

DEPENDENCIA ESPACIAL



Concepción González García (2008)

DEPENDENCIA ESPACIAL



Análisis Exploratorio de Datos Espaciales:

OBJETIVO:

> Revisar las herramientas para el análisis de dependencias de las variables de las que se han obtenido datos,
Como paso previo a la aplicación de técnicas de interpolación espacial.

Análisis de dependencias



Los procesos con variables dependientes que evolucionan en el tiempo y/o espacio requieren técnicas propias de análisis para modelizarlas y obtener predicciones de valores futuros, (modelos dinámicos**).**

Teoría estadística para el estudio de modelos dinámicos (i)

- Procesos estocásticos, aleatorios o probabilísticos.

- "***Estocástico***" (del griego *stokhastes* = adivino) ~ lo que está ligado al azar.

Estudio de dependencias

Estudio de la variación del proceso entre dos puntos de T

Teoría estadística para el estudio de modelos dinámicos (ii)

Un *proceso estocástico* es un conjunto (o familia) de variables aleatorias⁽¹⁾ indexadas por un conjunto T ,

$$\{ X_t \text{ ó } X(t) / t \in T \}, T \text{ de } R^n$$

*Si T es de R , será variación en una dimensión.
(caso de las series de tiempo)*

(1) una variable aleatoria (VA) es una variable que puede tomar ciertos valores de acuerdo a cierta distribución de probabilidades.

Términos en el tratamiento de la variación espacial

Procesos estacionarios:

Atendiendo a su comportamiento al movernos sobre el conjunto de índices:

• *Bajo traslaciones* : La covarianza sólo depende de h (vector que une s y t)

Bajo rotaciones o isótropo: la variación entre dos variables X_s y X_t sólo depende de la distancia entre s y t .

No estacionaridad: por variaciones de la dependencia según la dirección : **anisotropía**

Dependencia espacial

> La dependencia espacial se considera, desde el punto de vista de la Geoestadística, como la práctica de la teoría de las **variables regionalizadas**.

Variable regionalizada:

-Es una variable aleatoria caracterizada, además de por el conjunto de valores que puede tomar, por su posición en el espacio.

-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.

Variable regionalizada (muestra)

El valor medido en cada punto x_i
es considerado como una realización $z(x_i)$
de una VA $Z(x_i)$ cuya media es $m(x_i)$.

El conjunto de todas las mediciones $z(x)$ en el área de estudio de la variable regionalizada se considera como una realización particular del conjunto de variables aleatorias:

$$(Z(x), x \in \text{área de estudio})$$

Conceptos de Variable regionalizada (i)

Región: se refiere al espacio en el cual existe y se estudia el fenómeno natural, geográfico o sociológico.

Localización: Es el punto de una región en la cual se define una variable aleatoria regionalizada.

Soporte Geométrico: determinado por el elemento físico sobre el cual se realiza la determinación de la variable aleatoria regionalizada, esto es, la muestra unitaria, sobre la cual estudiaremos el atributo de interés.

Conceptos de Variable regionalizada (ii)

Momentos de primer orden:

Si la función de distribución de $Z(x_i)$ tiene una media definida, será una función de la localización x_i

$$m(x_i) = E\{Z(x_i)\}$$

Momento de segundo orden:

Si la varianza (Var) de $Z(x_i)$ existe, entonces se define como el momento de segundo orden y será también una función de la localización x_i

$$\text{Var} \{Z(x_i)\} = E\{[Z(x_i) - m(x_i)]^2\}$$

Conceptos de Variable regionalizada (iii)

Si la varianza de las variables $Z(x_i)$ y $Z(x_j)$ existe entonces su covarianza (Cov) también existe y es función de las localizaciones x_i y x_j

$$\text{Cov}[Z(x_i), Z(x_j)] = E\{[Z(x_i) - m(x_i)][Z(x_j) - m(x_j)]\}$$

$$\text{si } x_i = x_j ; \quad \text{Cov}[Z(x_i), Z(x_i)] = \text{Var} \{Z(x_i)\}$$

Conceptos de Variable regionalizada (iv)

La **función variograma** o función estructural se define como la varianza de la diferencia $Z(x_i) - Z(x_j)$

$$2 \gamma(x_i, x_j) = \text{Var}\{Z(x_i) - Z(x_j)\} = 2 \gamma(x_i, x_j)$$

El semivariograma $\gamma(x_i, x_j)$ se define como:

$$\gamma(x_i, x_j) = \frac{1}{2} E \left\{ [Z(x_i) - Z(x_j)]^2 \right\}$$

O bien, $\gamma(x_i, x_j) = \frac{1}{2} \text{Var}\{Z(x_i) - Z(x_j)\}$

Conceptos de Variable regionalizada (v)

Cuando la función es **estrictamente estacionaria** (la función de distribución del vector aleatorio es invariante respecto a cualquier traslación h del vector),

el semivariograma también es estacionario y se cumple que:

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

h = distancia entre puntos (lag)

Conceptos de Variable regionalizada (vi)

También se puede definir el correlograma estandarizando, la covarianza para los valores

$x_i - x_j = h = 0$ como:

$$\rho(h) = C(h)/C(0) \quad -1 \leq \rho \leq 1$$

donde:

$C(h)$ es la covarianza a la distancia h ,
 $C(0)$ es la covarianza en el origen.

