .

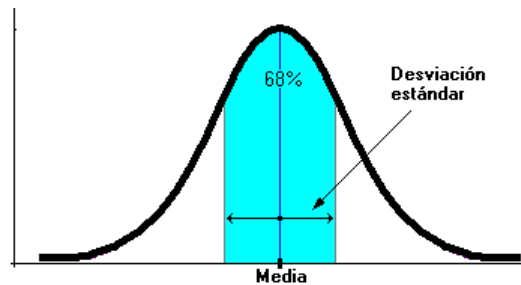
Figura 3-5: Graficas de Distribuciones Log-Normal.



Modificado de Díaz (2002).

Para conocer con detalle un conjunto de datos, no basta con conocer las medidas de tendencia central, sino que necesitamos conocer también la desviación que representan los datos en su distribución respecto de la media aritmética de dicha distribución (Ver Figura 3-6). El objetivo es tener una visión de los mismos más acorde con la realidad a la hora de describirlos e interpretarlos para la toma de decisiones (Caers, 2005).

Figura 3-6: Desviación estándar.



Coefficiente de variación (CV). Es una medida de dispersión útil para comparar dispersiones a escalas distintas pues es una medida invariante ante cambios de escala. Sirve para comparar variables que están a distintas escalas pero que están correlacionadas estadísticamente y sustantivamente con un factor en común (Journel *et al.*, 1978). Se define como (ver Ecuación (3.21)):

$$CV = \sigma/m \quad (3.21)$$

Coefficiente de Asimetría (medida de la simetría). Las medidas de asimetría son indicadores que permiten establecer el grado de simetría (o asimetría) que presenta una distribución de probabilidad de una variable aleatoria sin tener que hacer su representación gráfica.

Como eje de simetría consideramos una recta paralela al eje de ordenadas que pasa por la media de la distribución. Si una distribución es simétrica, existe el mismo número de valores a la derecha que a la izquierda de la media, por tanto, el mismo número de desviaciones con signo positivo que con signo negativo. Hay asimetría positiva (o a la derecha) si la "cola" a la derecha de la media es más larga que la de la izquierda, es decir, si hay valores más separados de la media a la derecha. Existe asimetría negativa (o a la izquierda) si la "cola" a la izquierda de la media es más larga que la de la derecha, es decir, si hay valores más separados de la media a la izquierda.

La medida de asimetría más utilizada parte del uso del tercer momento estándar por que nos interesa mantener el signo de las desviaciones con respecto a la media, para obtener si son mayores las que ocurren a la derecha de la media que las de la izquierda. El coeficiente de asimetría se define como (ver Ecuación (3.22)):

$$\alpha_1 = \frac{\mu_3}{\sigma^3} \quad (3.22)$$

Donde μ^3 es el tercer momento en torno a la media y σ es la desviación estándar.

- Si $\alpha_1 = 0$, la distribución es simétrica.
- Si $\alpha_1 > 0$, la distribución es asimétrica positiva o a la derecha.

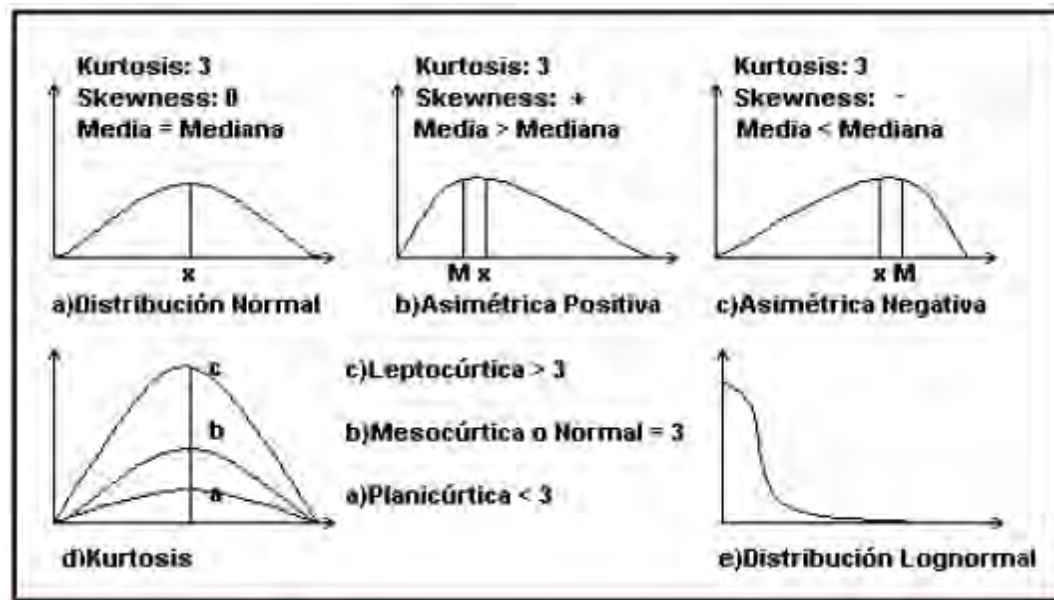
- Si $\alpha_1 < 0$, la distribución es asimétrica negativa o a la izquierda.

Coefficiente de curtosis (medida del achatamiento). Es una medida de la forma o apuntamiento de las distribuciones. Las medidas de curtosis estudian la mayor o menor concentración de frecuencias alrededor de la media y en la zona central de la distribución. Se define como (ver Ecuación (3.23)):

$$\alpha_2 = \frac{\mu_4}{\sigma^4} - 3 \tag{3.23}$$

Donde μ^4 es el 4º momento centrado o con respecto a la media \bar{x} y σ es la desviación estándar. La figura 24 muestra varios ejemplos de cómo es la simetría y curtosis en diferentes clases de distribuciones.

Figura 3-7: Simetría y curtosis de una distribución.



Modificado de Díaz (2002).

Estadística bivariada. Según Deutsch (2002) se define como el análisis en conjunto de dos variables aleatorias debido a que en la caracterización de un yacimiento es más importante conocer el patrón de dependencia que relaciona a una variable aleatoria X con otra variable aleatoria Y, como por ejemplo la porosidad y permeabilidad.

- **Función de Distribución de Probabilidad Bivariada.** La distribución de probabilidad conjunta de un par de variables aleatorias (X, Y) se define como (ver Ecuación (3.24)):

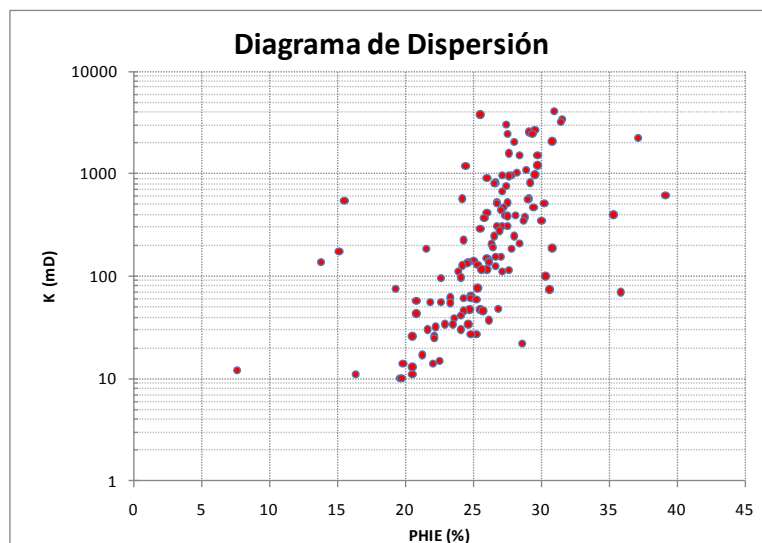
$$F_{XY}(x, y) = Pr\{X \leq x, Y \leq y\} \quad (3.24)$$

En la práctica se estima mediante la proporción de pares de valores de (X, Y) que se encuentran por debajo del umbral x, y, respectivamente (Caers, 2005).

- **Diagrama de Dispersión (*Scatter plot*).** Es el equivalente bivariado del histograma, donde cada par (x_i, y_i) es un punto. El grado de dependencia entre dos variables aleatorias X vs Y puede ser caracterizado por el diagrama de dispersión alrededor de cualquier línea de regresión (Caers, 2005).

