GEOESTADÍSTICA

Estimación de semivariogramas y KRIGEADO

Concepción González García (2008)

Imagen cortesía de la NASA

El análisis estructural

Está compuesto por:

- El cálculo del semivariograma experimental.
- El ajuste al semivariograma empírico de un modelo teórico conocido.

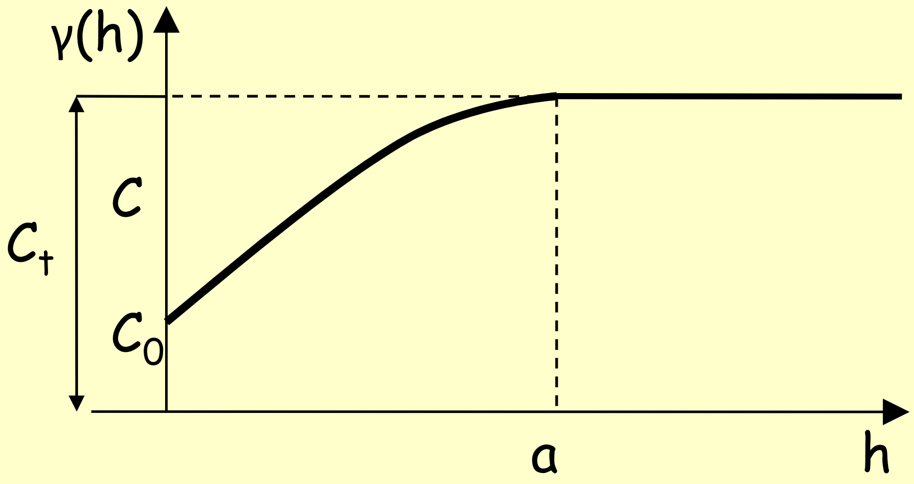
Objetivo: determinar los parámetros descriptivos del semivariograma que posteriormente serán usados en la estimación.

Modelado de semivariogramas

Parámetros del semivariograma:

Son tres elementos que caracterizan la variabilidad de un atributo:

- la discontinuidad en el origen (existencia de efecto de pepita: C_0)
- el valor máximo de variabilidad (meseta: C_t), y
- el área de influencia de la correlación (alcance: a),



