CAPÍTULO 2

2. MARCO TEÓRICO CONCEPTUAL DE LA GEOESTADÍSTICA.

2.1. Introducción.

En el presente capítulo, se detalla como se inició la Geoestadística, las principales teorías que componen esta ciencia, los personajes que desarrollaron estas teorías y cuales fueron los diferentes campos en los que empezó a ser aplicada.

Además, se presenta la definición teórica de la Geoestadística, se indican cuales son los pasos principales que se deben seguir para desarrollar un estudio Geoestadístico, y las teorías principales en las cuales se ha desarrollado y se desarrolla esta ciencia.

2.2. Reseña Histórica.

El uso de la Geoestadística surge a partir de la década de los años 50 con los estudios realizados por D.G. Krige y por otros Ingenieros mineros en las minas de oro de Rand, África del Sur, desarrollaron un procedimiento ponderado de estimación, con determinados coeficientes correctores empíricos.

El desarrollo posterior es atribuido, en paralelo, a Matèrn y Matheron, quien estuvo ligado directamente a la Escuela de Minas de París, desarrollo la teoría de las variables regionalizadas, que provee una buena base para los recursos naturales, está teoría incluye una técnica de estimación óptima llamada Kriging, esta técnica se hizo conocida debido a que Matheron hizo las publicaciones en textos en Inglés y campos como la Hidrogeología empezaron a implementar la Geoestadística.

En la ciencia del suelo, el reconocimiento de la variación espacial de propiedades del suelo, es probablemente tan viejo como su clasificación. Este enfoque de caracterizar la variabilidad de los suelos es principalmente cualitativo.

En los años setenta, los investigadores empezaron los estudios de las propiedades de suelo de una manera más cuantitativa (p. ej. Rogowski, 1972; Nielsen et al, 1973; Biggar y Nielsen, 1976), la mayoría de ellos confió en métodos estadísticos clásicos (p. ej. La función de la distribución de frecuencias con momentos asociados y coeficientes de variación; análisis de varianza para medir cambios sobre la distancia, etc.). Una suposición implícita en estos análisis es; que las observaciones de una variable del suelo son independientes con respecto a su ubicación en el campo.

Sin embargo, solamente el conocimiento de la distribución de frecuencias de las observaciones provee información sobre la variabilidad de las observaciones con respecto a las coordenadas del área donde se ha hecho un muestreo.

La primera publicación que aplicó Geoestadística apareció en 1978 (Campbell, 1978). En 1980, Webster y su colegas iniciaron una serie de publicaciones llamadas “Interpolación Optima y mapeo isarithmico de propiedades del suelo”, en la que Kriging era por primera vez usado en la ciencia del suelo (Burguess & Webster, 1980ª,b; Webster & Burguess,1980 Burguess et al,

1981; McBratney & Webster, 1983ª, webster & Oliver, 1989). Samra en 1989 realizó estimación de alturas en plantaciones de Melia azedarach, Mandallaz (Mandallaz, 1993), que comparó varios métodos de kriging en la estimación de densidad y área basal. Höck (Höck et al, 1994), combinó el uso de Sistemas de Información Geográficos con técnicas geoestadísticas en la estimación de índices de sitio de Pinus radiata en Nueva Zelandia.

En Chile, Corvalán (Corvalán et al, 1998) analizaron la aplicabilidad de estas técnicas en la estimación de variables dasométricas en bosques de Pinus radiata, concluyendo que constituyen un aporte importante en la planificación del manejo forestal.

La palabra Geoestadística, denota el estudio estadístico de fenómenos naturales, tiene varias interpretaciones como la de Matheon, que la define como: “la aplicación del formalismo de las funciones aleatorias al reconocimiento y estimación de fenómenos naturales” (Journel & Huijbregts, 1978). El término Geoestadística nos indica, como algunos interpretan incorrectamente, “estadísticas en la geología”. Una más reciente

(y orientado al SIG) interpretación de Geoestadística, podría ser: "*las estadísticas involucradas con datos geográficos*", también llamada "estadísticas espaciales".

2.3. Geoestadística.

La Geoestadística implica el análisis y la estimación de fenómenos espaciales o temporales, tales como: calidades de metal, porosidades, concentraciones de contaminantes.

