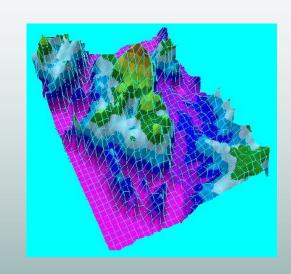
INSTITUTO MEXICANO DEL PETRÓLEO

GEOESTADÍSTICA APLICADA

Tema: Otros Métodos de Estimación Espacial



Instructores:

Dr. Martín A. Díaz Viera (mdiazv@imp.mx)

Dr. Ricardo Casar González (<u>rcasar@imp.mx</u>)

Contenido

- Introducción
- Interpoladores espaciales, definición
- Estimación vs representación
- Elementos de los interpoladores espaciales
 - La malla
 - Métodos de búsqueda
- Clasificación de los procedimientos de interpolación
- Métodos de interpolación:
 - El vecino más cercano
 - El vecino natural
 - Triangulación con interpolación lineal
 - Análisis de superficies de tendencias
 - Inverso de la distancia
 - Mínima curvatura
 - Kriging
- Comparación
- Otras técnicas Geoestadísticas
- Recomendaciones
- Conclusiones
- Referencias

Introducción

- La modelación de la distribución espacial de contaminantes en suelos se puede realizar usando esencialmente dos enfoques:
- Se resuelve un sistema de EDP de flujo y transporte que se ajustan a los valores medidos.
- A partir de unos pocos datos puntuales se predice la distribución espacial y/o temporal de los contaminantes.

Introducción

El segundo enfoque puede ser realizado de dos maneras:

 Estimaciones espaciales: Se estima el valor en ubicaciones no muestreadas.

 Simulaciones estocásticas espaciales: Se reproduce el comportamiento estadístico del fenómeno.

Introducción

- En particular, aquí se revisarán algunos de los métodos más comunes para la estimación de la distribución espacial de contaminantes en suelos.
- Abusando del lenguaje, en lo sucesivo nos referiremos a los estimadores espaciales como interpoladores.
- Estrictamente un interpolador requiere que la superficie estimada resultante pase exactamente por todos los valores de los puntos medidos

Interpoladores espaciales, definición

- La interpolación espacial es el procedimiento de estimar el valor de una propiedad en localizaciones no muestreadas en una área dentro de la cual existan observaciones.
- El razonamiento intuitivo de la idea de interpolación se establece en el ámbito de la Geografía mediante la ley de Tobler: las observaciones que se encuentran próximas unas con otras presentan valores similares en contraste con las observaciones que se encuentran alejadas, las cuales muestran valores poco parecidos.

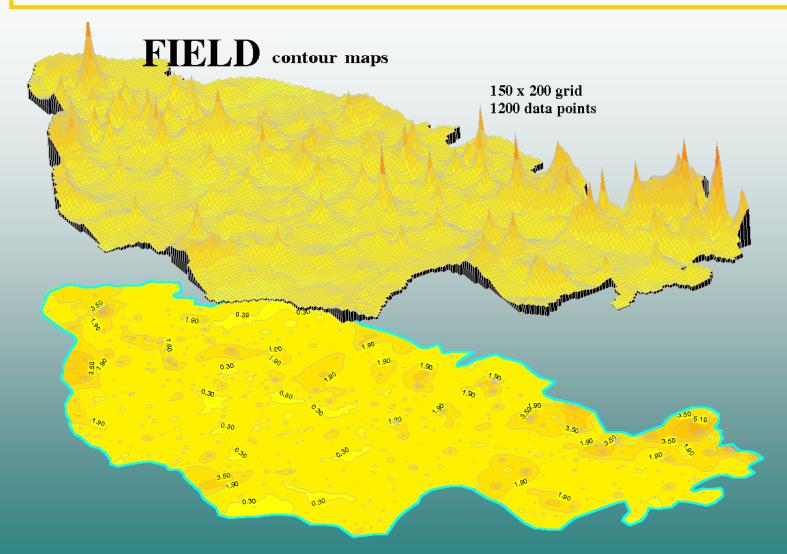
Estimación vs. Representación

- Resulta conveniente hacer la distinción entre el procedimiento de estimación y la forma de representación de los valores estimados. Usualmente se piensa en la representación (mapas) como si fuera el método de estimación (interpolador)
- La implementación de los métodos de representación gráfica (mapas) requieren como premisa que los valores estén ubicados en una malla regular, lo cual sólo se puede realizar mediante un procedimiento de interpolación.
- Existen una gran diversidad de métodos para la representación gráfica de los resultados de la estimación usando mapas (por ejemplo, de contornos, de imagen, de relieves, etc.)

Representación Gráfica

- Los contornos es una representación gráfica del resultado de la interpolación.
- En geografía, geociencias y otras disciplinas afines es importante describir variaciones espaciales de parámetros.
- Dichas variaciones pueden ser representadas en mapas a través de líneas de contornos o líneas de isovalores.
- Cualquier mapa que utilice líneas de contornos para representar las variaciones espaciales de una variable o parámetro puede ser considerado un mapa de contornos.
- Un ejemplo son los mapas topográficos que representen variaciones en la elevación del terreno.