Un diagrama de dispersión es un tipo de diagrama matemático que utiliza las coordenadas cartesianas para mostrar los valores de dos variables para un conjunto de datos. Los datos se muestran como un conjunto de puntos, cada uno con el valor de una variable que determina la posición en el eje horizontal y el valor de la otra variable determinado por la posición en el eje vertical (Ver Figura 3-8).

Figura 3-8: Diagrama de dispersión análisis bivariado.



- **Covarianza.** Según Deutsch (2002) se define la covarianza de manera análoga a los momentos centrales univariados, como (ver Ecuación (3.25)):

$$Cov(X, Y) = \sigma_{XY} = E\{(X - m_X)(Y - m_Y)\} \quad (3.25)$$

En estadística la covarianza es una medida de dispersión conjunta de dos variables estadísticas. La covarianza se estima así (ver Ecuación (3.26)):

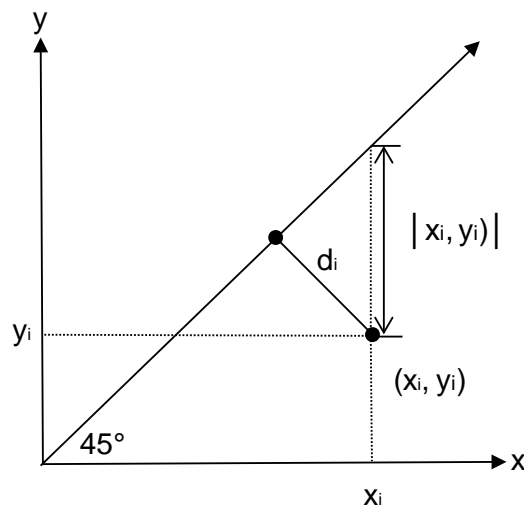
$$\sigma_{XY} = \frac{1}{N} \sum_{i=1}^N (x_i - m_X)(y_i - m_Y) = \frac{1}{N} \sum_{i=1}^N x_i y_i - m_X m_Y \quad (3.26)$$

- **Semivariograma.** Es una herramienta que permite analizar el comportamiento espacial de una variable sobre un área definida, obteniendo como resultado la influencia de los datos a diferentes distancias (Permite caracterizar la carencia de dependencia). Según Caers (2005) es el momento de inercia del diagrama de dispersión con respecto a una línea con pendiente de 45° y se define como (ver Ecuación (3.27)):

$$\gamma_{XY} = \frac{1}{N} \sum_{i=1}^N [d_i]^2 = \frac{1}{2N} \sum_{i=1}^N [x_i - y_i]^2 \quad (3.27)$$

De acuerdo a Goovaerts (1997) mientras mayor sea el valor del semivariograma más dispersos estarán los valores en el diagrama de dispersión y menor será la dependencia entre las dos variables aleatorias (Ver Figura 3-9).

Figura 3-9: Semivariograma.



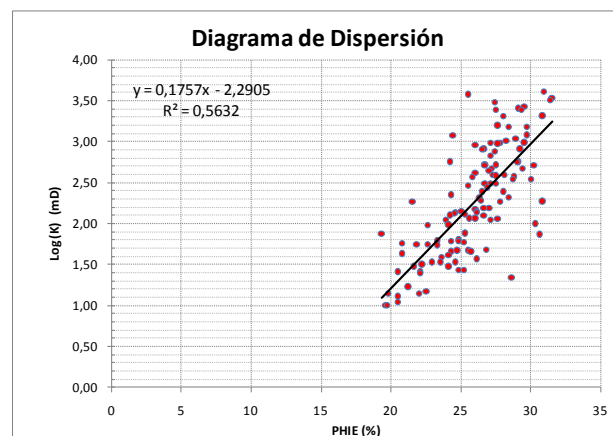
- **Coefficiente de correlación lineal de Pearson.** Caers (2005) define el coeficiente de correlación de *Pearson* como un índice que mide la relación lineal entre dos variables aleatorias cuantitativas (Ver Figura 27). A diferencia de la covarianza, la correlación de *Pearson* es independiente de la escala de medida de las variables. Se define como (ver Ecuación (3.28)):

$$r_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = \frac{Cov\{X,Y\}}{\sqrt{Var\{X\} Var\{Y\}}} \in [-1, 1] \quad (3.28)$$

El valor del índice de correlación varía en el intervalo [-1, +1], donde:

- Si $r = 1$, existe una correlación positiva perfecta. El índice indica una dependencia total entre las dos variables denominada relación directa: cuando una de ellas aumenta, la otra también lo hace en proporción constante.
- Si $0 < r < 1$, existe una correlación positiva.
- Si $r = 0$, no existe relación lineal. Pero esto no necesariamente implica que las variables son independientes: pueden existir todavía relaciones no lineales entre las dos variables.
- Si $-1 < r < 0$, existe una correlación negativa.
- Si $r = -1$, existe una correlación negativa perfecta. El índice indica una dependencia total entre las dos variables llamada relación inversa: cuando una de ellas aumenta, la otra disminuye en proporción constante

Figura 3-10: Diagrama de Dispersión con su correlación lineal.



Regresiones. Una regresión intenta establecer relaciones funcionales entre variables aleatorias. Los modelos de regresión nos permiten hacer predicciones o pronósticos a partir del modelo establecido. La más usada es la Regresión Lineal.

- **Regresión lineal.** Consiste en establecer una relación descrita mediante una recta. Dados n valores de de dos variables aleatorias X y Y donde:

- X es una variable independiente.
- Y depende de X en forma lineal.

Entonces la ecuación del modelo lineal es $Y = B_0 + B_1X$, o (ver Ecuación (3.29)):

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, N \quad (3.29)$$

Donde: B_0, B_1 =son los parámetros del modelo y e_i = errores o residuos del modelo. El residuo debe cumplir las siguientes condiciones:

- Valor esperado cero: $E\{e_i\} = 0$
- Varianza constante: $\text{Var}\{e_i\} = \sigma_e^2$
- No correlacionados: $\text{Cov}\{e_i, e_j\} = 0$
- Distribución normal: $e \sim N(0, \sigma_e^2)$.

La primera forma de regresiones lineales documentada fue el método de los mínimos cuadrados (Caers, 2005).

- **Método de mínimos cuadrados ordinarios.** Es una técnica de análisis numérico en la que dado un conjunto de pares (o ternas, etc), se intenta encontrar la función que mejor se aproxime a los datos, de acuerdo con el criterio de mínimo error cuadrático. Así este método consiste en hallar los parámetros del modelo de manera que la suma de los cuadrados de los errores (SCR) sea mínima. Dadas n observaciones (Y_1, Y_2, \dots, Y_n) de la variable dependiente Y , correspondientes a n experimentos, el problema que se plantea es la estimación de los parámetros $(B_0, B_1, B_2, \dots, B_{p-1})$ de manera que sea mínima la suma de los cuadrados de los errores (SCR) (ver Ecuación (3.30)):

$$SCR = \sum_{i=1}^N e_i^2 = \sum_{i=1}^N [y_i - \hat{y}_i]^2 = \sum_{i=1}^N [y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)]^2 \quad (3.30)$$

Entonces el sistema de ecuaciones a resolver para predecir el valor de la variable dependiente es (ver Ecuación (3.31)):

$$\frac{\partial SCR}{\partial \beta_0} = 0, \quad \frac{\partial SCR}{\partial \beta_1} = 0 \quad (3.31)$$

Coefficiente de determinación R². Este coeficiente en modelos lineales mide el grado de la bondad del ajuste. Este coeficiente es igual al coeficiente de correlación lineal al cuadrado y representa la proporción de varianza explicada por la regresión lineal. A continuación se enumeran algunos criterios de la bondad del ajuste:

- Si $R^2 \gg 1$, el ajuste es bueno (Y se puede calcular de modo bastante aproximado a partir de X y viceversa).
- Si $R^2 \gg 0$, las variables X y Y no están relacionadas y por tanto no tiene sentido hacer un ajuste lineal.
- Sin embargo, no es seguro que las dos variables no posean ninguna relación en el caso $R=0$, ya que si bien el ajuste lineal puede no ser procedente, tal vez otro tipo de ajuste sí lo sea.

