

CAPITULO 2

MARCO TEÓRICO CONCEPTUAL DE LA GEOESTADÍSTICA

INTRODUCCION

En el campo de las ciencias de la tierra es muy común encontrar variables distribuidas espacialmente, para el estudio de estas variables son usados diversos procedimientos geoestadísticos.

En este capítulo se muestra los inicios de la Geoestadística, las principales teorías en las cuales se ha desarrollado esta ciencia, los personajes más importantes que desarrollaron dichas teorías y cuales fueron los diferentes campos en los que empezó a ser aplicada.

Además, se detalla la definición teórica de la Geoestadística, y se indican cuales son los pasos a seguir para desarrollar un estudio de este tipo.

2. MARCO TEÓRICO CONCEPTUAL DE LA GEOESTADÍSTICA

2.1 RESEÑA HISTÓRICA

La Geoestadística surge en la década del 50 a partir de estudios previos de H. Sichel, D.G. Krige, y B. Matern, con el objetivo de obtener una mayor precisión en la estimación de las reservas minerales. Su punto de partida es el análisis de los fenómenos distribuidos en el espacio, por ejemplo la mineralización. La evolución de la Geoestadística se ha dividido en algunas generaciones [Matheron, y Kleingeld,1987], las cuales son:

La Geoestadística Lineal 1945-1965, la ciencia, ahora llamada Geoestadística, tuvo su comienzo en los trabajos de H.S. Sichel en 1947 y 1949, en la aplicación de la distribución lognormal en minas de oro, seguido por la contribución de D.G. Krige en la aplicación del análisis de regresión entre muestras, los cuales fijaron la base de la Geoestadística Lineal, además de la introducción de la teoría de funciones aleatorias por B. Matern en el estudio de la variación espacial. La generación de la Geoestadística Lineal culmina en el trabajo de G. Matheron en su tesis doctoral de 1965 titulada: "Las variables regionalizadas y su estimación".

La Geoestadística No Lineal 1966-1974, en Mayo de 1968, fue creado en Fontainebleau el Centro de Geoestadística, según Matheron (1987).

Importantes contribuciones fueron hechas en este período por los autores M. David, A.G. Journel, Ch. J. Hiujbregts, P. Delfiner, P. Chauvet y J.P. Chiles y muchos otros autores.

La tercera generación 1974-1987, estuvo dedicada a resolver problemas más complejos, Matheron introduce el modelo de Kriging, la primera aplicación práctica del Kriging fue desarrollada por A. Maréchal en 1975. Durante este período, fue también desarrollado por A.G.Jounel el Kriging de Indicadores y una forma especial de Kriging Disyuntivo.

En la actualidad, los métodos Geoestadísticos han sido extendidos a los más diversos campos de la Ciencias de la Tierra. El desarrollo de la informática moderna ha propiciado condiciones para su divulgación y su aplicación a un grupo cada vez mayor de problemas, pueden encontrarse en el mercado informático programas profesionales que ofrecen opciones para la aplicación de estas técnicas.

Hoy por hoy, los dominios de aplicación de la Geoestadística son amplios, ya que a partir del estudio de la variabilidad de sus variables, se obtienen elementos para predecir sus características. Además de la minería que es el campo que le dio origen, se pueden mencionar otras áreas de estudio tales como: el petróleo, (estimaciones de sismos), la pesca (estimación provisiones de peces, profundidad, temperatura del agua), la salud (distribución espacial de enfermedades) ingeniería civil (obras de grandes

dimensiones que exigen del conocimiento de la variabilidad espacial del terreno), las finanzas (relación entre el análisis técnico con el análisis económico), la cartografía, la hidrogeología, el medio ambiente, entre otros.

En su surgimiento y desarrollo hay tres elementos importantes que destacar: la consolidación de la Geoestadística con los trabajos de G. Matheron en 1965, D.G. Krige, el establecimiento de la escuela de Fontainebleau, y finalmente el desarrollo de la Geoestadística asociada con la informática.

2.2 GEOESTADISTICA

Geoestadística es una ciencia que comprende un conjunto de herramientas y técnicas que sirven para analizar y predecir los valores de una variable que está distribuida en el espacio de una forma continua, es conocida también como estadística espacial. La aplicación de la geoestadística está orientada a los SIG “Sistemas de información Geográfica” (Conjunto de programas informáticos que sirven para captar, almacenar, recuperar, transformar, Mostrar y analizar diversos tipos de datos espaciales), por esto se la puede definir también como estadística relacionada con datos geográficos.