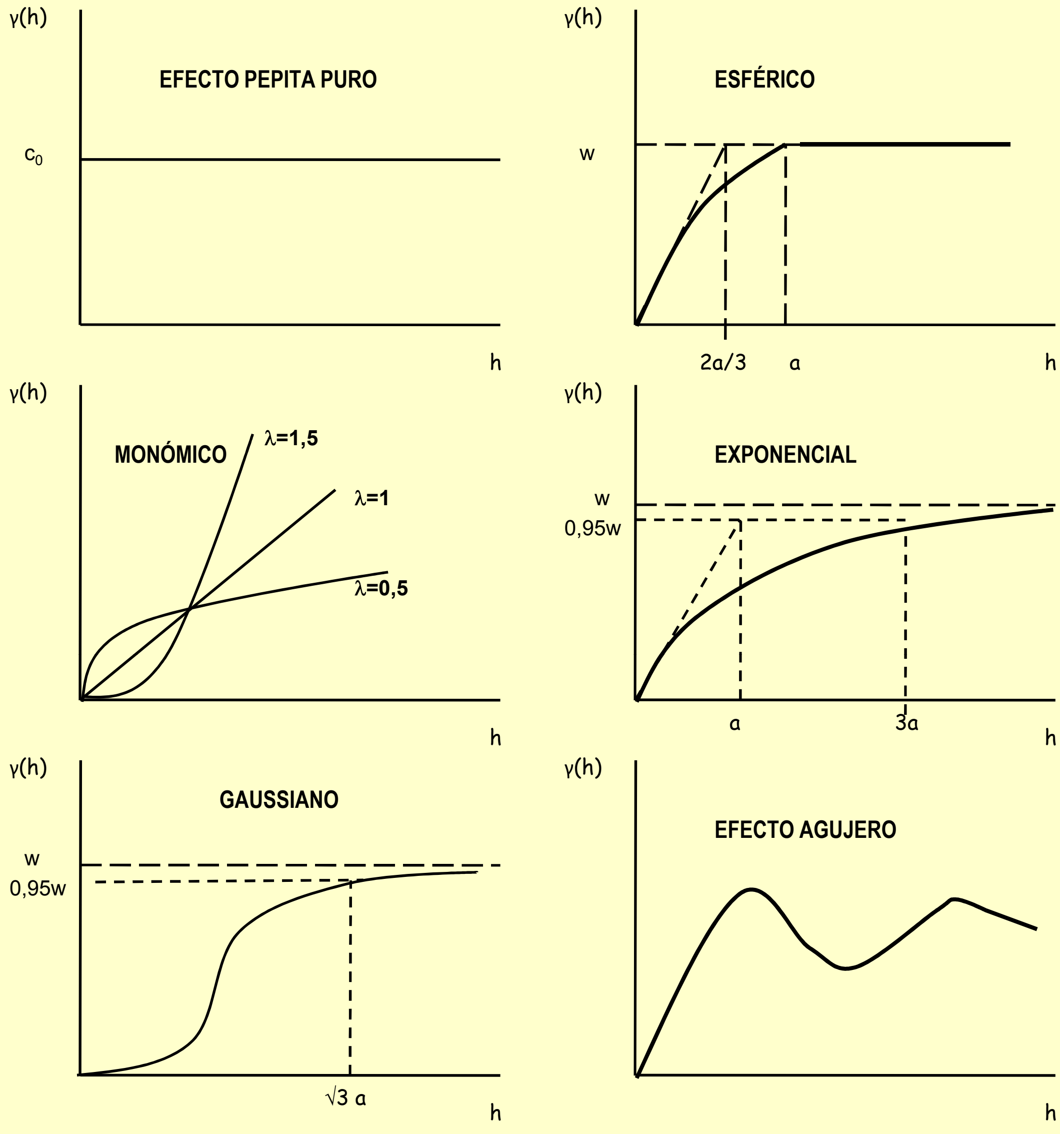
Parámetros del semivariograma

Modelos teóricos de semivariogramas

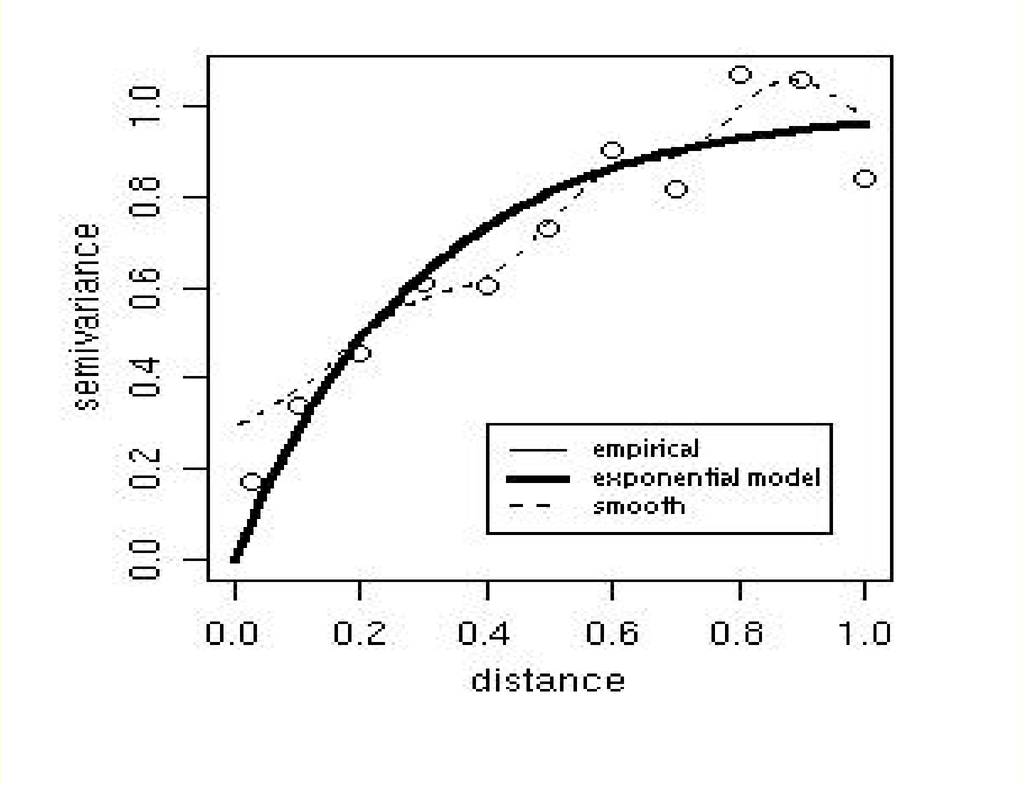
Características:

1.- Su comportamiento en el origen, el cual puede ser lineal, parabólico y con Efecto de Pepita y

2.- La presencia o ausencia de meseta



Se dibujan los variogramas, empírico y teórico, y se comparan visualmente. Por ejemplo, en la figura se muestra el modelo teórico de variograma , usado para simular una muestra de 100 datos, y dos variogramas estimados.



geoR : Package for Geostatistical Data Analysis
An illustrative session
Paulo J. Ribeiro Jr. & Peter J. Diggle
Last update: 26/Dez/2003

Modelado de semivariogramas (i)

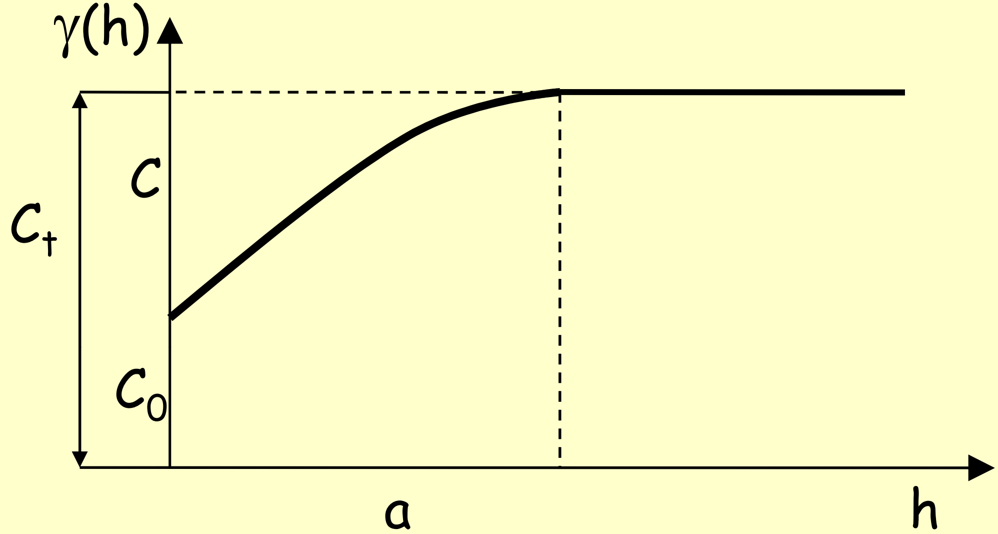
La selección del modelo y los parámetros apropiados a las características del semivariograma empírico, para ser usados en la interpolación *geoestadística* es el punto más importante del proceso planteado .

Sin un estudio de estructura espacial y la selección adecuada del modelo de semivariograma y sus parámetros, según muchos autores, el empleo del Krigeado puede tener un efecto negativo en la estimación.

Modelado de semivariogramas: Validación del modelo teórico

El ajuste de los modelos teóricos al semivariograma experimental, se realiza de forma visual o interactiva, variando los valores,

Co (efecto pepita),
C + Co (meseta) y
a (alcance),



Parámetros del semivariograma

hasta coincidir con los parámetros que mejor se ajustan, es conveniente validar el modelo seleccionado y los parámetros meseta y alcance escogidos

Modelado de semivariogramas: Validación del modelo teórico (ii)

El **método de validación cruzada**

Sea $Z(x)$ una función aleatoria estacionaria con semivariograma $\gamma(h)$, su función de covarianza $C(h)$ viene dada por $C(h) = \sigma_2 - \gamma(h)$ donde σ^2 es la varianza de $Z(x)$.

Sea Zx_1, Zx_2,...,Zx_n los valores de $Z(x)$ en n puntos medidos.