Representación Gráfica

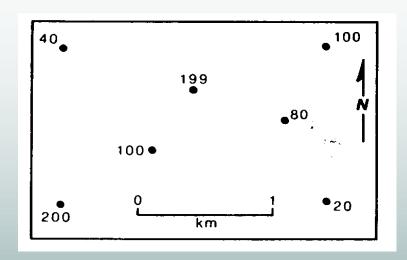


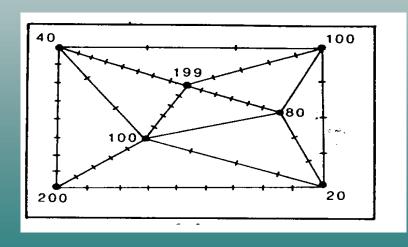
Tipos de Contornos

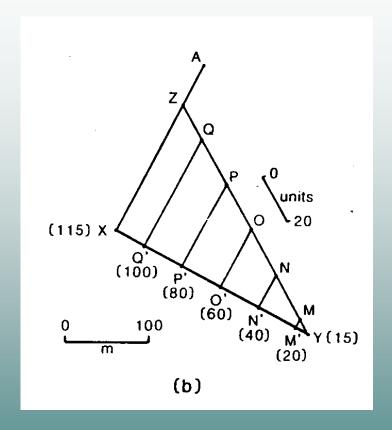
Tipos de Contornos:

- Contornos Reales: Esta superficie o contornos refleja los valores reales de los datos.
- Contornos de Tendencia: Esta superficie refleja una tendencia regional de los valores, se usa para realzar características regionales ocultas en un plano de datos reales.
- Contornos Residual: Este mapa se emplea para analizar características o anomalías particulares que en ocasiones se ocultan por la influencia de tendencias regionales. Se elabora primero una malla de tendencias y esta se resta de los datos originales para mostrar las anomalías locales o residuos.

- Una de las técnicas comúnmente utilizadas para crear contornos manualmente es el llamado método de triangulación.
- Se forman triángulos uniendo los puntos de control y se interpolan los valores de los puntos de control a lo largo de los vértices de los triángulos, para esto se esta suponiendo que los gradientes entre los puntos de control son constantes.
- Para interpolar se utiliza un método gráfico.

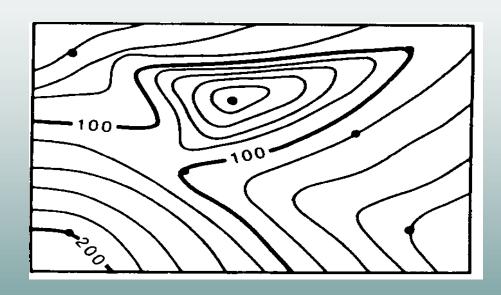






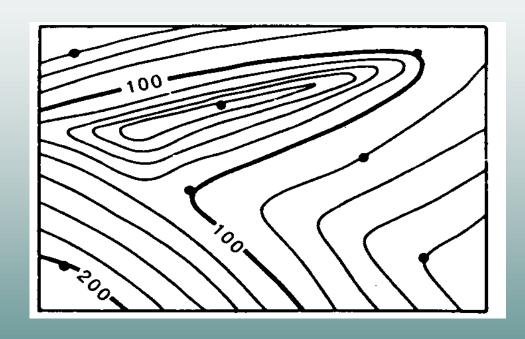
Método gráfico de interpolación lineal

Contornos Constantes:



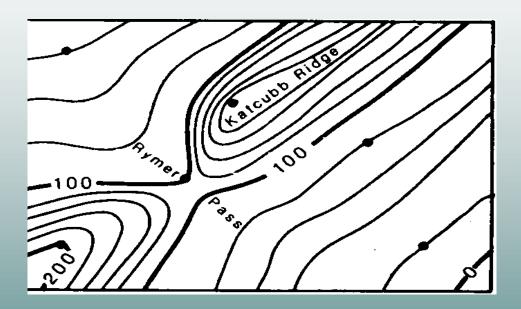
• Se asume que las pendientes entre dos puntos de control son constantes, se interpola entre dos puntos y se dibujan los contornos haciéndolos pasar por sus respectivos puntos de control.

Contornos Paralelos:



• Se asume que los contornos son paralelos entre si, se dibujan tratando en lo posible que sean paralelos.

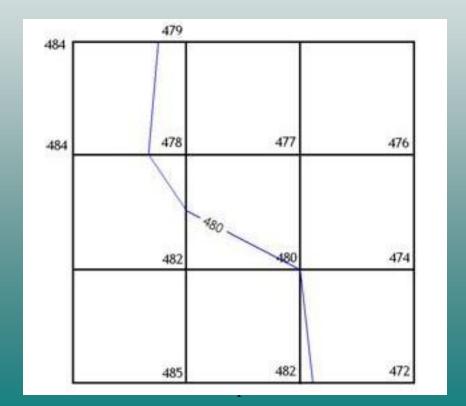
Contornos Libres:



• Esta opción considera que los puntos interpolados representan un muestreo aleatorio de la información y que los gradientes no son constantes ni paralelos, así que el intérprete dibuja los contornos siguiendo un criterio preconcebido.

Método de Creación de Contornos

 Un contorno se construye mediante segmento de línea recta de valor constante (por ejemplo, 480). Se puede construir a partir de pares de nodos en dirección NS y EW por interpolación lineal simple. El resultado es una polilínea con intersecciones a lo largo de la malla o cuadrícula y de valor constante

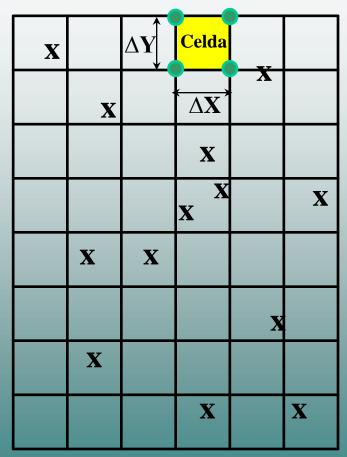


Elementos de los interpoladores

- La malla
- Métodos de búsqueda
- Discontinuidades (fallas, barreras, etc)
- Métodos de interpolación

La Malla

Malla



Nodos •

Muestra X

- Para efectuarse una interpolación o estimación debe de crearse una malla.
- La malla debe cubrir el área que posee información.
- La malla esta formada por celdas y nodos.
- La malla esta formada por un número determinado de celdas por X y por Y.
- Debe asignarse un tamaño a la celda $(\Delta X, \Delta Y)$.
- El tamaño de la celda determina la densidad de la malla.
- Es el grado de resolución con que se efectúa la estimación.