La palabra Geoestadística es normalmente asociada con la Geología, desde que esta ciencia tiene sus orígenes en la minería.

Hoy en día, la Geoestadística es un nombre asociado con una clase de técnicas, para analizar y predecir los valores de una variable que está distribuida en espacio o tiempo. Se asumen tales valores implícitamente, para ser puestos en correlación entre sí, y el estudio de semejante correlación normalmente se llama un "análisis estructural" o un "Variograma". Después del análisis estructural, se hacen estimaciones a las situaciones de

los sectores no muestreados usando la técnica de interpolación "Kriging" .

La Geoestadística, tiene como objetivo el caracterizar e interpretar el comportamiento de los datos que están distribuidos espacialmente “variables regionalizadas”.

Los pasos principales de un estudio Geoestadístico son:

- Análisis de los datos exploratorios.
- Análisis Estructural (Cálculo y modelación de los Variogramas).
- Estimaciones (Kriging o simulaciones).

2.4. Análisis de Datos Espaciales.

Alguna veces, se ha señalado a la Geoestadística como un juego de técnicas, para los análisis espaciales de datos Geoestadísticos, o datos con un índice espacial continúo.

Se la ha incluido en la familia de tipos de datos espaciales, datos de celosía (datos con una colección contable de sitios espaciales, p. ej. la distribución de mortalidades infantiles en diferentes pueblos) y diseños de datos de puntos espaciales (datos donde ambas localidades y magnitudes son aleatorias p. ej. una realización de sedimentos de cuerpos geológicos en el espacio).

Los datos espaciales se pueden clasificar en:

- Puntos.
- Líneas.

  Líneas aisladas (pista de aeropuerto), Líneas con Ramificaciones (río y sus afluentes), Redes (carreteras o líneas ferroviarias.)
- Regiones.

  Regiones aisladas (lago), Regiones adyacentes (países) o Regiones anidadas (Municipios contenidos en Estados).

2.5. Variable Regionalizada.

Dado que *x*, es un punto en el espacio R^3 y *z*(*x*) es el valor de la función para el punto *x,* en el que se está interesado. Está función es llamada *"Variable Regionalizada"*, p. ej. puede ser la consistencia de un deposito en un punto, o la gravedad especifica de los minerales. Tal función usualmente es altamente variable y no-continua y no puede ser estudiada correctamente. El estudio se lo realiza, examinando sus incrementos.

La idea básica de la teoría es considerar a tal función *z(x), do*nde x es un punto o un vector en R^n , como una realización de una función aleatoria Z(X). Solo se tiene una realización de aquella

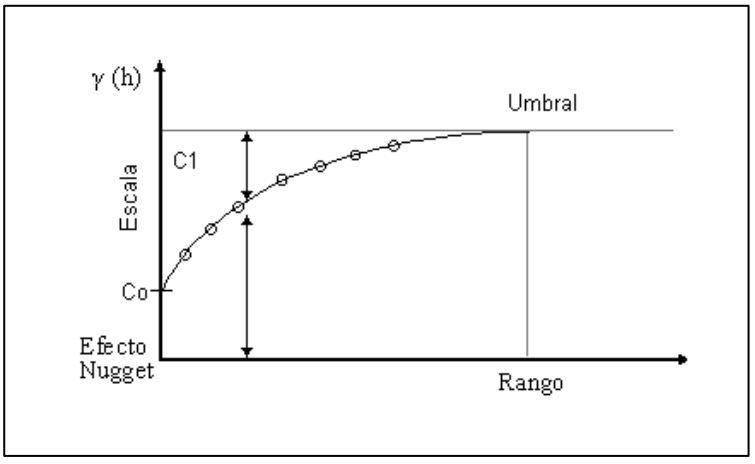
función aleatoria, el fin es encontrar las características de la función aleatoria Z(X) para hacer las estimaciones de posibles puntos desconocidos.

2.6. Variograma.

Es un estimador de la varianza poblacional, por lo tanto la población debe tener una tendencia de estacionariedad; el variograma esta relacionado con una dirección y distancia (*h*), es la principal herramienta básica, que da soporte a las técnicas de Kriging, permite representar cuantitativamente la variación de un fenómeno regionalizado en el espacio. El variograma modela como dos valores en el espacio o tiempo se ponen en correlación.