Para realiza un trabajo estadístico es necesario llevar a cabo los pasos siguientes:

1. Análisis Exploratorio de los datos:

Básicamente es una etapa de aplicación de la estadística, se estudian los datos puros sin tener en cuenta su distribución geográfica en el espacio. Se verifica la normalidad de los datos y se aplica la transformada de los datos si fuera necesario.

2. Análisis Estructural

Se estudia la continuidad espacial, es indispensable conocer la distribución geográfica de las variables para realizar esta etapa del estudio. Se calculan los variogramas que expliquen la variabilidad espacial y se ajustan a un variograma teórico.

3. Kriging

Estima la variable en los puntos no muestrales, considerando la estructura de correlación espacial seleccionada e integrando la información obtenida de forma directa en los puntos muestrales.

2.3 VARIABLE REGIONALIZADA

Es una función que describe un fenómeno natural geográficamente distribuido. El dato inicial obtenido es conocido como variable aleatoria, además de su valor este dato está distribuido en el espacio y tiene coordenadas geográficas, por esta característica es conocido como variable regionalizada.

Variable Regionalizada, matemáticamente es una función $z(x)$ donde x es un punto o un vector en R^n , representando a una función aleatoria $Z(x)$. La finalidad es encontrar las características de la función $Z(x)$ para hacer las estimaciones de posibles puntos desconocidos.

Variable Regionalizada presenta una estructura espacial de correlación (autocorrelación). Toda variable regionalizada está formada por dos componentes, una estructurada (marca la relación entre el valor de la variable y puntos próximos) y un segunda componente aleatoria (representa un valor desconocido).

2.4 VARIOGRAMA

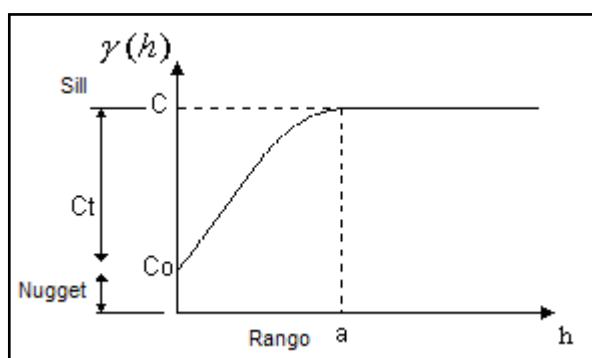
El variograma es una herramienta que permite analizar el comportamiento espacial de una variable sobre una zona dada y modela como dos valores en el espacio se ponen en correlación. Es un estimador de la varianza poblacional, por lo tanto debe tener una tendencia de estacionaridad y es un soporte para las técnicas del Kriging ya que permite representar cuantitativamente la variación de un fenómeno regionalizado en el espacio. El variograma está relacionado con la dirección y la distancia (h).

El variograma se ve limitado porque es un estadístico de dos puntos y además porque es extremadamente sensible a valores extremos.

El variograma está formado por los siguientes elementos:

Figura 2.1

Elementos del Variograma



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

2.4.1 EFECTO NUGGET (Co)

Por lo general, el variograma no tiende a cero como lo hace la distancia h , el efecto nugget es una discontinuidad de salto en el origen conocido también como efecto Pepita, representa la discontinuidad en el semivariograma para distancias que sean menores que la menor distancia dada entre los puntos muestrales. Esta discontinuidad se puede dar también debido a errores en la medición o a una pobre precisión analítica.

2.4.2 SILL (C)

El sill conocido también como La “Meseta” es el valor máximo que alcanza el semivariograma cuando la variable es estacionaria. Teóricamente, la meseta coincide con el valor de la varianza y por tanto un buen estimador de la misma será la varianza experimental de los datos.

2.4.3 RANGO (a)

Conocido también como Alcance, el Rango es la distancia a la cual el variograma se estabiliza y las muestras se relacionan espacialmente.