Otros conceptos importantes son:

- **Función Aleatoria (FA).** Es un proceso estocástico. Si a cada punto \underline{x} que pertenece a un dominio Ω en el espacio le hacemos corresponder una variable aleatoria Z, entonces el conjunto de variables aleatorias espacialmente distribuidas será una función aleatoria $Z(\underline{x})$ (Jensen *et al.*, 2003). Como ejemplo de una Función aleatoria está la distribución espacial de las facies o la porosidad en un yacimiento.

- **Variable Aleatoria Regionalizada.** Desde el punto de vista matemático una variable regionalizada es simplemente una función $f(x)$ que toma valores en todos los puntos x de coordenadas (x_i, y_i, z_i) en el espacio tridimensional, sin embargo, es muy frecuente que estas funciones varíen tan irregularmente en el espacio que impiden un estudio

matemático directo, y se hace necesario realizar un análisis de variabilidad de la información disponible (Jensen *et al.*, 2003).

- **Función de distribución de una FA.** Sea una función aleatoria $Z(x)$ definida en una región Ω , entonces se define el vector aleatorio $\{Z(x_1), Z(x_2), \dots, Z(x_n)\}$ se caracteriza por su función de distribución de probabilidad n-variada (ver Ecuación (3.32)):

$$F_{Z(\underline{x}_1), Z(\underline{x}_2), \dots, Z(\underline{x}_n)}(Z_1, Z_2, \dots, Z_n) = \Pr[Z(\underline{x}_1) \leq z_1, Z(\underline{x}_2) \leq z_2, \dots, Z(\underline{x}_n) \leq z_n] \quad (3.32)$$

El conjunto de todas las distribuciones para todo valor de n y para cualquier selección de puntos en W constituye la ley espacial de probabilidad de la función aleatoria. Esta función en la práctica es imposible de determinar y sólo se puede inferir los primeros momentos de la distribución de la FA $Z(\underline{x})$ (Díaz, 2002).

- **Momentos de una Función Aleatoria.** El momento de primer orden se conoce como la media de $Z(\underline{x})$ y se define como (ver Ecuación (3.33)):

$$m(\underline{x}) = E[Z(\underline{x})] \quad (3.33)$$

Los momentos de segundo orden son: La varianza y la Covarianza de $Z(\underline{x})$. La varianza se define como (ver Ecuación (3.34)):

$$\sigma^2(\underline{x}) = \text{Var}[Z(\underline{x})] = E\left[\{Z(\underline{x}) - m(\underline{x})\}^2\right] \quad (3.34)$$

Y la covarianza se define como (ver Ecuación (3.35)):

$$C(\underline{x}_i, \underline{x}_j) = E\left[\{Z(\underline{x}_i) - m(\underline{x}_i)\}\{Z(\underline{x}_j) - m(\underline{x}_j)\}\right] \quad (3.35)$$

Otro momento de segundo orden es el semivariograma de $Z(x)$. Igualmente se conoce como función de semivarianzas o variograma y se define como (ver Ecuación (3.36)):

$$2\gamma(\underline{x}_i, \underline{x}_j) = Var[Z(\underline{x}_i) - Z(\underline{x}_j)] \quad (3.36)$$

, entonces (ver Ecuación (3.37)):

$$\gamma(\underline{x}_i, \underline{x}_j) = \frac{1}{2} E \left[\{Z(\underline{x}_i) - Z(\underline{x}_j)\}^2 \right] \quad (3.37)$$

Estacionaridad de una Función Aleatoria. Según Davis (2003) una función aleatoria es estrictamente estacionaria si su función de distribución de probabilidad es invariante a cualquier traslación respecto a un vector h . Las funciones aleatorias se clasifican según su grado de estacionaridad en:

- **Estacionarias de segundo orden.** Sus momentos de primer y segundo orden no dependen de la posición, es decir (ver Ecuaciones (3.38), (3.39) y (3.40)):

$$E[Z(\underline{x})] = m \quad y \quad Var[Z(\underline{x})] = \sigma^2 \quad \forall \underline{x} \quad (3.38)$$

$$C(\underline{h}) \equiv C(\underline{x} + \underline{h}, \underline{x}) = E[Z(\underline{x} + \underline{h})Z(\underline{x})] - m^2 \quad (3.39)$$

$$\gamma(\underline{h}) = \gamma(\underline{x} + \underline{h}, \underline{x}) = \frac{1}{2} E \left[\{Z(\underline{x} + \underline{h}) - Z(\underline{x})\}^2 \right] \quad (3.40)$$

- **Aleatorias intrínsecas.** Cuando la FA no es estacionaria pero las diferencias $Z(\underline{x} + \underline{h}) - Z(\underline{x})$ son estacionarias de segundo orden (Hipótesis Intrínseca). En este caso el valor esperado se define como (ver Ecuación (3.41)):

$$E[Z(\underline{x} + \underline{h}) - Z(\underline{x})] = m \quad (3.41)$$

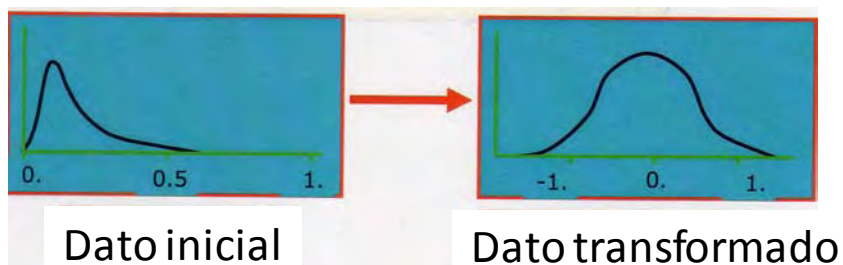
La varianza se define como (ver Ecuación (3.42)):

$$Var[Z(\underline{x} + \underline{h}) - Z(\underline{x})] = 2\gamma(\underline{h}) \quad (3.42)$$

- **No estacionarias:** Un función aleatoria es no estacionaria cuando no cumple con la Hipótesis intrínseca. Entonces el valor esperado depende de la posición y la varianza no es estacionaria.

Transformación “Normal Score” de los datos. Es un proceso que usa funciones matemáticas y otras técnicas para convertir datos desde un tipo de distribución a otra. Por ejemplo lognormal a normal (Ver Figura 3-11).

Figura 3-11: Transformación de datos.



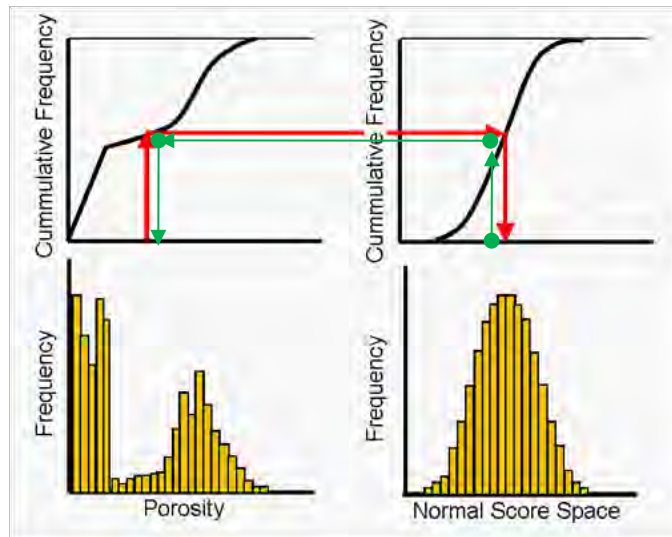
La simulación Secuencial Gaussiana usa un modelo de variograma espacial y datos de registros escalados normalizados para asignar valores a celdas de la malla 3D. Adicionalmente requiere:

- Estacionaridad (no tendencias espaciales).
- Distribución normal estandarizada (media a ser cero y desviación estándar a ser 1).
- Modelo de variograma espacial.