FIGURA 2.6.

GRÁFICO DE VARIOGRAMA.



El Variograma está conformado por los siguientes elementos:

Efecto Nugget (Co): Generalmente el variograma no tiende a cero como lo hace la distancia *h*, el *Efecto Nugget* indica la discontinuidad del semivariograma para distancias que sean menores, que la menor distancia que se da entre los puntos muestrales. Parte de esta discontinuidad puede también ser dividida en errores de medición, una pobre precisión analítica, o más a menudo por mineralización altamente errática a baja escala.

Umbral (Sill = C1 + Co): Es el valor alcanzado por el variograma en el R*ango*. El Sill de un variograma, puede ser denotado como la varianza de la muestra.

Escala (C1): Es el valor tal que; C1 + Co = Sill.

Rango (a): Es la distancia dentro de la cual, las muestras están correlacionadas espacialmente.

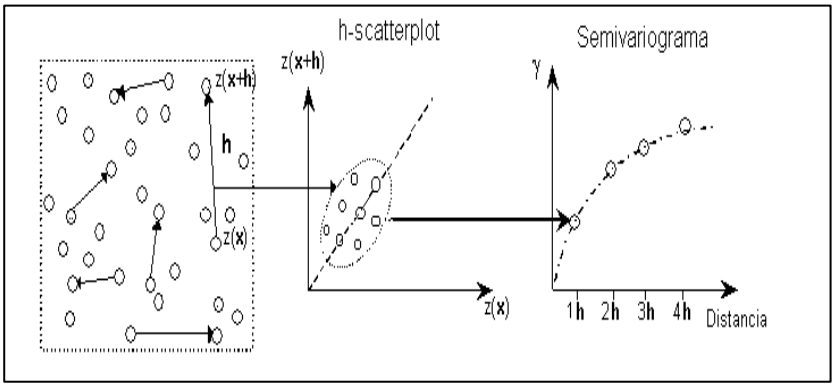
2.6.1. Cálculo del Variograma.

Para calcular el variograma se debe definir un incremento lag, tomar todos los pares de puntos que tengan esa

distancia entre ellos, calcular el cuadrado de las diferencias para cada par de puntos, resumir todas las diferencias y luego dividir para dos veces el numero de pares de puntos que tienen esta distancia entre sí, este resultado es el valor del variograma. Hacer lo mismo para otros incrementos lag.

FIGURA 2.6.1.

CONSTRUCCIÓN DEL VARIOGRAMA DESDE DATOS MUESTRALES.



FUENTE: (AI-GEOESTATS).

El variograma es definido como:

$$2\,?(h) = \text{Var.}\,[Z(x+h) - Z(x)] = E[\{Z(x+h) - Z(x)\}^2]$$

Puede ser estimado a partir de una muestra por:

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

Muchos autores definen al variograma de una forma distinta a la ecuación descrita arriba, considerando que usualmente al variograma también se lo conoce como semivariograma, el cual es definido como:

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

Puede ser estimado a partir de una muestra por:

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

2.6.2. Modelos Teóricos del Variograma.

Son varios los modelos básicos de los Variogramas, que son capaces de explicar los diferentes comportamientos que pueden presentar las variables regionalizadas; entre los modelos más usuales que se dan tenemos:

- Modelo Esférico.

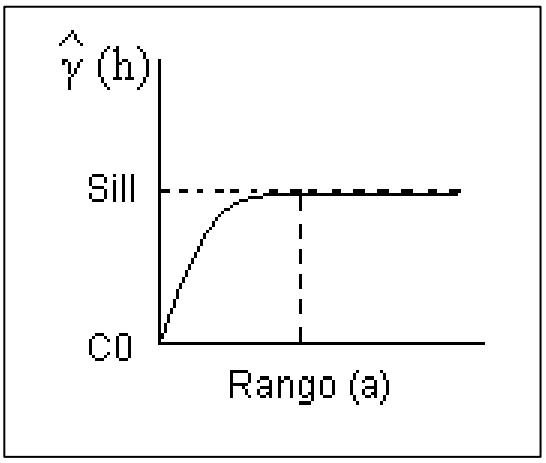
  Está representado por la ecuación siguiente:

$$\hat{\gamma}(h) = \begin{cases} 0 & ;|h| = 0 \\ C0 + C1\left[\frac{3}{2}\left(\frac{|h|}{a}\right) - \frac{1}{2}\left(\frac{|h|}{a}\right)^3\right] & ; 0 < h < a \\ C0 + C1 & ;|h| > a \end{cases}$$