2.4.4 ESCALA (Ct)

Es el valor tal que $Co + C = Sill$

2.5 CALCULO DEL VARIOGRAMA

El variograma es una función que se calcula mediante el análisis de los datos distribuidos en el área de estudio. El cálculo del variograma depende de la distancia entre los puntos establecida previamente conocido como lag, el proceso consiste en tomar todos los pares de puntos que tengan la distancia lag entre ellos y calcular el cuadrado de las diferencias para cada par de puntos, luego hallamos la sumatoria de todas las diferencias y lo dividimos para 2 veces la cantidad de pares que tienen esa distancia lag entre sí. Esta operación nos da como resultado el valor de la semivarianza.

La semivarianza está definida como una esperanza matemática del cuadrado de la diferencia entre pares de puntos separados por la distancia h

$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]^2\} = \text{Var}[Z(\mathbf{x}) - Z(\mathbf{x} + \mathbf{h})]$$

La semivarianza puede ser estimado a partir de una muestra por:

$$2\gamma(h) = \frac{1}{Np(h)} \sum_{i=1}^{Np(h)} [Z(x_i) - Z(x_i + h)]^2$$

donde: $N_p(h)$: es el número de pares a la distancia h .

h : es el incremento

$Z(x_i)$: son los valores experimentales.

x_i ; localizaciones donde son medidos los valores $z(x_i)$.

El gráfico de la semivarianza contra la distancia, es conocido como VARIOGRAMA.

2.6 MODELOS TEÓRICOS DEL VARIOGRAMA

Los modelos teóricos se ajustan a un modelo de variograma experimental y son capaces de explicar los diferentes comportamientos que pueden tener las variables en el espacio. Los modelos más comunes son El Efecto Pepita Puro, Gaussiano, Potencia y Exponencial.

Las dos características más importantes en el modelado de Variogramas son:

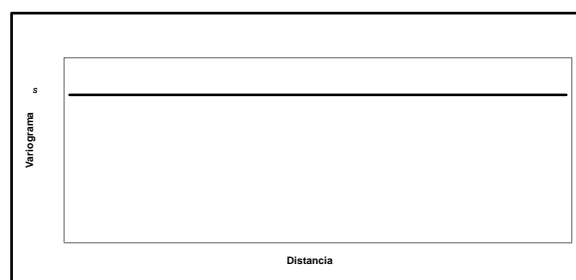
- 1.- Su comportamiento en el origen, el cual puede ser lineal, parabólico y con efecto de pepita
- 2.- La presencia o ausencia de meseta.

2.6.1 EFECTO PEPITA PURO

Figura 2.2

Gráfico del Efecto Pepita

$$\gamma(h) = \begin{cases} 0 & \text{si } |h| = 0 \\ s & \text{si } |h| \neq 0 \end{cases}$$



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

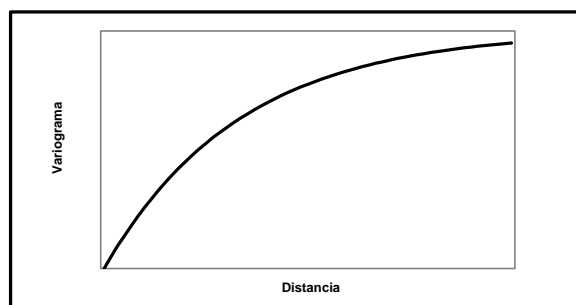
El modelo Efecto Pepita representa un fenómeno completamente aleatorio, conocido también como ruido blanco, en el cual no hay correlación espacial en las muestras independiente de cual sea la distancia h que las separe. S representa el valor del sill.