La transformación de los datos de entrada debe hacerse antes de la simulación y se hace usando una transformación normal “score”, así mismo los resultados se transforman de regreso usando la misma transformación. Normal “score” es un tipo de transformación que usa histogramas o una función fija a un histograma para convertir a una distribución normal (media cero, desviación estándar uno).

La Figura 3-12 muestra el proceso de transformación “Normal Score”. En este ejemplo los datos de porosidad generan un histograma de distribución binomial con su curva de Frecuencia acumulada. Cada dato original se transforma a su equivalente en la curva de frecuencia acumulada de una distribución normal. De esta forma todos los datos son transformados. De igual forma los resultados de la simulación se regresan usando la misma transformación.

Figura 3-12: Metodología Transformación Normal Score.



Modificado de Petrel (2008).

3.2.2 Análisis de la relación espacial (estructural)

Según Caers (2005) y Deutsch (2002) es el proceso de estimación y modelación de la función que describe la correlación espacial de las variables. Este análisis es uno de los temas más importante de la geoestadística puesto que se encarga de la caracterización de la estructura espacial de una propiedad o fenómeno regionalizado. Es el proceso en el marco del cual se obtiene un modelo geoestadístico para la función aleatoria que se estudia

En el análisis estructural se estima y modela una función que refleje la correlación espacial de la variable regionalizada a partir de la adopción razonada de la hipótesis más adecuada acerca de su variabilidad. De acuerdo a Davis (2003) para este análisis se necesita estimar y modelar la función de semivarianzas o semivariograma.

- **Definición de semivariograma o variograma.** El variograma o semivariograma es una herramienta que permite analizar el comportamiento espacial de una variable sobre un área definida, obteniendo como resultado la influencia de los datos a diferentes distancias. A partir de los datos proporcionados por el variograma teórico se pueden

realizar estimaciones. El variograma es la herramienta más importante de la geoestadística (Deutsch, 2002).

Jensen *et al.* (2003) define el variograma como la varianza de las diferencias de los valores de la variable regionalizada en las localizaciones separadas una distancia h . Si tenemos una Función aleatoria $Z(\underline{x})$ que cumpla con la hipótesis intrínseca entonces se define la semivarianza como (ver Ecuación (3.43)):

$$\gamma(\underline{h}) = \frac{1}{2} Var[Z(\underline{x}) - Z(\underline{x} + \underline{h})] = \frac{1}{2} E \left[\{Z(\underline{x}) - Z(\underline{x} + \underline{h})\}^2 \right] \quad (3.43)$$

Para estimar el semivariograma se usa la siguiente ecuación (ver Ecuación (3.44)):

$$\gamma(\underline{h}) = \frac{1}{2N(\underline{h})} \sum_{i=1}^{N(\underline{h})} [Z(\underline{x}_i) - Z(\underline{x}_i + \underline{h})]^2 \quad (3.44)$$

Donde:

- $N(\underline{h})$ es el número de pares a la distancia h
- \underline{h} es el incremento
- $Z(x_i)$ son los valores experimentales
- X_i : localizaciones donde son medidos los valores de $Z(x_i)$

Forma del Variograma. La Figura 3-13 ilustra la forma general de un variograma y muestra sus principales características. Las partes de un variograma son:

- **Varianza.** Definida por el cálculo de la diferencia entre puntos para la 1ra, 2da, 3ra, etc (distancia *Lag*). También se define como el promedio del grado de diferencia entre pares de puntos.
- **Lag:** distancia que separa los datos (puntos).
- **Alcance o Range.** Es el punto sobre el eje "X" que corresponde a donde la varianza se vuelve plana. Este define la distancia de correlación. Según Davis (2003) es la distancia a la cual el variograma se estabiliza o distancias para la cual los valores de la variable dejan de estar correlacionados.

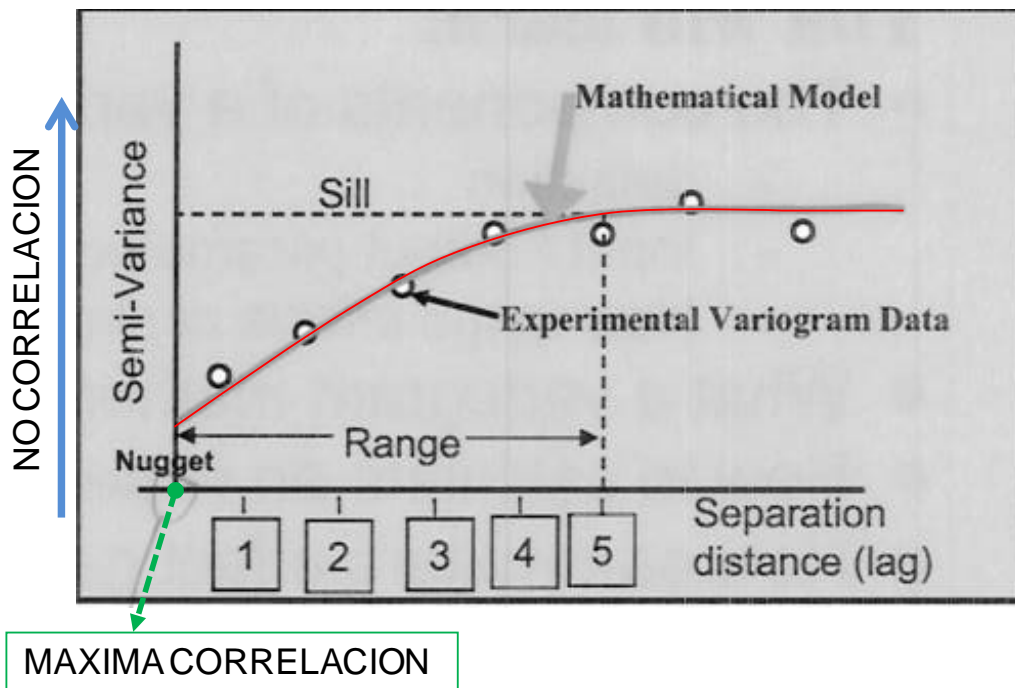
- **Meseta o Sill.** Muestra la máxima diferencia, también se refiere a “La variación total del sistema”. La meseta es el punto sobre el eje “Y” que corresponde a donde los puntos de la variación se vuelven planos horizontalmente. Otra definición según Davis (2003) es el valor constante que toma el variograma en distancias mayores al rango, Varianza de la función aleatoria $Z(x)$.

- **Efecto pepita o “nugget effect” (discontinuidad en el origen).** Indica el grado de disimilaridad en la distancia cero. El efecto puede aparecer debido a errores en las medidas pero este también puede representar heterogeneidad a pequeña escala (Goovaerts, 1997).

Los principales usos del variograma son:

- Detectar direcciones de anisotropía: es decir en situaciones en que la variación depende de la dirección.
- Zonas de influencia y su extensión (correlación espacial).
- Variabilidad con la distancia.

Figura 3-13: Representación gráfica de un Variograma.



Modificado de Jensen *et al.* (2003).