La validación cruzada consiste en suprimir el i-ésimo valor medido $Z(x_i)$ y estimarlo a partir del resto de los datos.

El valor estimado $Z^*(x_i)$ se calcula por krigeado: procedimiento explicado más adelante.

Modelado de semivariogramas: Validación del modelo teórico (iii)

El **método de validación cruzada** (cont).

Si se repite este proceso para los n puntos, se pueden calcular n errores de validación:

$$E(xi) = Z^*(xi) - Z(xi) \qquad i = 1, 2, \ldots, n.$$

Así se van probando diferentes valores de los parámetros del semivariograma hasta que los errores de validación cumplen una serie de criterios estadísticos:

- error medio nulo,
- error cuadrático medio pequeño,
- $\mathrm{Corr}\{Z(xi), Z^*(xi)\} = 1$, ...

Modelado de semivariogramas: Validación del modelo teórico (iv)

El **Ajuste automático**

Método presentado por algunos autores, que sugieren una forma particular de aplicar el método de los mínimos cuadrados y así obtener el modelo y sus parámetros teniendo en cuenta que el modelo obtenido sea definido positivo

El ajuste realizado de forma automática no tiene por qué reportar mejores resultados en el proceso de estimación.

Distintos autores recomiendan validar el modelo seleccionado de acuerdo al estimador a utilizar.

Independientemente de la forma utilizada en la elección del modelo teórico y sus parámetros, como criterio se recomienda emplear el método de la validación cruzada con el estimador a utilizar en el proceso de estimación.

Modelado de semivariogramas: Análisis de anisotropía

Análisis sobre el comportamiento de la variabilidad del atributo en estudio.

El semivariograma describe las características de continuidad espacial de la variable regionalizada en una dirección, pero este comportamiento pueden variar según la dirección que se analice.

Se exige por este motivo un análisis del comportamiento de la continuidad en distintas direcciones:

Análisis de Anisotropía.

Modelado de semivariogramas: Análisis de anisotropía (ii)

Cuando el semivariograma calculado en diferentes direcciones (norte-sur, este-oeste, y en direcciones intermedias de 45° o de 22.5°, con tolerancia de 22.5º), muestra similar comportamiento, se dice que el fenómeno es **Isotrópico**.

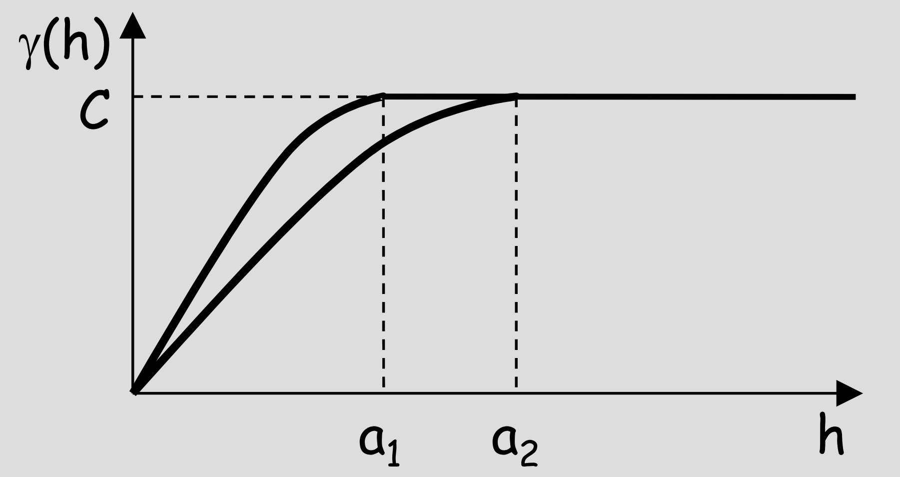
Cuando muestran diferentes comportamientos es **Anisotrópico**