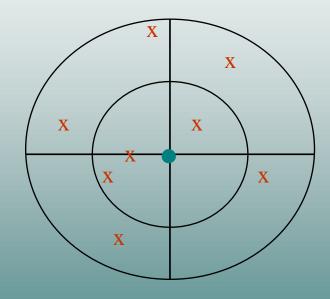
Métodos de búsqueda

- Los métodos para identificar el conjunto de datos que se emplearan en la estimación de un punto desconocido son llamados métodos de búsqueda.
- Los métodos de búsquedas están formados por:
 - Una vecindad de búsqueda
 - Una regla de búsqueda
- La selección del método de búsqueda es muy importante ya que de éste depende fuertemente el resultado de la interpolación.

Métodos de búsqueda

- Vecindad de búsqueda
 - Anillos y sectores

- Reglas de búsqueda:
 - Puntos más cercanos



Métodos de búsqueda

Anillos y Sectores:

- Este método asegura el uso de puntos de control hacia todas las direcciones a partir de un nodo que será estimado.
- Esta técnica es adecuada cuando los datos están agrupados.
- Se define un radio de influencia o de búsqueda, es decir el radio máximo de búsqueda de puntos de control alrededor de un nodo.
- Se define el número de sectores radiales y el número de puntos de control para cada sector dentro del radio de búsqueda.

- Interpoladores puntuales y por áreas o bloques
- Interpoladores globales y locales
- Interpoladores exactos y aproximados
- Interpoladores graduales y abruptos
- Interpoladores estadísticos y determinísticos

- Interpoladores puntuales y por áreas o bloques:
- Los datos conocidos (temperaturas, porcentaje de algún contaminante, tipo o alguna propiedad del suelo) se emplean para interpolar los nodos de una malla. La interpolación se efectúa de punto a punto, es el método más empleado en los GIS La malla interpolada se usa como datos de entrada para algún algoritmo de creación de contornos computarizado.
- La interpolación por área o por bloque se refiere a que dado un conjunto de datos mapeados en un conjunto de área fuente, determinar el valor de los datos para diferentes conjuntos de áreas, ejemplo: Dado un conteo de población de una zona determinada, estimar la población para un distrito en particular.

- Interpoladores globales y locales:
- Interpoladores globales: se determinan mediante una sola función, la cual es empleada en toda la región. Un cambio en un valor de entrada afecta toda el área. Tienden a producir superficies suavizadas, sin cambios abruptos.
- Interpoladores locales: utilizan un algoritmo específico para porciones menores del conjunto total de datos. Un cambio en un valor de entrada solo afecta el resultado de un área reducida o ventana. En ocasiones un interpolador local se extiende hasta incluir el total del área y de esta manera cambia su carácter de local a global.

- Interpoladores exactos y aproximados:
- Interpoladores exactos: respetan los datos sobre los cuales se basa la interpolación. La superficie pasa a través de todos los puntos de los cuales se conoce su valor. En algunas aplicaciones es importante respetar los puntos conocidos.
- Interpoladores aproximados: la superficie no pasa necesariamente a través de todos los puntos de los cuales se conoce su valor. Se utiliza cuando hay incertidumbre sobre los datos conocidos. Cuando se sospecha de la presencia de tendencias globales. Presentan un efecto de suavizado, con el fin de reducir el error en la superficie resultante.

- Interpoladores graduales y abruptos:
- Los interpoladores graduales: producen superficies con cambios graduales, el ejemplo típico son los interpoladores como los promedios móviles ponderados con la distancia (inverso de la distancia). Pero, cuando se tiene un número reducido de puntos de control, el resultado puede ser una superficie con cambios abruptos.
- Los interpoladores abruptos: producen superficies con cambios abruptos, son útiles cuando se requiere incluir barreras en el proceso de interpolación, tal es el caso de barreras impermeables en el subsuelo o fallas geológicas.

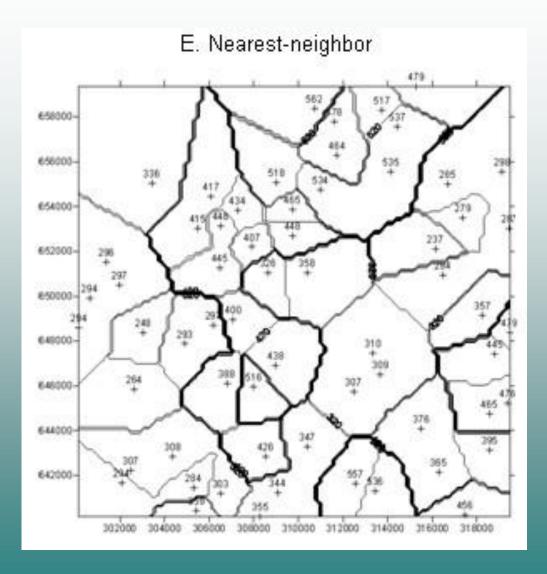
- Interpoladores estadísticos y determinísticos:
- Los interpoladores estadísticos: incorporan conceptos de estadística. De los interpoladores estadísticos podemos incluir el Análisis de Superficies de Tendencia (trend surface analysis), análisis de Fourier y Kriging, este último garantiza una mínima variancia de la estimación y permite estimar el grado de incertidumbre de los valores estimados.
- Los interpoladores determinísticos: no incorporan conceptos de probabilidad y estadística en el proceso. Como ejemplos tenemos el método inverso de la distancia (promedios móviles ponderados con la distancia) y el método de Splines.

Métodos de interpolación

El vecino más cercano

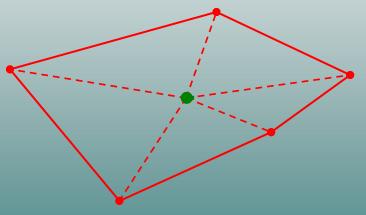
- El vecino más cercano (Nearest Neighbor)
- Esta técnica simplemente asigna el dato más cercano para cada nodo que se calcula.
- Es útil para datos regularmente espaciados en una retícula, en la cual el muestreo es lo suficientemente denso.
- Se debe tener cuidado al aplicarlos ya que es un método sumamente simple.