Modelos de Variogramas. Los modelos de variogramas que se trabajaran en este estudio son:

- **Modelo esférico.** La terminología “variograma esférico” se explica porque $C - g(h)$ es proporcional al volumen de la intersección de dos esferas de diámetro a , separadas por h . El alcance corresponde al parámetro a , la meseta al parámetro C (Jensen *et al.*, 2003). Se define como (ver Ecuación (3.45)):

$$\gamma(h) = \begin{cases} C \left[\frac{3h}{2a} - \frac{1}{2} \frac{h^3}{a^3} \right], & \text{si } h \leq a \\ C, & \text{en caso contrario} \end{cases} \quad (3.45)$$

- **Modelo exponencial.** Modelo similar al esférico pero asciende más abruptamente, tiene más curvatura y alcanza la meseta C sólo en forma asintótica. El alcance práctico está donde $g(h)$ vale $0.95 C$ y corresponde al parámetro a (Jensen *et al.*, 2003). Este modelo es el más usado pues genera estimaciones más acordes con lo que sucede en el Yacimiento. Se define como (ver Ecuación (3.46)):

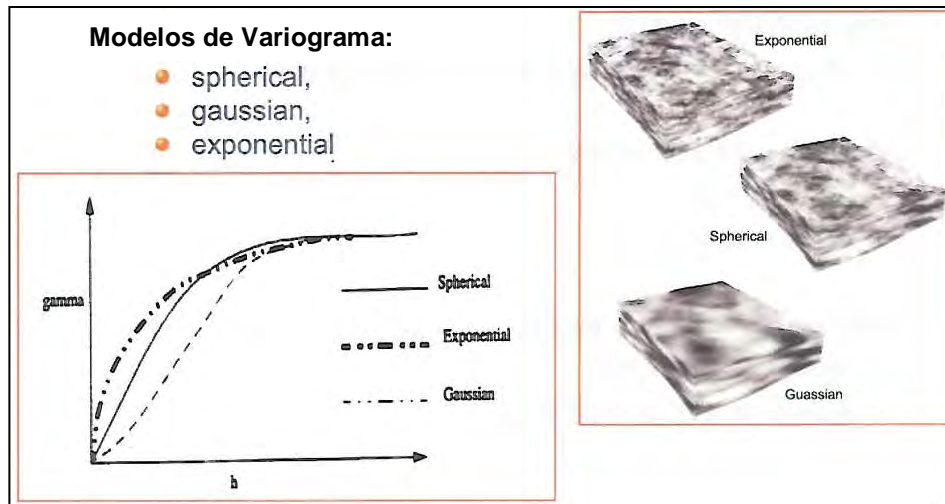
$$\gamma(h) = C \left(1 - \exp\left(-\frac{3h}{a}\right) \right) \quad (3.46)$$

- **Modelo gaussiano.** El modelo implica continuidad de corta escala debido a su comportamiento parabólico en el origen. Adecuado para variables como espesor. El alcance práctico se define cuando $g(h)$ vale $0.95 C$ y corresponde al parámetro a (Jensen *et al.*, 2003). Se define como (ver Ecuación (3.47)):

$$\gamma(h) = C \left(1 - \exp\left(-\frac{3h^2}{a^2}\right) \right) \quad (3.47)$$

La Figura 3-14 muestra los modelos de variograma esférico, exponencial y gaussiano, así como también su efecto sobre los modelos 3D del subsuelo. Se observa como el modelo gaussiano tiende a suavizar las estimaciones mientras que en el exponencial se tiende a generar demasiado ruido en las estimaciones.

Figura 3-14: Modelos de Variograma y sus efectos.



Modificado de Jensen *et al.* (2003).

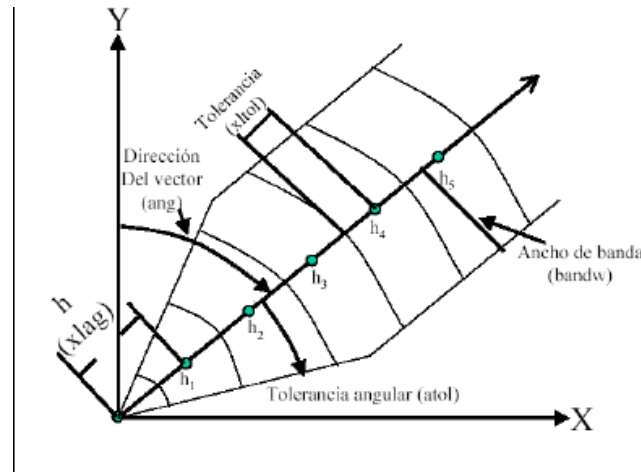
Para la construcción del variograma experimental se deben tener en cuenta los siguientes parámetros (ver Figura 3-15):

- Rango y dirección en la dirección principal
- Rango y dirección en la dirección menor
- Rango y dirección en la vertical
- *Sill*
- *Nugget*
- Tipo de Variograma
- Radio de investigación
- Angulo de tolerancia
- Tolerancia del *Lag*
- Número de *Lags*
- Distancia del *lag*

Comúnmente calculamos variogramas experimentales en direcciones ortogonales: vertical, horizontal dirección principal y horizontal dirección menor.

Es conveniente hacer un análisis de los tipos de anisotropía más comunes: anisotropía geométrica y la anisotropía zonal:

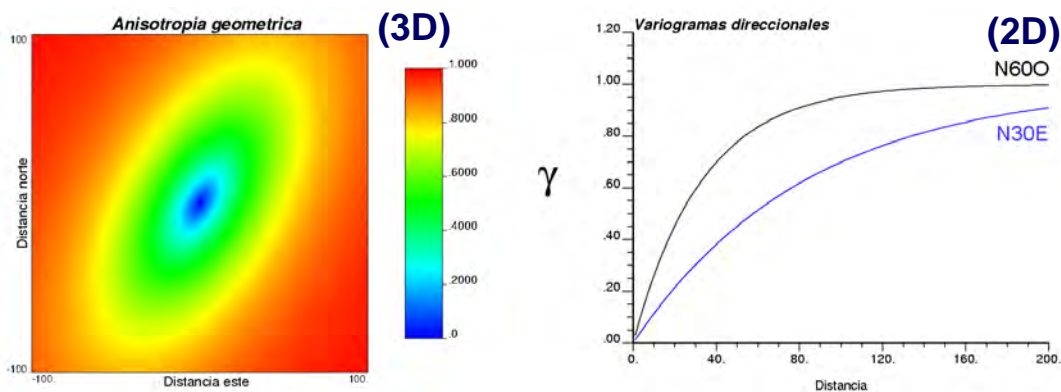
Figura 3-15: Parámetros para la construcción de un variograma.



Modificado de Jensen *et al.* (2003).

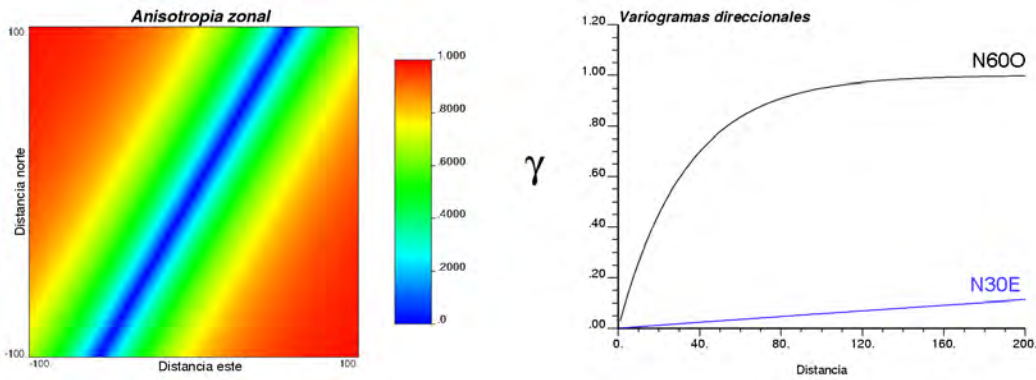
- **Anisotropía Geométrica.** El mapa variográfico dibuja elipses (2D) o elipsoides (3D). El modelamiento sólo requiere especificar las direcciones principales (ortogonales entre sí) y los alcances correspondientes (Jensen *et al.*, 2003). La Figura 3-16 muestra el efecto de anisotropía geométrica 2D y 3D.
- **Anisotropía Zonal.** El mapa variográfico dibuja bandas. Se trata de un caso límite de anisotropía geométrica, donde el alcance en una dirección se vuelve muy grande. A la escala de trabajo, la meseta cambia según la dirección (Jensen *et al.*, 2003) (Ver figura 3-17).
- **Anisotropías Complejas.** Se obtiene formas más complejas de anisotropía al mezclar anisotropías geométricas y/o zonales de orientación y razón diferentes (Jensen *et al.*, 2003) (Ver figura 3-18)

Figura 3-16: Anisotropía Geométrica 2D y 3D.