Este modelo, presenta una tendencia creciente curvilínea, hasta que llega al S*ill*, para tomar una tendencia lineal. El modelo Esférico alcanza el valor S*ill*, en la distancia *a* (*rango* actual).

FIGURA 2.6.2.1.

GRÁFICO MODELO ESFÉRICO.



- Modelo Exponencial:

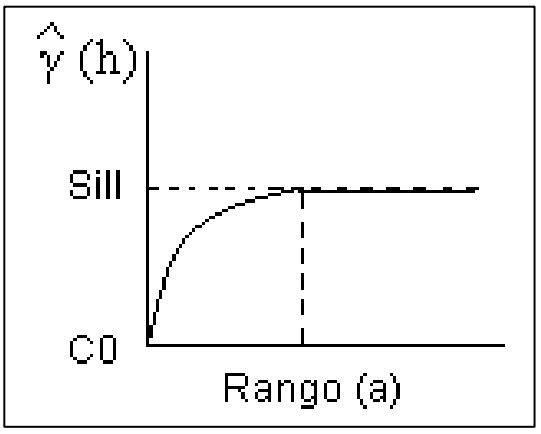
Está representado por la ecuación siguiente:

$$\hat{\gamma}(h) = \begin{cases} 0 & ;|h| = 0 \\ C0 + C1\left[1 - \exp\left(-\frac{|h|}{a}\right)\right] & ; h \neq 0 \end{cases}$$

El modelo Exponencial, alcanza el valor S*ill* asintóticamente, el *rango (a)* es la distancia a la cual el valor del modelo, es el 95 % del S*ill*.

FIGURA 2.6.2.2.

GRÁFICO MODELO EXPONENCIAL.



- Modelo Gaussiano:

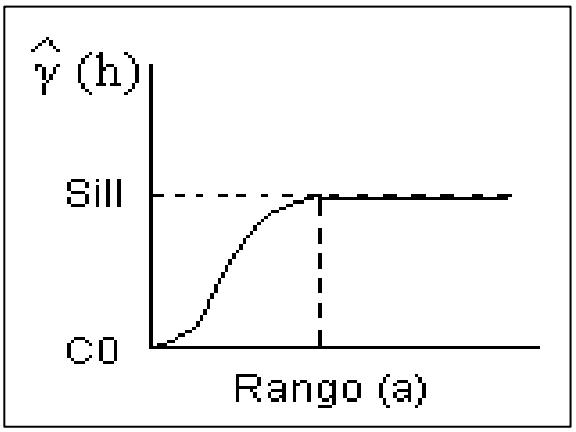
Está representado por la ecuación siguiente:

$$\hat{\gamma}(h) = \begin{cases} 0 & ;|h| = 0 \\ C0 + C1\left[1 - \exp\left(-\left(\frac{|h|}{a}\right)^2\right)\right] & ; h \neq 0 \end{cases}$$

El modelo Gaussiano, alcanza el valor S*ill* asintóticamente, el *rango (a)* es la distancia a la cual el valor del modelo, es el 95 % del S*ill*.

FIGURA 2.6.2.3.

GRÁFICO MODELO GAUSSIANO.



- Modelo Lineal.