2.6.2 MODELO EXPONENCIAL

Figura 2.3

Gráfico del Modelo Exponencial

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



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

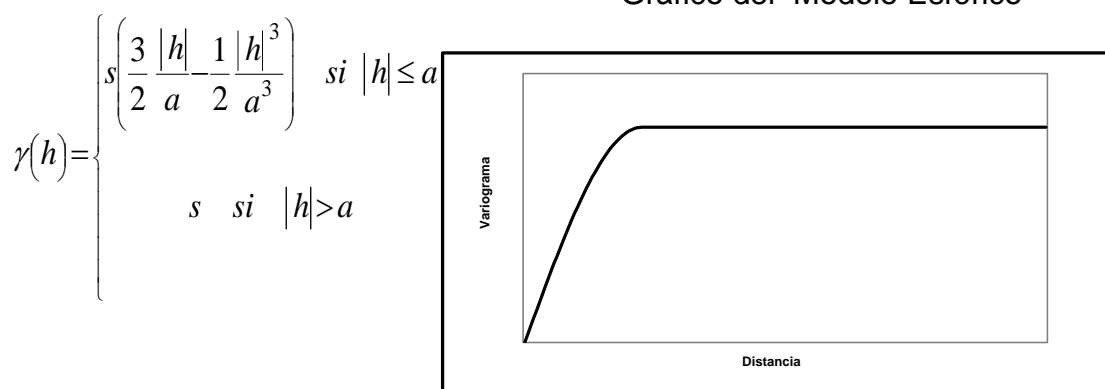
Autor: Evelyn Véliz

En este modelo podemos observar que crece inicialmente más rápido y después se estabiliza de forma asintótica. Como la meseta no se alcanza a una distancia finita, se usa con fines prácticos el “alcance efectivo” o “alcance práctico”, se considera que el rango a , es la distancia para la cual el valor del variograma es el 95% de la meseta.

2.6.3 MODELO ESFÉRICO

Figura 2.4

Gráfico del Modelo Esférico



Fuente: Geoestistics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

Por lo general es de todos los modelos el más utilizado para análisis geoestadístico, es una expresión polinomial simple, en el gráfico se puede observar un crecimiento casi lineal y después a cierta distancia finita del origen se alcanza una estabilización, la meseta.

2.6.4 MODELO GAUSSIANO

Figura 2.5

Gráfico del Modelo Gaussiano

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



Fuente: Geoestatics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

Este es un modelo continuo, inicialmente presenta un comportamiento parabólico en el origen, después al igual que en el modelo Exponencial se alcanza la meseta de forma asintótica. Constituye el único modelo estacionario con un punto de inflexión.

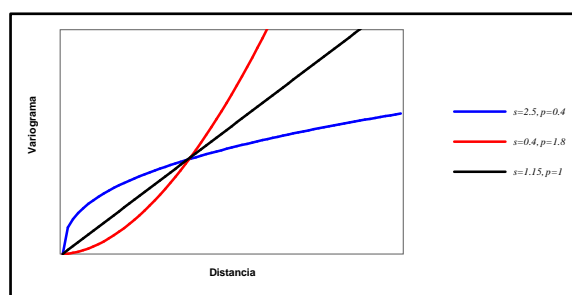
2.6.5 MODELO POTENCIA

Figura 2.6

Gráfico del Modelo Potencia

$$\gamma(h) = s |h|^p$$

$$0 \leq p < 2$$



Fuente: Geoestatics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

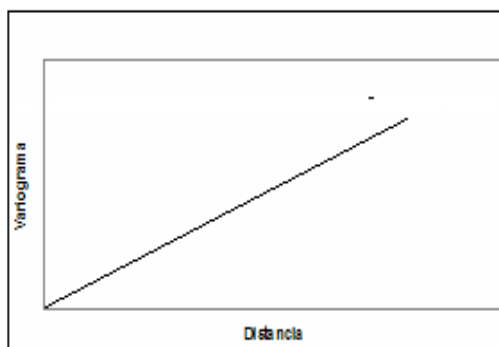
Este es un modelo sin meseta, su forma se representa con la intersección de tres líneas como podemos ver en la figura. Se aproxima a un comportamiento parabólico conforme p tiende a 2.

2.6.6 MODELO LINEAL

Figura 2.6

Gráfico del Modelo Lineal

$$y(h) = A(h) + B$$



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

El modelo Lineal no tiene sill ni rango. El sill continúa aumentando mientras la distancia h aumenta, es utilizado para modelar fenómenos que presentan capacidad infinita de dispersión.

2.7 ANISOTROPÍA

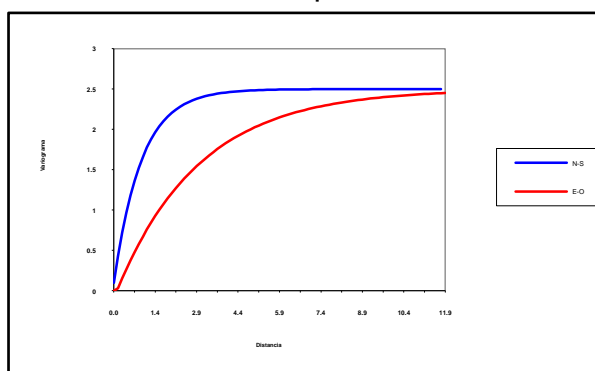
Anisotropía se da generalmente cuando el variograma experimental es calculado en diferentes direcciones y este presenta distintos comportamientos con la variación de la distancia, para conocer si tenemos existencia de anisotropía o no es recomendable hacer uso del diagrama de rosa. Existen 3 tipos de anisotropía Geométrica, zonal e híbrida.