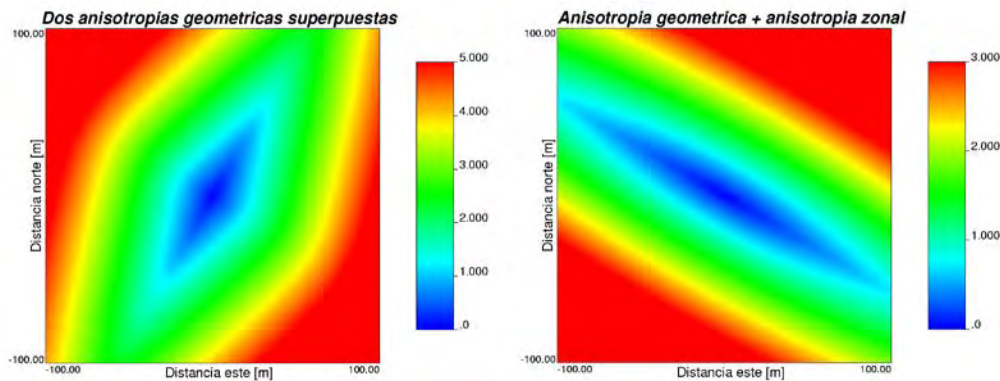
Modificado de Díaz (2002).

Figura 3-17: Anisotropía Zonal.



Modificado de Díaz (2002).

Figura 3-18: Anisotropías complejas.



Modificado de Díaz (2002).

3.2.3 Predicción de valores

Es la técnica de predecir valores en puntos donde no tenemos muestra y se puede hacer aplicando técnicas de predicción por interpolación o usando métodos de simulaciones condicionales. Es decir que a partir de información conocida de una propiedad, medida en ciertas ubicaciones, predecir el valor en localizaciones en donde se desconoce o se carece de muestreo, usualmente en una malla 2D o 3D. En geoestadística el método de predicción espacial que se usa es el *kriging*.

- **Kriging.** Esta es una técnica de predicción espacial desarrollada por G. Matheron en los sesentas a partir de los trabajos de D. G. Krige quién fue pionero en el uso de la

correlación espacial para propósitos de predicción, *Matheron* le asigna el nombre de *kriging* en honor a Krige (Deutsch, 2002). El término *kriging* ha sido usado para definir el “mejor predictor lineal insesgado” de su acrónimo en inglés B.L.U.P. (Best Linear Unbiased Predictor). Las ventajas del kriging son:

- Es insesgado debido a que el valor esperado del error es cero.
- Garantiza la mínima varianza de la predicción porque reduce al mínimo la varianza del error de la predicción.

El *kriging* es una técnica de predicción que resuelve un sistema de ecuaciones lineales donde los valores del variograma son parámetros conocidos (media y varianza) y los pesos son parámetros desconocidos. El *kriging* hace asunciones tales como estacionaridad en los datos. La estacionaridad asume que la propiedad se comporta de la misma forma en todas las locaciones del volumen de estudio, es decir no hay tendencias inherentes. *Kriging* puede ser usado para crear tanto superficies 2D como propiedades 3D usando variogramas.

- **Predicción del Kriging.** Sobre una superficie 2D o una malla 3D la variable conocida (porosidad o permeabilidad) están en los pozos del campo donde fue medida. Todas las otras celdas de malla deben ser predichas así (ver Ecuación (3.48)):

$$Z(x_x) = \sum_i^n \lambda_i z(x_i) \quad (3.48)$$

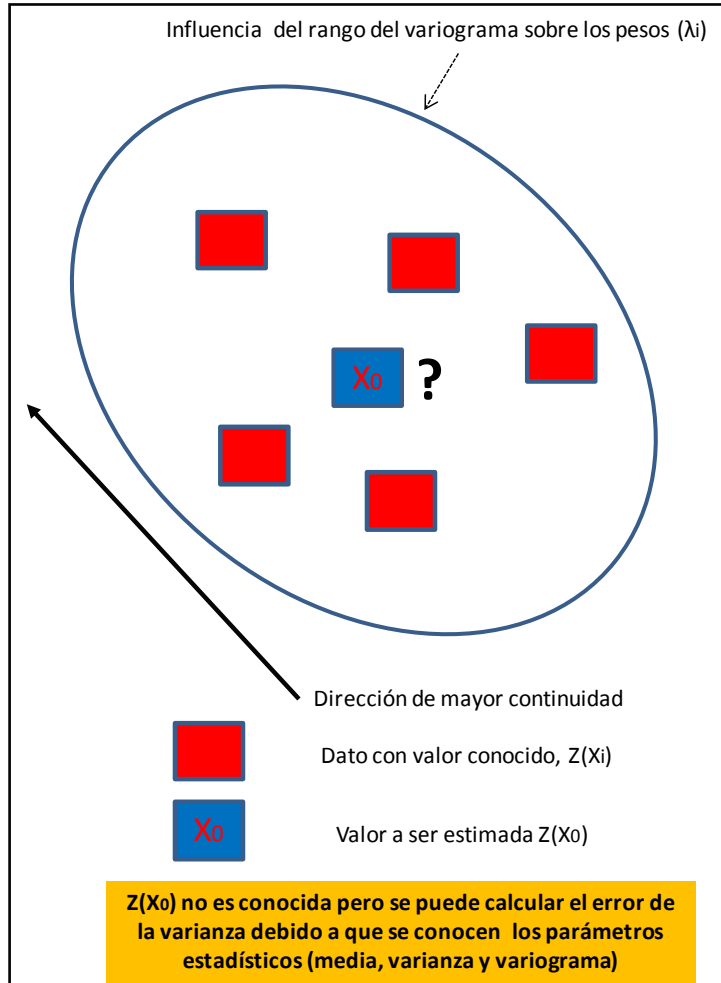
Donde:

- $Z(x_0)$ es el valor desconocido y es predicho como una sumatoria de pesos de los valores conocidos.
- λ_i Peso del *kriging* y es calculado desde el modelo de variograma.
- $Z(x_i)$ valor conocido como por ejemplo datos de pozos existentes.

En general el *kriging* es un algoritmo que promedia el movimiento de los pesos. La clave está en cómo calcular los pesos (λ_i) . La anterior ecuación corresponde a un *kriging* ordinario y combina valores conocidos, después aplica un factor de peso y de esta forma

una predicción puede ser hecha en una locación desconocida. *Kriging* usa predicciones lineales de peso (Ver figura 3-19).

Figura 3-19: Modelo de predicción por *Kriging*.



Modificado de Petrel (2008).

Error de la varianza: La pregunta es: qué tan buena es la predicción?. Esta predicción debe ser medida para escoger el mejor predictor (B.L.U.P.). El valor verdadero no es conocido exactamente aún, pero el error de la varianza puede ser calculado así (ver Ecuación (3.49)):

$$\text{Error de la varianza mínimo} = \text{Var}[Z^*(U) - Z(u)^2] \quad (3.49)$$

Entonces:

- La predicción que minimiza el error de la varianza por construcción es llamado *kriging*.
- *Kriging* usa el variograma para entender la variabilidad de los datos sobre una distancia.
- Este conocimiento permite a *kriging* calcular los pesos que minimizan el error de la varianza.

Tipos de *Kriging*: a continuación se hace un resumen de los principales tipos de *kriging*.

- ***Kriging simple*:** el valor medio es conocido y se define como (ver Ecuación (3.50)).

$$Z(x_0) = \sum_i^n \lambda_i Z(x_i) + [1 - \sum_i^n \lambda_i] m \quad (3.50)$$

Donde:

- La suma de los pesos (λ_i) puede ser menor que 1.
- M=media y se asume constante y conocida.
- Es bueno para una alta densidad de datos
- ***Kriging ordinario*:** predicción local de la media y se define como (ver Ecuación (3.51)):

$$Z(x_0) = \sum_i^n \lambda_i z(x_i) \quad (3.511)$$

Donde:

- La suma de los pesos (λ_i) es siempre 1.
- Este re-predice la media en cada locación (media local). Esta es constante pero desconocida.
- Este es más inestable que el *kriging simple* pero es útil cuando hay muchos pozos.
- Es confiable para puntos que muestran una tendencia.
- **Kriging Indicador:** Permite una distribución discreta de la propiedad honorando el histograma predefinido. Se usa para modelar facies en una malla 3D.