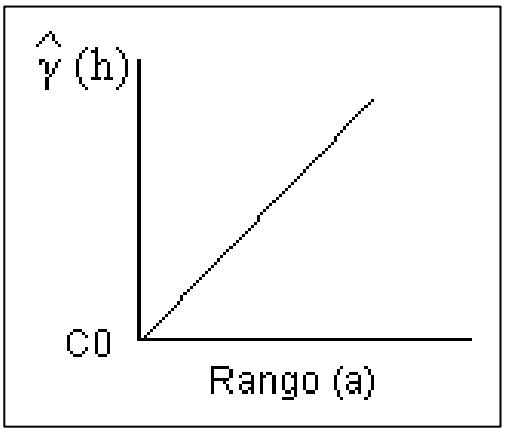
Está representado por la ecuación siguiente:

$$\hat{\gamma}(h) = A(h) + B$$

No posee valor S*ill*, continua aumentando mientras la *distancia (h)* aumenta, es utilizado para modelar fenómenos que presentan capacidad infinita de dispersión.

FIGURA 2.6.2.4.

GRÁFICO MODELO LINEAL.



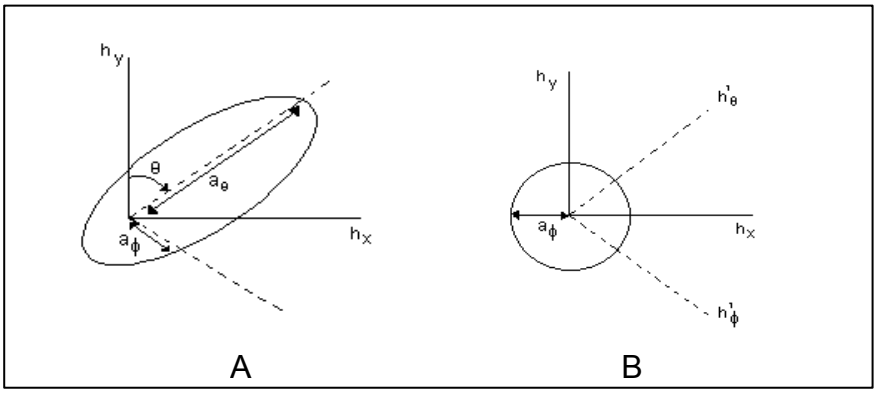
2.7. Anisotropía.

Si al calcular Variogramas en diferentes direcciones, los valores del *rango (a*) presentan cambios, esto nos indica que se está dando una anisotropía geométrica, en cambio si la diferencia muestra cambios en el valor del *Sill* en diferentes direcciones, entonces hay una anisotropía zonal. Algunos Variogramas son combinaciones de anisotropía geométrico y zonal.

Se debe calcular varios Variogramas en diferentes direcciones, y hacer uso de un diagrama de rosa, para descubrir la existencia de anisotropía. La proporción de la anisotropía es la proporción entre el rango más pequeño y el rango más grande, utilizando un proporción de uno, indica un variograma isotrópico, es decir el mismo variograma en todas las direcciones (círculo isotrópico).

FIGURA 2.7.

A. ROSA ELÍPTICA B. CÍRCULO ISOTRÓPICO.



FUENTE: GEOSTATISTICS FOR NATURAL RESOURCES EVALUATION.

2.8. Kriging.

El Kriging es conocido como el método interpolador Geoestadístico, es un estimador lineal insesgado, presenta dos propiedades básicas que son: Hacer que la suma de errores tienda a cero, y que el cuadrado de las desviaciones sea mínimo.

Tiene como objetivo estimar el valor de la variable Z, para un punto x_0 que no ha sido considerado anteriormente, realiza una suma ponderada sobre todos los sectores que conforman la zona de estudio de interés, tomando los vecinos más cercanos al punto de interés x_0.

$$\hat{z}(x_0) = \lambda_1 z(x_1) + \lambda_2 z(x_2) + \ldots + \lambda_n z(x_n)$$

$$\hat{z}(x_0) = \sum \lambda_i * z(x_i)$$

El proceso del Kriging es asignar pesos a los vecinos más cercanos, considerados para la estimación, la diferencia del Kriging con otros métodos de interpolación, es que utiliza un método semejante a la interpolación por media móvil ponderada, a diferencia que los pesos son asignados a partir de un análisis espacial, basados en el semivariograma experimental.

Un método de interpolación será exacto cuando, pase por los puntos muestrales, lo más cercano posible a ellos.

Es importante que un modelo para semivariograma que ha sido ajustado, represente una tendencia a los modelos antes

descritos, para que las estimaciones obtenidas por medio del Kriging sean más exactas y más confiables.