El vecino más cercano



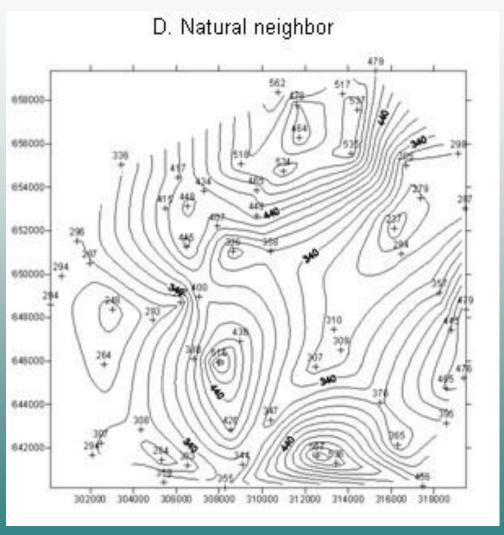
Vecino Natural

- Primero se forman triángulos con los puntos con datos.
- El vecino natural de un punto dado serán todos los puntos cuyos triángulos comparten vértices.
- Los vecinos naturales forman un polígono cerrado (polígonos de Thiessen) alrededor del punto de control.



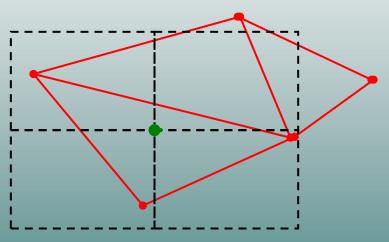
• El algoritmo de interpolación consiste en obtener un promedio ponderado de las observaciones que sean vecinos naturales, donde los pesos son proporcionales al área.

Vecino Natural



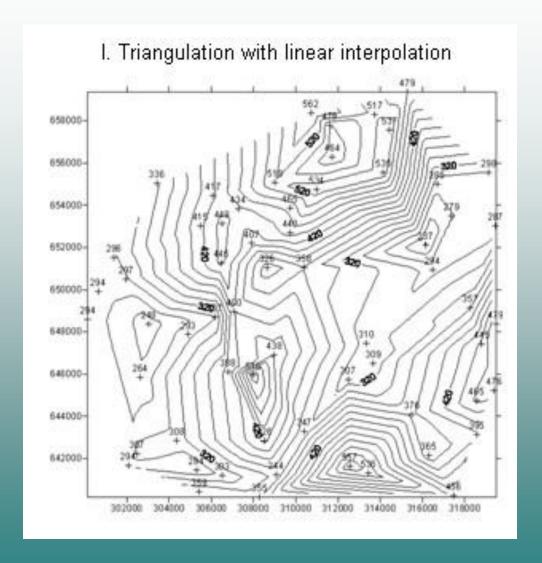
Triangulación con interpolación lineal

- Es un método exacto de interpolación basado en la generación previa de una malla irregular de triángulos (triangulación de Delaunay) cuyos vértices coinciden con los puntos conocidos.
- La interpolación se realiza suponiendo que dichos puntos pertenecen a la superficie plana de primer orden que se apoya en los vértices de cada triángulo.



 Los valores máximos y mínimos se conservan. En ocasiones los mapas presentan el llamado "efecto de borde", cuando los triángulos son demasiado oblicuos.

Triangulación con interpolación lineal



Análisis de superficies de tendencias

- Este método es un interpolador global, el cual crea una superficie mediante una aproximación polinomial.
- El análisis de tendencias realiza el ajuste polinomial de los datos, matemáticamente se trata de minimizar la suma de los cuadrados de las desviaciones en el dominio de Z, para lo cual hay que definir el orden del polinomio.
- Se supone que la tendencia general de la superficie es independiente de los errores aleatorios que pudieran encontrarse en cada punto conocido.
- La demanda computacional de este método es baja y el efecto de bordes puede ser severo. Generalmente se encuentra en los software de interpolación.

Análisis de superficies de tendencias

 Un polinomio de primer grado define un plano, la ecuación será de la forma:

$$Z(x,y) = b_0 + b_1 x + b_2 y$$

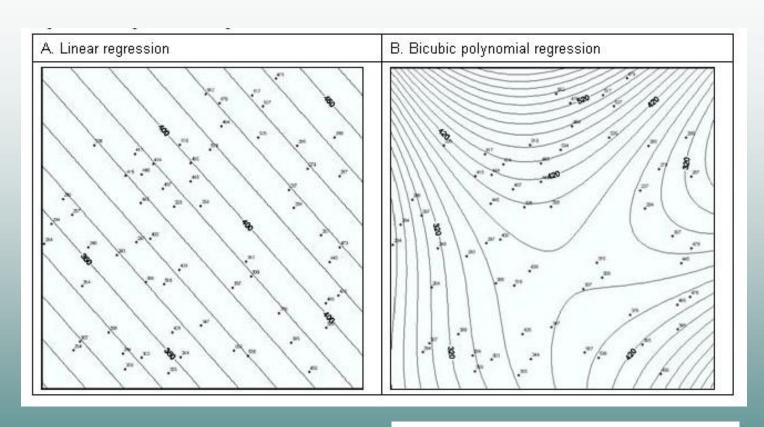
- En donde (x, y) representan las coordenadas cartesianas y describe una superficie plana.
- Un polinomio de segundo grado presenta la forma

$$Z(x,y) = b_0 + b_1 x + b_2 y + b_3 x^2 + b_4 y^2 + b_5 xy$$

y puede describir una elevación (cerro) o una hondonada (valle)

 Esta metodología permite crear dos superficies: siendo una la componente regional y otra la residual. El objetivo es visualizar tendencias generales o bien detalles estructurales.