- **Geoestadística multivariada.** Según Armstrong (1998) la predicción conjunta de variables aleatorias regionalizadas, más comúnmente conocida como *Cokriging* (*Kriging* Conjunto) es el análogo natural del *kriging* de una función aleatoria. Mientras que el *kriging* utiliza la correlación espacial para determinar los coeficientes en el estimador lineal, el *Cokriging* utiliza la correlación espacial y la correlación entre funciones aleatorias al mismo tiempo.

Las aplicaciones que han recibido una mayor atención en la geoestadística de la industria del petróleo son los casos donde dos o más variables están muestreadas, pero una está menos muestreada que las otras o existe la presencia de errores de muestreo (Goovaerts, 1997).

Existe un número de dificultades prácticas, la más importante de todas es la ausencia de modelos estándar para las covarianzas cruzadas o covariogramas. Uno de los modelos más simples, el modelo lineal "estricto", no produce una varianza de predicción menor que el *kriging* separado excepto en los dos casos mencionados anteriormente (Caers, 2005).

- **Simulación geoestadística.** Una variable regionalizada $z(x)$ es interpretada como una realización de una cierta función aleatoria $Z(x)$. Esta función aleatoria $Z(x)$ puede ser caracterizada por una función de distribución de probabilidad y, en dependencia de su grado de estacionaridad, por una función de correlación (función de covarianzas o semivarianzas).

Según Deutsch (2002) la simulación estadística consiste en obtener nuevas realizaciones "artificiales" $Z_s(x)$ del campo aleatorio $Z(x)$ de manera tal que éstas reflejen las mismas propiedades estadísticas que se esperan que posea la función aleatoria $Z(x)$. Pero como por lo general no conocemos con precisión las propiedades estadísticas de $Z(x)$ y cuando más lo que podemos hacer es inferirlas a través de una sola realización o muestra de la función aleatoria. Lo que se hace es intentar obtener realizaciones simuladas $Z_s(x)$ que sean estadísticamente equivalentes a la muestra que se posee de la función aleatoria.

La equivalencia estadística en un sentido estricto significa que todas las realizaciones $Z_s(x)$ tengan la misma distribución de probabilidad de la función aleatoria $Z(x)$ que se simula, pero en la mayoría de los casos nos tenemos que conformar con que al menos posean los mismos momentos de primer y segundo orden que inferimos a partir de una muestra de $Z(x)$.

Según Jensen *et al.* (2003) la simulación es un proceso estocástico. En estadística, y específicamente en la teoría de la probabilidad, un proceso aleatorio o proceso estocástico es un concepto matemático que sirve para caracterizar; es una sucesión de variables aleatorias (estocásticas) que evolucionan en función de otra variable, generalmente el tiempo. Cada una de las variables aleatorias del proceso tiene su propia función de distribución de probabilidad y, entre ellas, pueden estar correlacionadas o no. Cada variable o conjunto de variables sometidas a influencias o impactos aleatorios constituye un proceso estocástico.

Los objetivos de la simulación son (Caers, 2005):

- Obtener nuevas realizaciones “artificiales” de una función aleatoria (FA) de manera tal que éstas posean las mismas propiedades estadísticas de la FA.
- Como no se conoce con precisión las propiedades estadísticas de las FA y cuando más lo que podemos hacer es inferirlas a través de una sola realización o muestra de la misma.
- Intenta obtener realizaciones simuladas que sean estadísticamente equivalentes a la muestra que se posee de la FA.
- La equivalencia estadística en un sentido estricto significa que todas las simulaciones tengan la misma distribución de probabilidad de la FA que se simula.
- Reproducir los momentos de primer y segundo orden que inferimos a partir de una muestra de la FA.

En algunos casos el deseo es quedarse solamente con aquellas simulaciones que en los puntos muestrales los valores simulados coinciden con los valores experimentales. A estas realizaciones de la FA se les conoce como: “Simulaciones Condicionales”. La Tabla 3-1 hace un resumen de los métodos de simulación.

Tabla 3-1: Métodos de simulación.

| METODO | CONDICIONADO | GAUSSIANO | MALLA REGULAR |
|----------------------|--------------|-----------|---------------|
| Indicador Secuencial | SI | NO | NO |
| Gaussiano Truncado | SI | SI | NO |
| Plurigaussiano | SI | SI | NO |
| Booleano | NO | NO | NO |
| Bandas Rotantes | NO | SI | NO |
| Secuencial Gaussiano | SI | SI | NO |
| Recocido Simulado | SI | NO | SI |

Otra clasificación se basa según la geometría:

- Basados en celdas (pixel)
- Basado en objetos (simulaciones booleanas)

Este trabajo se enfoca en dos tipos de simulación condicional. El primer método es el Secuencial Indicador (SIS) usado para poblar las mallas 3D con facies o tipos de roca y el Secuencial Gaussiano (SGS) usado para poblar las mallas 3D con propiedades petrofísicas (porosidad y permeabilidad).

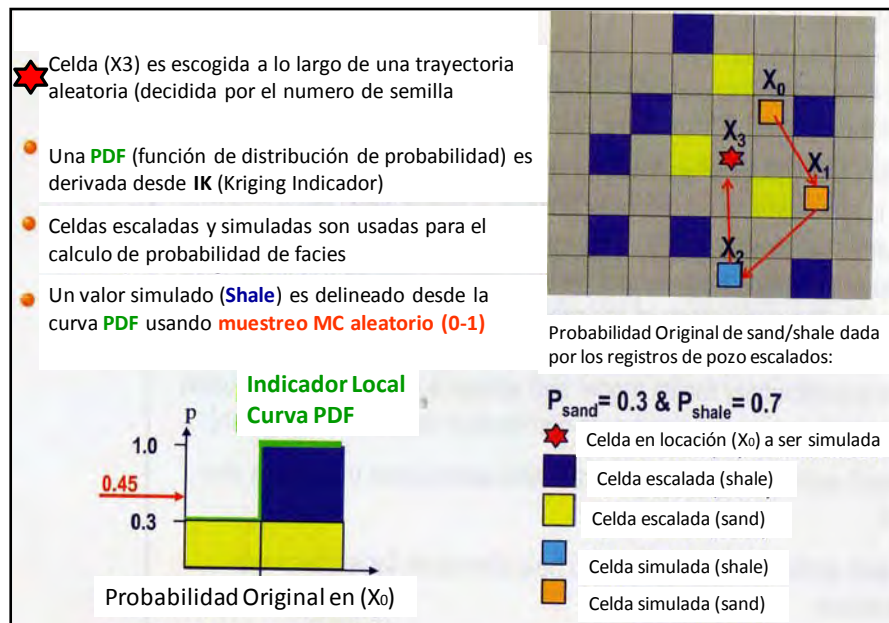
- **Método Secuencial Indicador (SIS).** Se usa para funciones aleatorias discretas. Es un algoritmo de modelamiento estocástico basado en píxeles que usa las celdas escaladas con datos reales como base para la fracción de los tipos de facies a ser modeladas. Usa como estimador el *kriging* indicador y el variograma restringe la distribución y la conexión de cada tipo de facies. Este algoritmo es ampliamente usado para modelar facies con formas no definidas o cuando se tienen pocos datos de entrada disponibles.

La idea es usar los datos de litología de los pozos que se encuentra dispersa (arena y lutita) para predecir la probabilidad de litologías lejos de los pozos. El procedimiento del método es que cada celda de la malla es visitada secuencialmente siguiendo una trayectoria aleatoria (dada por el número de semilla). Para cada nueva celda la distribución de probabilidad (PDF) arena/*shale* local es calculada usando los datos

escalados y las celdas ya simuladas como puntos de control. Una estimación de la probabilidad condicional arena/shale es estimada usando *kriging* indicador (IK).

La estimación está dada por el variograma y subsecuentemente el variograma de la combinación lineal de los pesos de datos desde las celdas escaladas. Un valor simulado (arena o *shale*= 0 o 1) es dibujado aleatoriamente desde la curva PDF local (Ver Figura 3-20). Finalmente el valor de la celda simulada será usado para la celda próxima para actualizar la PDF como un punto de control adicional junto con la data dura (celdas escaladas).

Figura 3-20: Teoría de Secuencial indicador (SIS).