2.7.1 ANISOTROPÍA GEOMÉTRICA

Es aquella en la que el variograma en distintas direcciones presenta el mismo sill pero rangos distintos. Presenta mayor continuidad espacial en la dirección de mayor rango y menor continuidad espacial en la dirección de menor rango.

Figura 2. 7

Gráfico de Anisotropía Geométrica



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

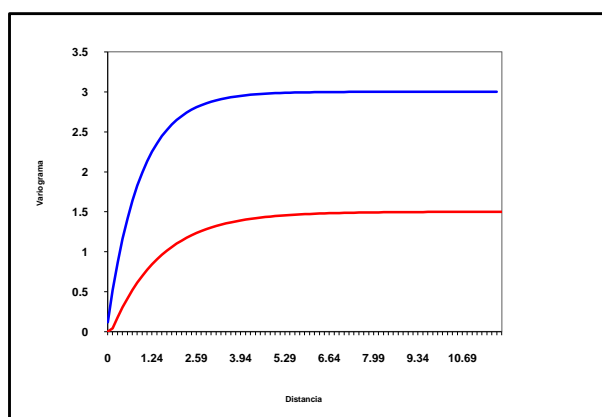
Autor: Evelyn Véliz

2.7.2 ANISOTROPÍA ZONAL

Es aquella en la que el variograma en distintas direcciones presenta el mismo rango pero diferente sill.

Figura 2. 8

Gráfico de Anisotropía Zonal



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

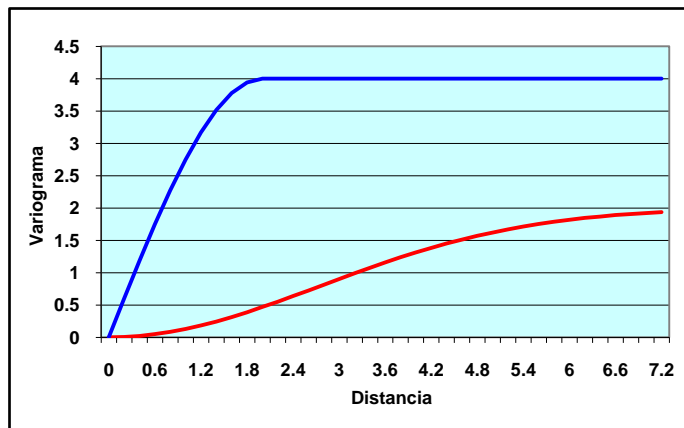
Autor: Evelyn Véliz

2.7.3 ANISOTROPÍA HÍBRIDA

Es aquella en la que el variograma en distintas direcciones presenta rangos diferentes y distintos sill.

Figura 2. 9

Gráfico de Anisotropía Híbrida



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

2.8 KRIGING

El kriging es un método de interpolación Geoestadístico, que está asociado con las siglas en inglés B.L.U.E (Best linear unbiased estimator), es lineal porque los estimadores se forman por la combinación lineal de los datos disponibles; tiene 2 características básicas:

1. Minimiza la varianza del error (el cuadrado de las desviaciones)
2. Es insesgado porque trata que la suma de los errores sea cercana a 0

El kriging utiliza el variograma para predecir valores desconocidos de variables distribuidas espacialmente, a partir de datos observados en lugares conocidos. Kriging es por lo tanto, el método para calcular el valor de una variable Z en un punto X_0 que ha sido considerado anteriormente, realiza una combinación lineal tomando los vecinos más cercanos al punto de interés X_0 , utiliza la fórmula que es la suma ponderada sobre todos los sectores que conforman la zona de estudio.

$$Z(X_0) = \lambda_1 z(x_1) + \lambda_2 z(x_2) + \dots + \lambda_n z(x_n)$$

$$\hat{Z}(x_0) = \sum_i^n \lambda_i Z(x_i) \quad \text{y} \quad \sum_{i=1}^n \lambda_i = 1$$

Donde: $z(x_i)$: Valores en los sitios donde hubo medición

λ_i : Vector de ponderación que minimiza el error de predicción

El kriging consiste en asignar pesos a los valores observados más cercanos, los pesos son calculados de manera que minimice la varianza de estimación resultante, teniendo en cuenta las características geométricas del problema, estos pesos son asignados a partir de un análisis espacial, basado en el Variograma experimental, la diferencia del kriging con otros métodos de interpolación es que utiliza un método semejante a la interpolación por media móvil ponderada.