Análisis de superficies de tendencias



$$z(x,y) = a_1 + a_2 x + a_3 y$$

$$z(x,y) = a_1 + a_2x + a_3y + a_4x^2 + a_5y^2 + a_6xy + a_7x^2y + a_8xy^2 + a_9x^3 + a_{10}y^3$$

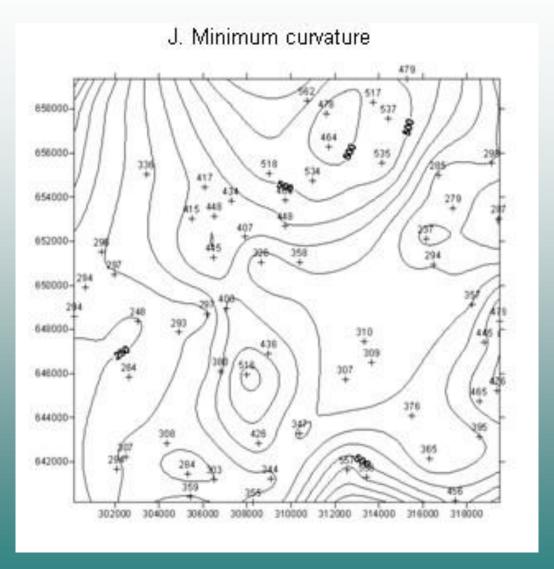
Mínima Curvatura

- Mínimo de curvatura es similar al método de spline bi-cúbico. Sin embargo, la interpolación no es exacta, pero cerca de la exacta, y está diseñado para asegurar que la cantidad de curvatura de la superficie es tan pequeña como sea posible.
- Esto se logra a través de un proceso de múltiples etapas. la primera etapa de regresión lineal simple y la extracción de los residuos. El procedimiento utiliza estos residuos en lugar de la interpolación de los datos originales de puntos. Al finalizar, a la interpolación de los residuos se le añade la superficie de regresión.
- El proceso de interpolación es iterativo que busca suavizar la superficie interpolada según un parámetro pre-especificado. El procedimiento es bastante complejo y se describe en detalle en Smith y Wessel (1990).

Mínima Curvatura

- Spline en matemáticas se refiere a una función, que tiene valores específicos en un número finito de puntos y consiste de segmentos de funciones polinomiales unidas en estos punto, lo cual permite su uso para la aproximación e interpolación de funciones.
- Este método se considera un interpolador local y exacto. Utiliza polinomios con el fin de proporcionar una serie de parches o piezas que resultan en una superficie que presenta la primera y segunda derivadas continuas.
- Se asegura continuidad en: elevación (que la superficie no tenga picos), en gradiente (que la superficie no tenga cambios de pendiente abruptos) y en curvatura (que la superficie posea mínima curvatura).

Mínima Curvatura



Inverso de la distancia

- Este algoritmo es uno de más sencillos y eficientes, el valor de cada nodo es determinado a partir de un conjunto de datos vecinos determinados de acuerdo al algún método de búsqueda.
- Utiliza los puntos vecinos al nodo que se va a calcular y les asigna un peso o valor de acuerdo a su proximidad o lejanía, utilizando la siguiente formula:

$$w_i^k = 1/(d_i^k)^e$$

en donde:

 W_i^k - peso asignado al *i*-ésimo dato vecino al nodo \underline{x}_k $d_i^k = |\underline{x}_i - \underline{x}_k|$ - distancia entre el *i*-ésimo dato vecino y el nodo \underline{x}_k e- exponente entero asignado por el usuario

 Nótese que entre mayor es el exponente e, menor será el peso para una distancia dada, así mismo entre mayor es la distancia menor será el peso w.

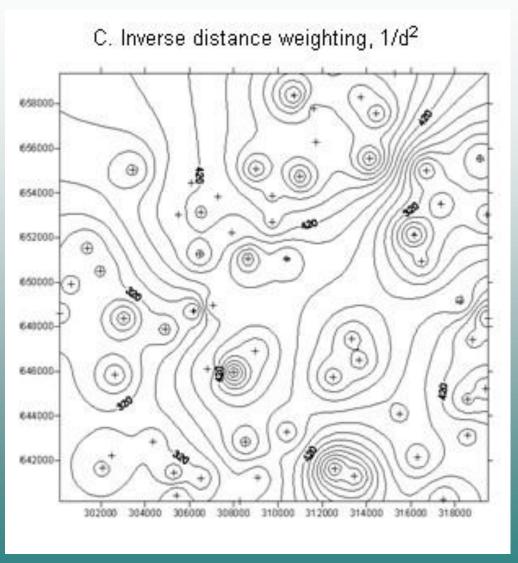
Inverso de la distancia

 Una vez calculados los pesos para cada punto vecino el valor de Z estimado se calcula mediante una combinación lineal:

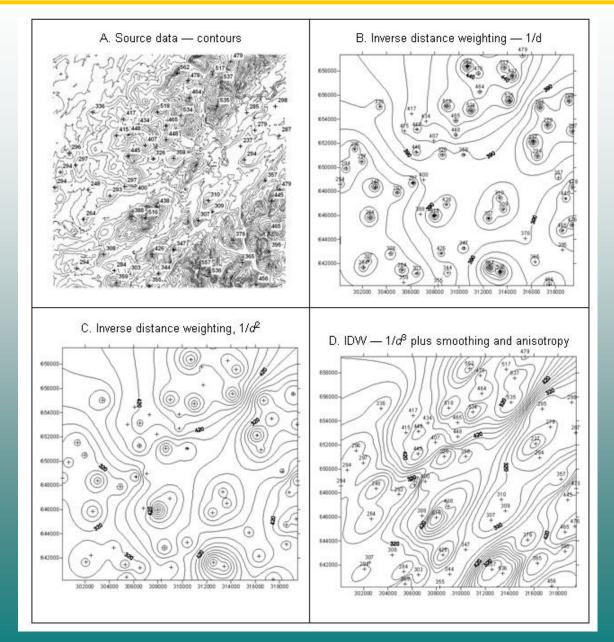
$$Z^*\left(\underline{x}_k\right) = \left(\sum_{i=1}^N w_i^k Z\left(\underline{x}_i\right)\right) / \sum_{i=1}^N w_i^k$$

- Con lo cual se le da más importancia (peso) a los valores cercanos al nodo por estimar.
- Nótese que aunque algunos nodos pueden compartir los mismos valores de Z, pero cada nodo tiene su particular conjunto de pesos únicos.
- Para un mejor funcionamiento de este método es importante tener cuidado en la selección de los parámetros que afectan la manera como se determinan los puntos vecinos a considerar (método de búsqueda).