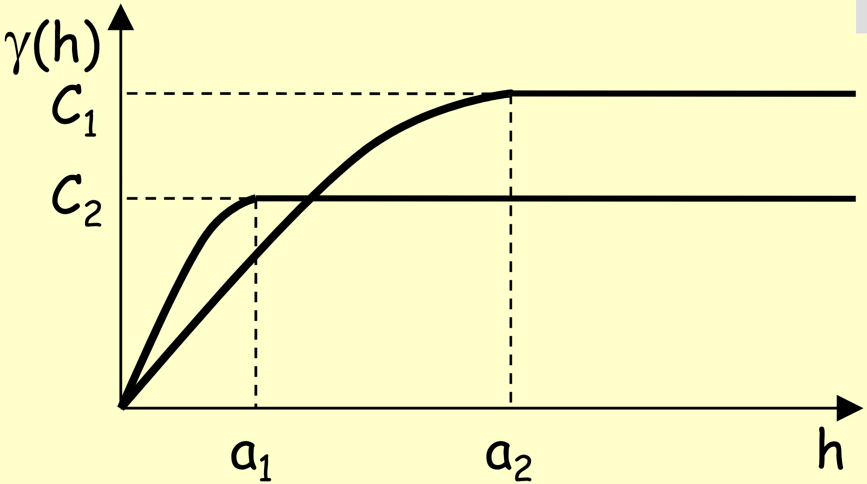
Tipos de Anisotropías:

Las más comunes son la Geométrica y la Zonal.

Modelado de semivariogramas: Análisis de anisotropía (iii)

Anisotropía Geométrica: cuando los semivariogramas en diferentes direcciones tiene la misma meseta pero distintos alcances.



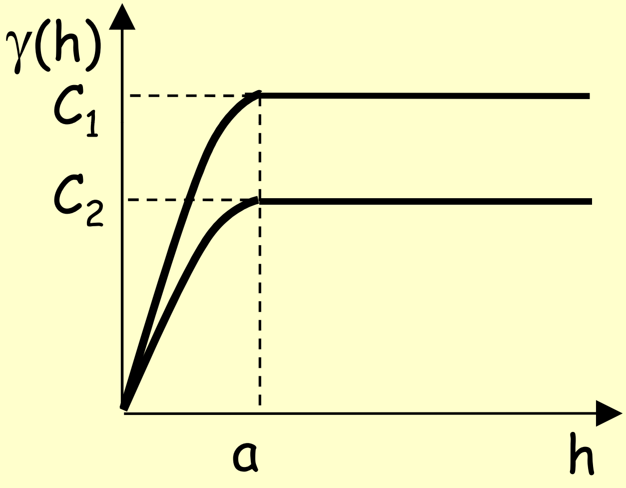


Anisotropía Zonal: cuando los semivariogramas en diferentes direcciones tiene diferentes mesetas y alcances.

En estos casos conviene realizar transformaciones de coordenadas con el objetivo de obtener modelos Isotrópicos

Modelado de semivariogramas: Análisis de anisotropía (iv)

Efecto proporcional: Cuando en el cálculo del semivariograma se detecta que existe una relación lineal entre el valor medio de las muestras usadas en el cálculo de cada $\gamma(h)$ y la desviación estándar correspondiente (heterocedasticidad).



Efecto proporcional

Este efecto se puede detectar en un gráfico de los valores de Xm contra σ, es decir, que el coeficiente de variación (σ/Xm) sea aproximadamente constante, ocurre cuando los datos presentan una distribución lognormal.

semivariograma relativo:

$$F(h) = \gamma(h)/Xm^2(h)$$

Modelado de semivariogramas:

Problemas en el modelaje de semivariogramas

- **Anisotropía geométrica:** semivariogramas direccionales con la misma meseta pero diferentes alcances; ésta puede ser corregida por una transformación linear de coordenadas que permita reducir una elipse a un circulo.

-**Anisotropía zonal:** tanto las mesetas como los alcances son diferentes para los semivariogramas direccionales, puede ser corregido separando el semivariograma en sus componentes isotrópicos horizontal y anisotrópico vertical.

- **Tendencia de los datos:** los valores medidos aumentan o disminuyen rápidamente en la zona estudiada con el aumento de la distancia. Esto puede ser resuelto aplicando polinomios a la ecuación del semivariograma (análisis de tendencia).

Modelado de semivariogramas:
Problemas en el modelaje de semivariogramas

- **Efecto proporcional:** Indica que la desviación estándar local es proporcional al cuadrado de la media local y que los datos presentan una distribución lognormal, puede ser resuelto dividiendo cada valor del semivariograma local por el cuadrado de la media local (semivariogramas relativos).