Modificado de Petrel (2008).

- **Método Secuencial Gaussiano (SGS):** Se usa para funciones aleatorias continuas. Es un método estocástico basado sobre *kriging* ordinario pero es capaz de capturar valores extremos en un yacimiento heterogéneo. Este puede honrar los datos de entrada, distribuciones de entrada, variogramas y tendencias. Estos métodos requieren que la FDP multivariada de la función aleatoria a simular sea Gaussiana. Como La mayoría de los fenómenos de ciencias de la tierra no presentan histogramas simétricos y mucho menos gaussianos se necesita transformar a la FA de manera que resulte su FDP normal.

Principios de la simulación gaussiana: La Figura 3-21 ilustra los siguientes pasos:

- **Paso 1:** Gráfico de columnas de histograma de la propiedad (porosidad contra frecuencia). Graficar el histograma como una curva de frecuencia acumulada (CDF). Posteriormente se debe hacer la transformación Normal “score” de los datos, entonces cambia la curva CDF a una curva mas suavizada que la simulación gaussiana usará.
- **Paso 2.** *Kriging* (simple/ordinario) es usado para calcular el valor desconocido en un punto de la malla usando el dato vecino y el dato ya simulado. *Kriging* simple es usado para calcular la desviación estándar en el punto de la malla.
- **Paso 3.** Más valores simulados son suministrados durante la simulación. La curva de frecuencia acumulada (CDF) será actualizada continuamente condicionada a los nuevos valores simulados. Este proceso produce una curva de frecuencia acumulada calculada (CCDF) continuamente actualizada. Un valor simulado es delineado (aleatoriamente usando simulación de Monte Carlo) desde la CCDF.
- **Paso 4.** La trayectoria aleatoria decide cual celda visitará la próxima vez. Después de la simulación el dato será transformado de regreso desde la distribución normal a la distribución original.

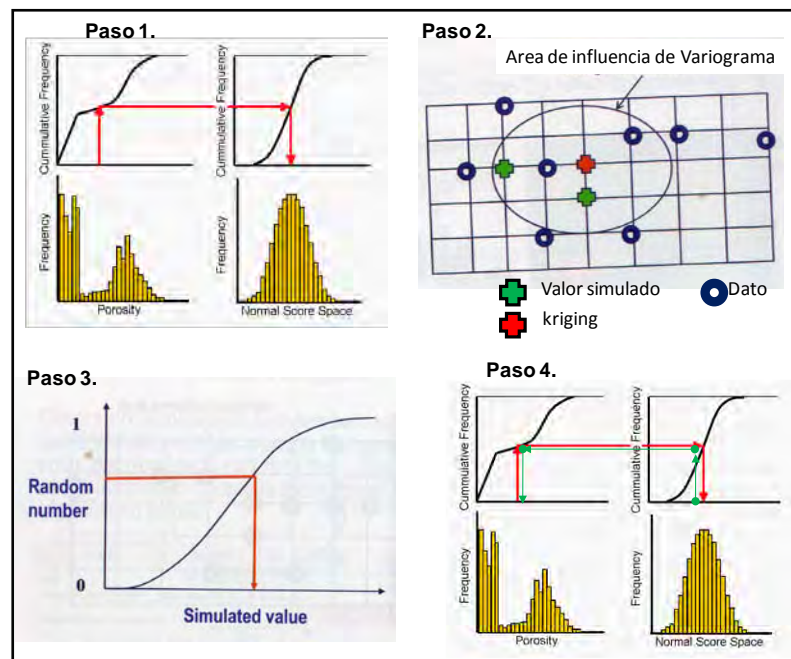
Predicción versus Simulación. El objetivo de una predicción es ofrecer en cada punto x un valor predicho. El predictor no tiene que reproducir la variabilidad espacial de los valores reales $Z(x)$. En el caso del *kriging* este produce un suavizado de las dispersiones reales, dicho de otra forma, se subestima la variabilidad de los valores reales (Davis, 2003).

Mientras que el objetivo de la simulación es reproducir las propiedades estadísticas de la función aleatoria $Z(x)$. La simulación posee los mismos momentos experimentales (media, covarianza o variograma, así como el histograma) que los valores reales los cuales caracterizan las principales propiedades estadísticas de $Z(x)$. En particular, en las simulaciones condicionadas en cada punto x el valor simulado no es el mejor estimador posible de $Z(x)$. Por el contrario, se puede mostrar que la varianza del error de la

estimación de $Z(x)$ mediante simulación condicional es exactamente el doble de la varianza del error usando *kriging* (Jensen *et al.*, 2003)

La Tabla 3-2 hace una comparación entre los métodos de simulación y *kriging*. Es importante anotar como la simulación honora los histogramas de distribución de la función mientras que la predicción no.

Figura 3-21: Pasos Simulación Gaussiana (SGS).



Modificado de Petrel (2008).

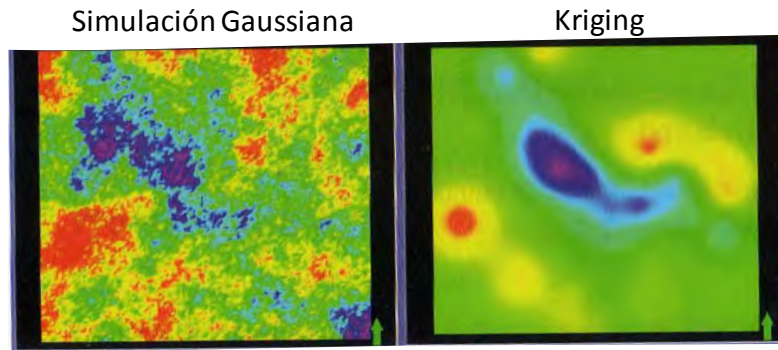
Tabla 3-2: Comparación simulación y *kriging*.

| | SIMULACIÓN | KRIGING |
|-------------|--------------------------------------|------------------------------|
| SALIDA | Realizaciones equiprobales multiples | Un modelo determinístico |
| PROPIEDADES | Honora los pozos | Honora los pozos |
| | Honora los histogramas | Honora los variogramas |
| | Honora los variogramas | |
| IMAGEN | Ruidosa | Suavizada lejos de los pozos |
| | Alguna variabilidad por todas partes | |
| USO | Simulación de flujo | Mapeo |
| | Calculo de incertidumbre | Volumétricos |

Modificado de Petrel (2008).

La Figura 3-22 muestra los resultados de los modelos tanto por simulación (SGS) como por *kriging*. Se puede observar como el *kriging* suaviza los modelos mientras que con la simulación puede detectar pequeñas variabilidades en el yacimiento.

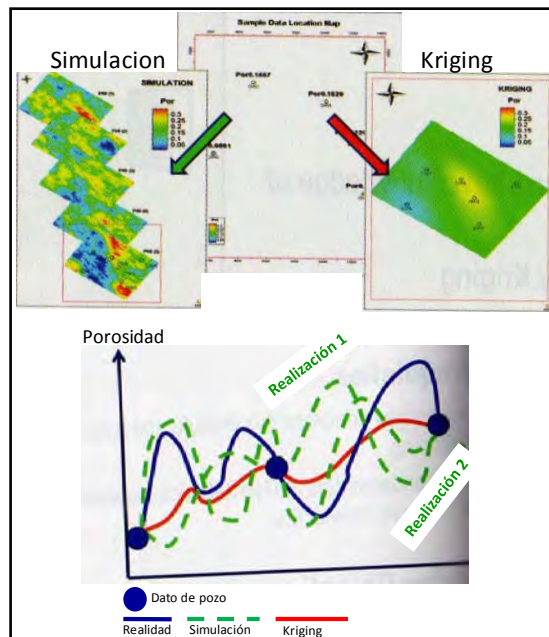
Figura 3-22: Comparación resultados simulación contra *kriging*.



Modificado de Petrel (2008).

Otro punto importante es que con simulación se pueden hacer múltiples realizaciones que pueden ser usadas para hacer análisis de incertidumbre mientras que *kriging* no permite hacerlas (Ver Figura 3-23).

Figura 3-23: Ventajas de la simulación.



Modificado de Petrel (2008).