Inverso de la Distancia



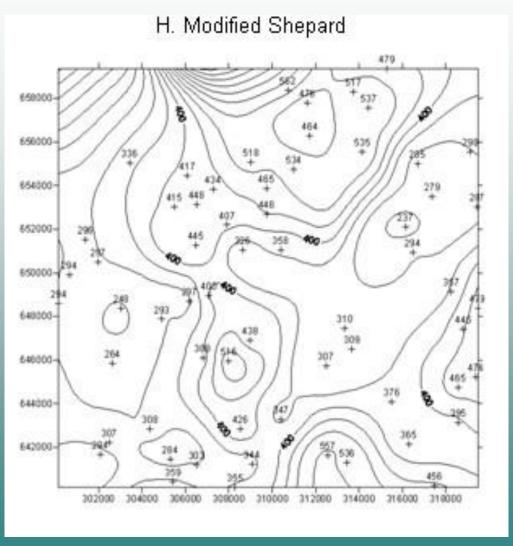
Inverso de la Distancia



Shepard Modificado

- Esencialmente es el método de inverso de la distancia
- Hay muchas variantes del método de Shepard original de interpolación.
- La variante introducida por Shepard implica el uso de dos potencias diferentes: un valor de potencia menor (por lo general 2) para los dato más cercanos y una potencia mayor (por lo general 4) para los datos más alejados
- El software Surfer utiliza un procedimiento más complejo, basado en un ajuste local polinomial cuadrático en la vecindad de cada dato. Esto se conoce como el método cuadrático de Shepard modificado.
- El método puede ser exacto o aproximado dependiendo de si se aplica un procedimiento de suavizado.

Shepard Modificado



- El Kriging es un término que ha sido acuñado para designar al "mejor estimador lineal insesgado" (BLUE, en inglés).
- Esta es una técnica de estimación 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.

- El estimador Kriging se considera óptimo ya que:
- Es insesgado, es decir, el valor esperado del error es cero.

$$oldsymbol{E}\left[oldsymbol{Z}_{k}^{*}
ight]=oldsymbol{E}\left[oldsymbol{Z}_{k}
ight]$$

 Garantiza la mínima varianza de la estimación, es decir, reduce al mínimo la varianza del error de la estimación.

$$\min \left\{ Var \left[Z_k - Z_k^* \right] \right\}$$

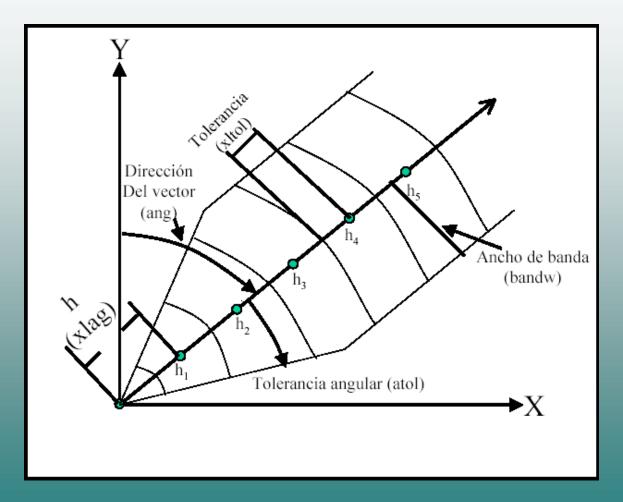
- Pero para que su desempeño sea óptimo requiere que se realice un análisis geoestadístico el cual consta de tres etapas:
 - Análisis exploratorio de los datos
 - Análisis variográfico
 - Estimación espacial

- Análisis exploratorio de los datos
- El análisis exploratorio de datos (AED) es un paso previo e indispensable para la aplicación exitosa de cualquier método estadístico.
- En particular permite la detección de fallos en el diseño y toma de datos, el tratamiento y/o la evaluación de datos ausentes, la identificación de valores atípicos y la comprobación de los supuestos requeridos por parte de las técnicas geoestadísticas.

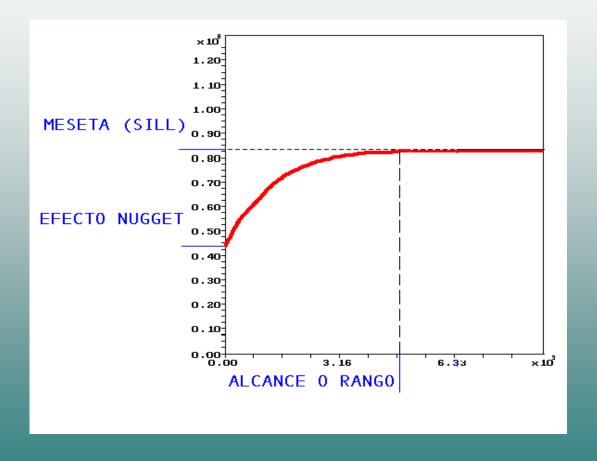
- Análisis variográfico
- Consiste en estimar y modelar una función (el variograma) que refleje la autocorrelación espacial de la variable que se estudia.
- El variograma se calcula como sigue:

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

Consideraciones para el cómputo del variograma:

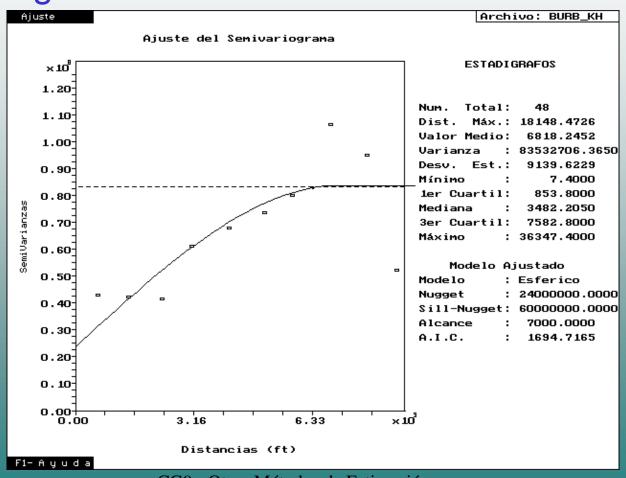


Forma general del variograma acotado

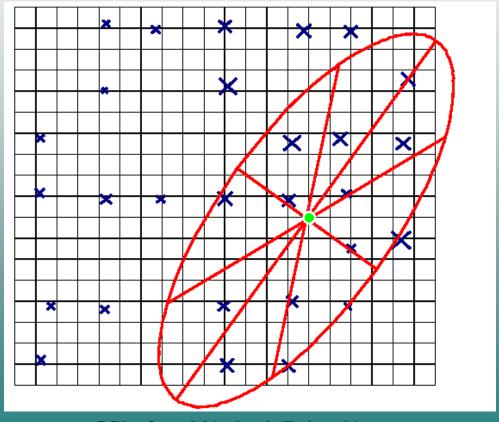


 Modelación variograma: del

- Modelo esférico
- Modelo exponencial
- Modelo Gaussiano



 Mediante el variograma, Kriging permite tomar en cuenta la anisotropía del fenómeno:

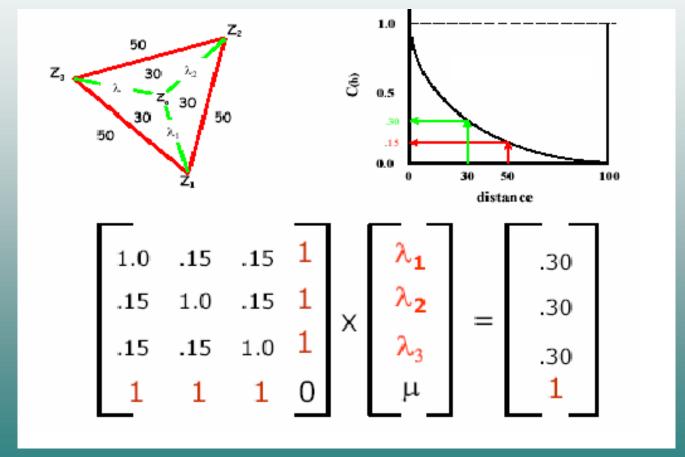


- Estimación espacial
- Ecuaciones del Kriging en un punto \underline{x}_k

$$\begin{bmatrix} \gamma_{11} & \gamma_{12} & \dots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \dots & \gamma_{2n} & 1 \\ \dots & \dots & \dots & \dots \\ \gamma_{n1} & \gamma_{n2} & \dots & \gamma_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \dots \\ \lambda_n \\ -\mu \end{bmatrix} = \begin{bmatrix} \gamma_{k1} \\ \gamma_{k2} \\ \dots \\ \gamma_{kn} \\ 1 \end{bmatrix}$$

Estimación espacial

$$oldsymbol{Z}_k^* = \sum_{i=1}^N \lambda_i oldsymbol{Z}_i$$



Clasificación de los tipos de Kriging:

- Según la forma del estimador:
- Lineales:

Simple

Ordinario

Universal

Residual

No lineales:

Disyuntivo

Indicador

Probabilístico

- Según el supuesto de la distribución de probabilidad:
- Paramétrico:

Multigaussiano

Disyuntivo

Lognormal

No paramétrico:

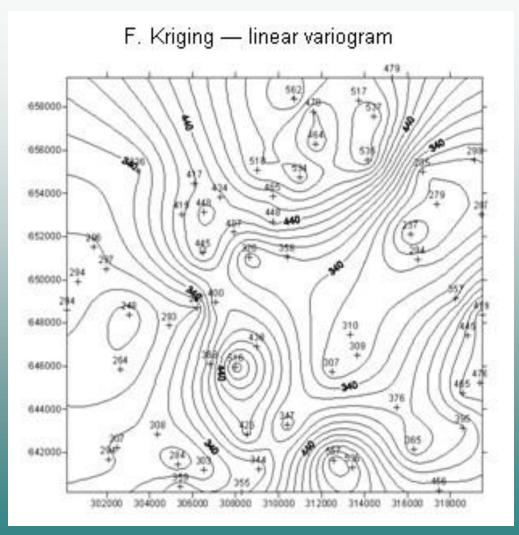
Simple

Ordinario

Universal

Residual

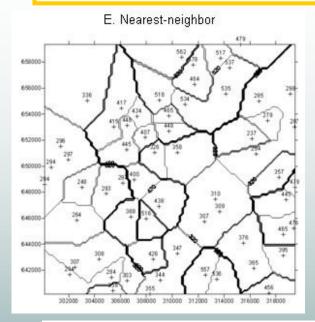
Indicador

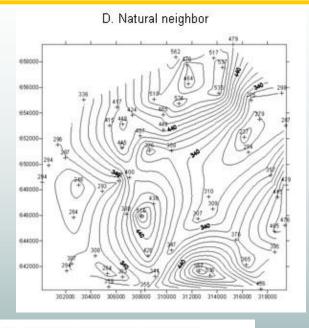


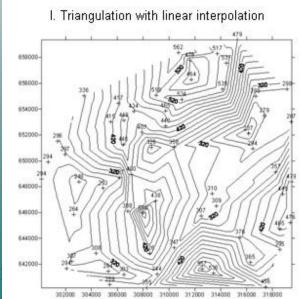
Resumen:

- Kriging es un estimador lineal que garantiza insesgadez (suma de errores igual a cero) y mínima varianza.
- Kriging nos da un mapa de la incertidumbre en la estimación.
- Kriging requiere de un modelo de correlación espacial, el cual se realiza mediante variogramas.
- Kriging requiere de una masa estadística (mínimo 30).
- Kriging Ordinario requiere estacionaridad, es decir que la variable no presente cambios en el valor medio (presencia de tendencia) dentro del área a estimar.