-**Estructuras anidadas**: diferentes procesos operan a diferentes escalas. Por ejemplo:

. A muy pequeñas distancias la variabilidad puede estar presente debido a cambios de una composición mineral a otra.

• A pequeñas distancias la variabilidad puede estar presente debido a errores.

• A grandes distancias la variabilidad puede estar presente debido a casos transitorios de desgaste mineral.

Se puede resolver aplicando varios modelos simultáneamente

Modelado de semivariogramas:
Problemas en el modelaje de semivariogramas

- **Efecto hueco**: indica que muy pocos pares están disponibles para la comparación a una distancia específica.
Se resuelve recuperando más casos para la distancia definida.

- **Periodicidad:** indica que el comportamiento del semivariograma repite por sí mismo periodicidades.

Ejemplo:
El valor de la meseta puede aumentar o disminuir sistemáticamente, o un caso en que los valores son tomados alternativamente a través de diferentes estratos, como piedras areniscas, esquistos, etc.

Se puede resolver si no es un enmascaramiento de la realidad.

El fenómeno es real cuando hay zonas similares que aparecen espaciadas regularmente

Estimación

Todo lo expresado hasta aquí tiene un único objetivo, conocer la información disponible para realizar estimaciones de valores desconocidos a partir, no sólo de los conocidos, sino también de su estructura de continuidad espacial

Algunos ejemplos: Triangulación ,
Inverso de la distancia

Estas dos técnicas de estimación utilizan, en el proceso de estimación, directamente los valores muestreados y refieren pesos de acuerdo a las distancias entre los datos, sin tener en cuenta la continuidad espacial de la información disponible.

Estimación:
La interpolación con método de Krige

El **krigeado** (de krigeage en francés, en inglés kriging), interpolador de la geoestadística, utiliza los resultados discutidos del análisis estructural.

El **krigeado** es una técnica de estimación que,
• proporciona el mejor estimador lineal insesgado.
(BLUE, en inglés, Best Linear Unbiased Estimator)

•proporciona un error de estimación conocido como **varianza de krigeado** que depende del modelo de variograma obtenido y de las localizaciones de los datos originales

$\downarrow$

posibilidad de hacer análisis sobre la calidad de las estimaciones.

Planteamiento del problema del krigeado

OBJETIVOS:

Generar superficies que incorporen las características estadísticas analizadas en los datos observados.

•El krigeado consiste en efectuar una ponderación, es decir, atribuir un peso a cada valor observado,
-los pesos son calculados de manera que minimice la varianza de estimación resultante, teniendo en cuenta las características geométricas del problema (Matheron, 1970).

Al minimizar la varianza de estimación se garantiza el uso óptimo de la información disponible

Ecuaciones del krigeado (i)

Se dispone de los valores muestreados $Z(x_i)$, $i=1,\ldots,n$,

- Se trata de estimar un valor de la característica observada en la zona de estudio, $Z(v)$, mediante una combinación lineal de $Z(x_i)$,

$$Z^*(v) = \sum \lambda_i Z(x_i)$$

$Z^*(v)$ es el valor estimado

λ_i son los pesos de krigeado

de modo que los λ_i sean obtenidos de tal forma que proporcione un estimador:

- insesgado $E[Z^*(v) - Z(v)] = 0$ y
- de varianza $Var[Z^*(v) - Z(v)]$ mínima

Ecuaciones del krigeado: Resultados (ii)

1. Mapa de predicciones en todas las ubicaciones de interés s_0 .

$$Z(s_0) \rightarrow Z^*(s_0) = \sum_{i=1}^{n} \lambda_i z(s_i)$$

2. Varianza estimada del error de predicción .

3. Media de los residuos de la predicción .

$$\frac{\sum_{i=1}^{n} \left[Z(s_i) - Z^*(s_i) \right]}{n} \rightarrow 0$$

Ecuaciones del krigeado: Resultados (ii)

4. Cuadrado medio del error.

$$\frac{\sum_{i=1}^{n}\left[Z(s_i) - Z^*(s_i)\right]^2}{n} \rightarrow \textit{mínimo}$$

5. Error cuadrático medio estandarizado .

$$\frac{1}{n}\left[\frac{\sum_{i=1}^{n}\left[Z(s_i) - Z^*(s_i)\right]^2}{\sigma^2(s_i)}\right] \rightarrow 1$$

Ecuaciones del krigeado: Resultados (ii)

6. Coeficiente de correlación de Pearson entre observaciones y predicciones .

$$r\,(z(s_i);z^*(s_i)) \rightarrow 1$$

7. También es necesario comparar:

   > Máximo y mínimo de los errores de predicción

   > Diagramas de cajas, tanto de las predicciones como de los errores, así como de todas las variables medidas.