Un método de interpolación será exacto cuando pase por los puntos muestrales, lo más cercano posibles a ellos. Para que las estimaciones de kriging sean más exactas y más confiables, es importante que un modelo para semivariograma que ha sido ajustado, represente una tendencia a los modelos antes descritos. Las estimaciones Kriging pueden ser por punto o por bloque.

2.8.1 KRIGING SIMPLE

En el kriging simple se supone que la media poblacional es nula o conocida, asume que las medias locales son relativamente constantes e iguales a la medida poblacional. La media poblacional es usada como un factor en cada estimación local, a lo largo con las muestras en la vecindad local.

2.8.2 KRIGING ORDINARIO

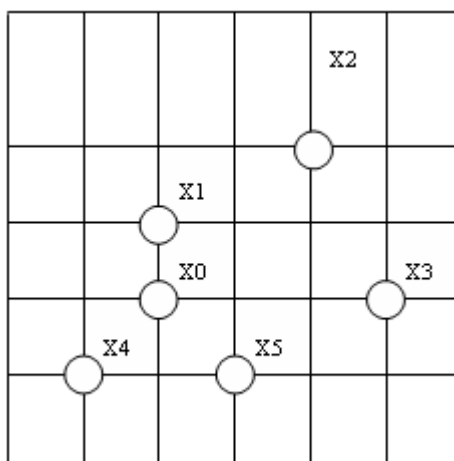
Es el método más utilizado para situaciones medioambientales, este método es la técnica más popular que elimina la necesidad de conocer el valor medio. Este método asume que las medias locales no son conocidas, por lo cual solo usa las muestra en la vecindad local para la estimación. La media varía suave y localmente.

2.8.3 KRIGING DE PUNTO

Como sabemos hay características que toman valores en puntos, es aquí donde este tipo de kriging estima los valores de los puntos en los nodos de las cuadrículas.

Figura 2. 10

Gráfico de Kriging de Punto



Fuente: Geostatistics for Natural Resources Evaluation, Goovaerts

Autor: Evelyn Véliz

2.8.4 KRIGING DE BLOQUE

Estima el valor promedio de los bloques rectangulares, que están centrados en los nodos de las cuadrículas, los bloques son de 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.

2.9 VARIABILIDAD ESPACIAL

Se entiende por variabilidad espacial de los modelos los cambios que ha sufrido el área de cultivo. Para tener información de estos datos es necesario trabajar con el GPS y obtener la posición del punto en latitud y longitud, para obtener las coordenadas geográficas, además es necesario recopilar otros datos de interés que mantengan una relación espacial como puede ser la calidad del suelo, cantidad de agua, densidad del cultivo, etc. Estos cambios pueden ser fácilmente vistos en un mapa de rendimiento, por ejemplo, para lo cual es necesario recopilar datos en posiciones precisas.

2.10 EFECTO DEL RANGO EN LAS ESTIMACIONES

Un valor grande para el rango “a” representa un comportamiento más continuo, las estimaciones dan como resultado mapas bastante lisos para la variable de interés.

2.11 EFECTO DEL MODELO EN LAS ESTIMACIONES

El modelo que mejor se ajuste al Variograma experimental, mostrará los mapas con más suavización.

2.12 EFECTO DEL SILL EN LAS ESTIMACIONES

El valor del sill es independiente al valor de las estimaciones, pero si afecta a la variación de las estimaciones, por lo que un sill más alto indica mayor variación en las estimaciones.

2.13 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 si, son estimadas.

- Independiente de los valores de los datos. Para un modelo de covarianza dado, la configuración de los datos idénticos podría producir la misma varianza kriging
- El mapa de varianzas, indica más incertidumbre en los sectores del área de estudio, en el que los datos están más dispersos de dicha área.