Las estimaciones mediante el método Kriging pueden ser, por punto o por bloque.

2.8.1. Métodos de Estimación del Kriging.

Existen dos métodos de Kriging para realizar las estimaciones.

2.8.1.1. Kriging Ordinario.

Es el método más apropiado, para situaciones medioambientales, este método asume que las medias locales, no están necesariamente relacionadas lo más cercanamente a la media poblacional, por lo cual solo usa las muestras en la vecindad local para la estimación.

2.8.1.2. Kriging Simple.

Asume que las medias locales son relativamente constantes e iguales a la media poblacional, la cual

es conocida. La media poblacional es usada como un factor en cada estimación local, a lo largo con las muestras en las vecindad local.

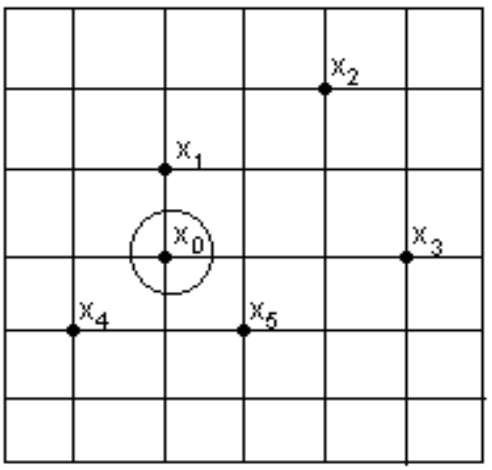
Hay dos tipos de Kriging, el de Punto y el de Bloque, lo cuales generan unas cuadrículas de interpolación.

2.8.1.3. Kriging de Punto.

Estima los valores de los puntos en los nodos de las cuadrículas.

FIGURA 2.8.2.

REPRESENTACIÓN TÉCNICA PUNTO KRIGING.



2.8.1.4. Kriging de Bloque.

Estima el valor promedio de los bloques rectangulares, que están centrados en los nodos de las cuadrículas, los bloques son le tamaño y forma de las celdas de las cuadrículas, este tipo de Kriging no resulta ser un buen interpolador, ya que no estima el valor de un punto.

2.8.2. Efecto del Rango en las Estimaciones.

Un valor grande para el *rango (a)*, significa un comportamiento más continuo. Las estimaciones dan como resultado mapas bastante lisos para la variable de interés.

2.8.3 Efecto del modelo en las Estimaciones.

Considerando la forma que presenta el Variograma en los primeros lags, un modelo Gaussiano es más continuo que un modelo Esférico con un mismo efecto en *6h*, para la variable mejor correlacionada, se muestran los mapas con más suavización.

2.8.4. Efecto del Sill en las Estimaciones.

El cambiar el valor de *Sill*, no cambia los valores de las estimaciones, por lo que los mapas de estimaciones seguirán siendo los mismo, afecta a la variación de las estimaciones, un *sill* más alto indica, mayor variación en las estimaciones.

2.8.5. Error de Estimación.

Dado que no son estrictamente equivalentes, hay un error de estimación involucrado en los procedimientos. La aceptación de un método de estimación será dado por la magnitud de los errores involucrados, el mejor método a considerar de be ser el que de los errores más pequeños, considerando todos los bloques o puntos en la estimación.

2.8.6. La Varianza en la Estimación.

La varianza indica la dispersión que presentan los valores estimados con respecto a los valores reales.

El Kriging no solo provee una estimación de mínimos cuadrados, también está ligado a la varianza del error.

La varianza del error es:

- Dependiente en el modelo de la covarianza.

  La precisión de la estimación podría depender de la complejidad de la variabilidad espacial de z, modelado por la covarianza.

- Dependiente en la configuración de los datos.

  La localización de los datos y sus distancias entre sí, son estimados.

- Independiente de los valores de los datos.

  Para un modelo de covarianza dado, la configuración de dos datos idénticos podría producir la misma varianza Kriging, sin importar que los datos estuvieren.

El mapa de varianzas, indica más incertidumbre en los sectores del área de estudio, en el que los datos están más dispersos del área de estudio .