Ecuaciones del krigeado: Resultados (ii)

DIFERENTES TIPOS DE PREDICCIÓN

Valores de la variable considerada	→ Prediction Map
Residuos de cada predicción	→ Prediction Std. Error Map
Probabilidad de que la variable supere o no supere un umbral definido	→ Probability Map
Cuantiles	→ Quantil Map

Ecuaciones del krigeado (iii)

Teniendo en cuenta las hipótesis de la geoestadística: estacionaridad (sin tendencia, autocovarianza y autocorrelación espacial dependiendo sólo de la distancia entre puntos)

- **Krigeado Simple:** para una función aleatoria estacionaria, de esperanza nula o conocida,

Estimador: $Z^*(v) = \sum \lambda i \; Z(xi) + m(1 - \sum \lambda i).$

Sistema: $\sum \lambda i \; C(xi, xj) = C(xj, v); \quad j = 1,\dots,n$

Varianza de krigeado:

$$\sigma^2 = C(v,v) - \sum \lambda i \; C(xi, v)$$

Ecuaciones del krigeado (iii)

Krigeado o kriging Simple

- Estacionariedad de segundo orden
- Media constante y conocida
- Dificilmente aplicable por requerir demasiado conocimiento de la variable
- Siempre insesgado (sin restricciones)
- No es preciso que cumpla que $\sum_{i=1}^{n} \lambda_i = 1$

Ecuaciones del krigeado (iii)

Para una función aleatoria estacionaria de esperanza desconocida

- **Krigeado Ordinario**

En términos de la covarianza

Estimador: $Z^*(v) = \sum \lambda i\ Z(xi)$

Sistema: $\sum \lambda i\ C(xi,\ xj) - \mu = C(xj,v);$

$i,j = 1,\ldots,n \qquad \sum \lambda i = 1$

Varianza de krigeado:

$$\sigma^2 = C(v,v) - \sum \lambda i\ C(xi,\ v) + \mu$$

Ecuaciones del krigeado (iv)

Para una función aleatoria estacionaria de esperanza desconocida

- **Krigeado Ordinario (cont)**

En términos del semivariograma

Estimador: $Z^*(v) = \sum \lambda_i Z(x_i)$

Sistema: $\sum \lambda_i \gamma(x_i, x_j) + \mu = \gamma(x_j, v)$

$j = 1,\ldots,n; \qquad \sum \lambda_i = 1$

Varianza de krigeado:

$$\sigma^2 = \sum \lambda_i (x_i, v) - \gamma(v,v) + \mu$$

El sistema krigeado y la varianza de krigeado dependen sólo: del modelo estructural C(h) o γ(h) obtenido y de la geometría del soporte de observación.

Krigeado Ordinario

Media constante pero desconocida

Puede usarse con estacionariedad de segundo orden o estacionariedad intrínseca

Para ser un estimador insesgado (centrado) se debe cumplir que $\sum \lambda_i = 1$

El caso no estacionario, Krigeado Universal (KU)

Tendencia en los datos: los valores medidos aumentan o diminuyen en alguna dirección en el área de estudio.