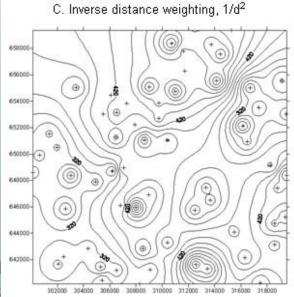
Comparación



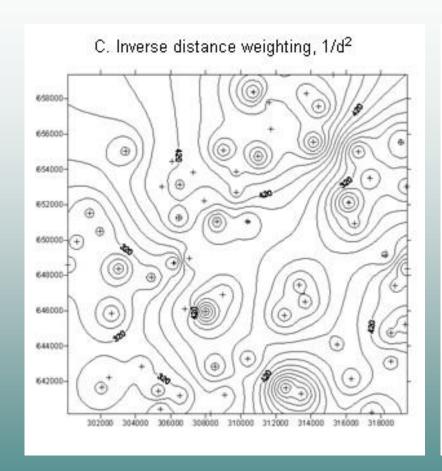


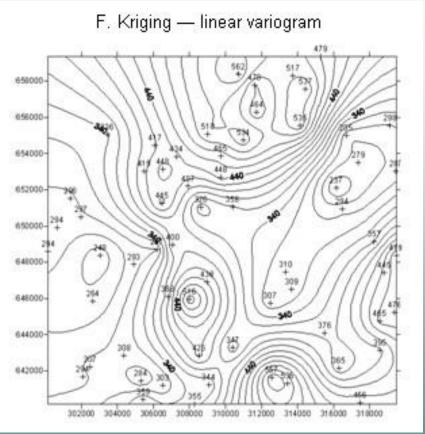


03/1



Comparación





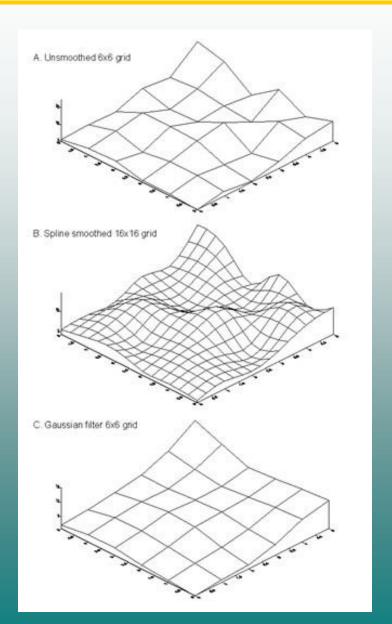
Otras técnicas Geoestadísticas

- Estimaciones multivariadas: El Cokriging es la extensión natural del Kriging cuando se tiene mas de una variable correlacionada entre sí. El Cokriging utiliza la correlación espacial y la correlación entre variables al mismo tiempo.
- Las simulaciones geoestadísticas incorporan los conceptos de probabilidad y de aleatoriedad, la interpolación resultante puede ser considera como una de muchas que pueden ser producidas con los mismos datos de entrada. La idea básica detrás de las simulaciones estocásticas consiste en obtener nuevas realizaciones "artificiales" de la propiedad de manera tal que éstas posean las mismas propiedades estadísticas. Dependen menos del número de datos. Con simulaciones es posible realizar análisis de riesgo.

Métodos de Suavizado

- Las superficies pueden ser suavizadas de diferentes maneras.
- Suavizar una superficie se refiere a diversos procedimientos para modificar la malla existente y lograr superficies más suaves.
- El suavizado se realiza con dos métodos principalmente:
 - Aumentar la resolución de la malla, normalmente se utilizan funciones spline bicúbicas.
 - Suavizar mediante la aplicación de un proceso de filtrado.

Métodos de Suavizado



Recomendaciones

- La selección del método de interpolación más adecuado tiene que ver con los objetivos que se quieren alcanzar. Uno de los aspectos a considerar es el tamaño de la muestra y su distribución espacial.
- Una muestra de 10 o menos datos, por lo general, no produce más que una tendencia general.
- Para una muestra pequeña (menos de 250 datos), se recomienda el método de Triangulación con interpolación lineal.
- Para una muestra mediana (de 250 a 1000 datos), se recomienda Triangulación, Kriging o Inverso de la distancia, en este último debe elegirse adecuadamente el método de búsqueda.
- Para una muestra grande (más de 1000 datos), se recomienda Mínima curvatura, Triangulación o Kriging.

Conclusiones

Interpoladores exactos:

- Mínima curvatura
 - Los valores máximo y mínimo no necesariamente coinciden con los máximos y mínimos de los datos. Se recomienda para interpolar superficies suaves. Trabaja bien con muestras grandes.

Kriging

 Cuando los criterios que se involucran en kriging no se satisfacen en su totalidad no es recomendable utilizarlo. Trabaja bien con suficientes muestras.

Interpoladores aproximados:

- Análisis de superficies de tendencias
 - Produce superficies redondeadas. La demanda computacional de este método es alta y el efecto de bordes es severo.
- Inverso de la distancia
 - Es un método muy efectivo. Lo afecta la selección del método de búsqueda. Tiende a producir el efecto de ojo de buey alrededor de los datos.

Referencias

- Notas del Curso: GEOESTADÍSTICA APLICADA, Instituto de Geofísica, UNAM, Instituto de Geofísica y Astronomía, Ministerio de Ciencia, Tecnología y Medio Ambiente de Cuba, 135 págs., 2002.
- Presentaciones del Curso: GEOESTADÍSTICA APLICADA, (Martín A. Díaz Viera, Ricardo Casar González), Instituto Mexicano del Petróleo, 480 págs., 2004.
- Geospatial Analysis: A Comprehensive Guide to Principles, Techniques and Software Tools, Mike de Smith, Mike Goodchild, Paul Longley, June 2009.
- Geostatistics for Environmental Scientists, by Richard Webster, Margaret A. Oliver, Wiley, 2007.
- ArcGIS Desktop Help 9.3,(http://webhelp.esri.com/)
- HELP de SURFER 7, (http://www.goldensoftware.com/)