Modelar la tendencia o deriva mediante una componente determinística función polinomial de las coordenadas:

$$m(x) = \sum_{l=0}^{K} a_l f^l(x)$$

a_l son coeficientes y f^l es la función que describe la tendencia.

$$Z(x) = m(x) + \varepsilon(x).$$

componente estocástico $\varepsilon(x)$

El caso no estacionario, Krigeado Universal (KU)

Así pueden obtenerse derivas simples (para una deriva simple el KU se reduce al Krigeado Ordinario) , lineales, cuadráticas, etc.,

El sistema de KU:

$$\sum_{\beta=1}^{N} \lambda_\beta \gamma(x_\alpha, x_\beta) + \sum_{l=0}^{K} a_l f^l(x_\beta) = \gamma(x_\alpha, x_o)$$

$$\sum_{\alpha=1}^{N} \lambda_\alpha f^l(x_\alpha) = f^l(x_o)$$

Con varianza de estimación:

$$\sigma_{KU}^2 = \sum_{\alpha=1}^{N} \lambda_\alpha \gamma(x_\alpha, x_\beta) + \sum_{l=0}^{K} a_l f^l(x_o)$$

El caso multivariante

Los conceptos presentados hasta aquí, extendidos a más de una variable, se denominan Geoestadística Multivariada

Situaciones prácticas de variables de interés, insuficientemente muestreadas, pero que se conoce su correlación con otras variables en la zona de estudio.

Utilizando esta correlación es posible estimar una variable de interés a partir de la información de la propia variable además de las correlacionadas con ellas.

El Co-Krigeado (o cokriging) es una extensión o generalización del krigeado cuando más de una de las variables disponibles guardan relación entre sí.

El caso multivariante:Cokriging

Se requiere conocimiento:

- del modelo de semivariograma de cada una de las variables
- del semivariograma cruzado entre las variables

Semivariograma cruzado

$$\gamma_{AB}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \left[Z_A(x_i) - Z_A(x_i + h) \right] \left[Z_B(x_i) - Z_B(x_i + h) \right]$$

Z_A y Z_B son variables correlacionadas, Z_A la variable de interés y Z_B la variable auxiliar o secundaria.

El semivariograma directo toma sólo valores positivos, el cruzado puede tomar valores negativos (correlación inversa entre las variables)

Geoestadística no Lineal

En ocasiones aparecen situaciones con características que las técnicas lineales no permiten modelar,

> por ejemplo, datos con alta asimetría.

En estos casos se pueden realizar transformación a los datos, y obtener configuraciones de estos que si pueden ser explicados por el krigeado.

Krigeado de Indicadores
Krigeado Disyuntivo
Krigeado de Probabilidades
Krigeado Lognormal

Geoestadística no Lineal

La idea de estos procedimientos es

1. Realizar transformaciones en los datos originales hasta encontrar homogeneidad en la información

2. Utilizar alguna técnica de Krigeado

3. Realizar la transformación inversa.

Geoestadística no Lineal (i)

KRIGIN INDICADOR

- La variable de interés es el cumplimiento o no de una condición.
- Hay que definir un umbral
- No tiene supuestos, por lo que se clasifica como un método de estimación no paramétrico.
- Se estima la probabilidad de que en cada punto se cumpla o no la condición.

Geoestadística no Lineal (ii)

KRIGIN INDICADOR

Se redefine la variable a través de una variable indicadora

$$I(s_i, z_l) = \begin{cases} 1 & \text{Si } Z(s_i) \leq z_l \\ 0 & \text{Otro caso} \end{cases}$$

<u>Sus propiedades:</u>

$$E(I(s_0, z_l)) = 1\,\Pr(I(s_0, z_l) = 1) + 0\,\Pr(I(s_0, z_l) = 0)$$

$$\Pr(I(s_0, z_l) = 1) = \Pr(Z(s_0) \leq z_l) = F(z_l)$$

Geoestadística no Lineal (iii)

KRIGIN INDICADOR

Además:

$$0 \leq I^{*}\left(s_0, z_l\right) \leq 1$$

$$I^{*}\left(s_0, z_l\right) \leq I^{*}\left(s_0, z_l'\right) \qquad \text{cuando}$$

$$z_l \leq z_l'$$

FUENTES

http://descargas.cervantesvirtual.com/servlet/SirveObras/46860175104026839600080/006458_8.pdf Cap.7: Sistemas de Información Geográfica: Pasado, presente y futuro (tesis doctoral)

www.geogra.uah.es/~joaquin/curso-quito/**SIG**-OdelT.pdf

www.monografías .com. Elementos de Geoestadística. CUADOR GIL, J.Q. Universidad de Pinar del Río (Cuba).