Existen relaciones entre estas medidas de dependencia:

$$\gamma(h) = C(0) - C(h) \quad \text{con } \gamma(0) = 0$$

$$\rho(h) = 1 - \gamma(h)/C(0)$$

Conceptos de Variable regionalizada (vii)

- Cuando la covarianza se expresa en función de h (distancia entre dos puntos) se llama **autocovarianza**:

$$\gamma(h) = C(s_i, s_{i+h}), \quad \forall i \in T$$

La distancia h de separación se llama **retardo (lag)**.

Hipótesis de la Geoestadística (i)

La Geoestadística se construye asumiendo condiciones de estacionaridad, debido a que la forma en que se presenta la información es muy diversa.

Estacionaridad Estricta: si la función de distribución de probabilidades de las variables aleatorias regionalizadas $Z(x_i)$ es la misma, independiente de la localización x_i

Estacionaridad de Segundo Orden:
más frecuente en la práctica, supone que

- 1) $E\{Z(x_i)\} = m$, existe y no depende de la localización x_i .
- 2) La función covarianza, $Cov\{Z(x_i) - Z(x_j)\}$, exista y sólo dependa de la longitud del vector $h = x_i - x_j$ o sea.

$$C(h) = Cov\{Z(x_i), Z(x_j)\} = E\{Z(x_i), Z(x_i+h)\} - m^2$$

Hipótesis de la Geoestadística (ii)

Estacionaridad de Segundo Orden (cont.)

Esta hipótesis requiere la estacionaridad sólo para la media y para la función de covarianza de la variable aleatoria regionalizada.

La segunda condición implica, estacionaridad de la varianza y del variograma.

Se demuestra que,

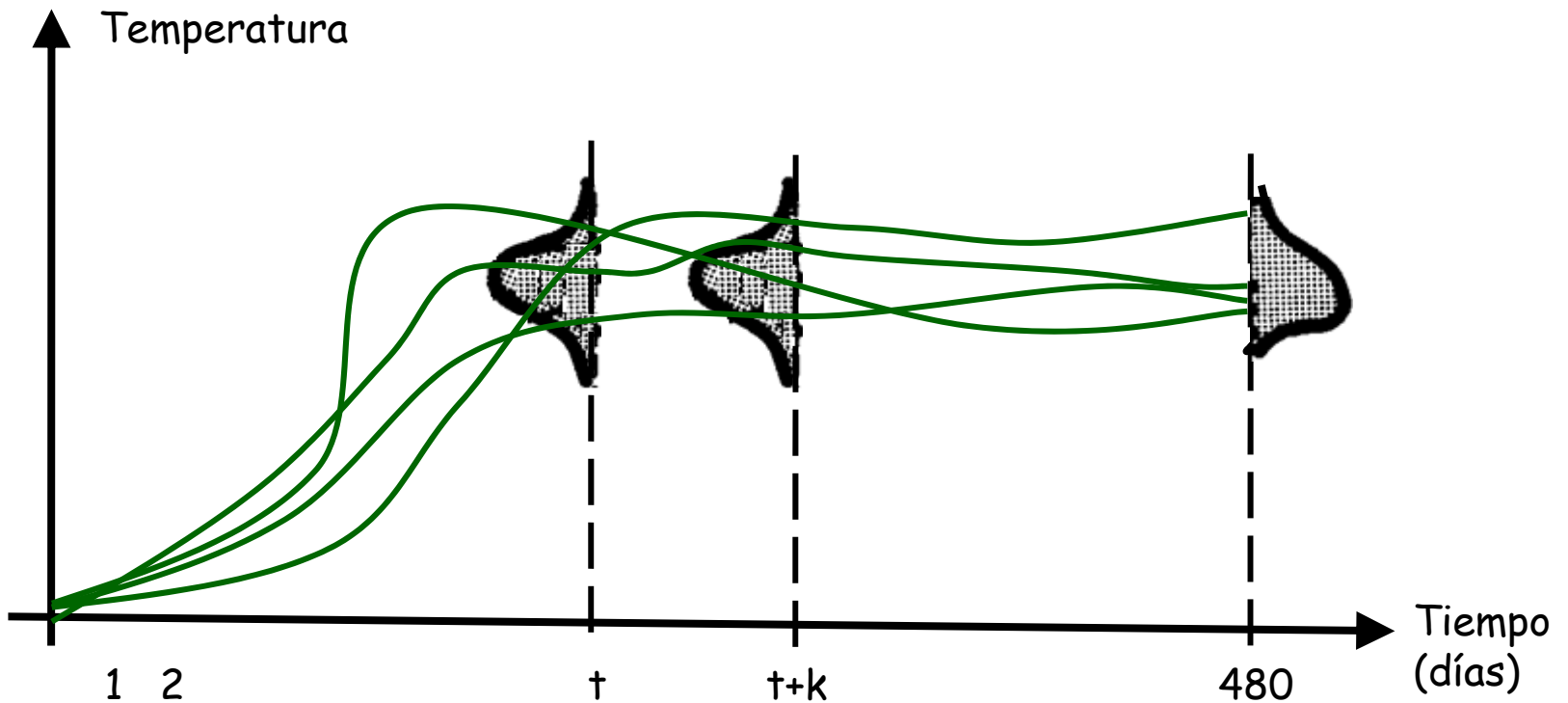
$$\gamma(h) = C(0) - C(h)$$

De aquí que $\gamma(h)$ y $C(h)$, son dos herramientas que permiten expresar la correlación entre las variables aleatorias regionalizadas $Z(x_i)$ y $Z(x_i+h)$, separadas por el vector h .

Concepto de Estacionaridad: Ejemplos gráficos en R (i)

Proceso **no estacionario** en sentido estricto

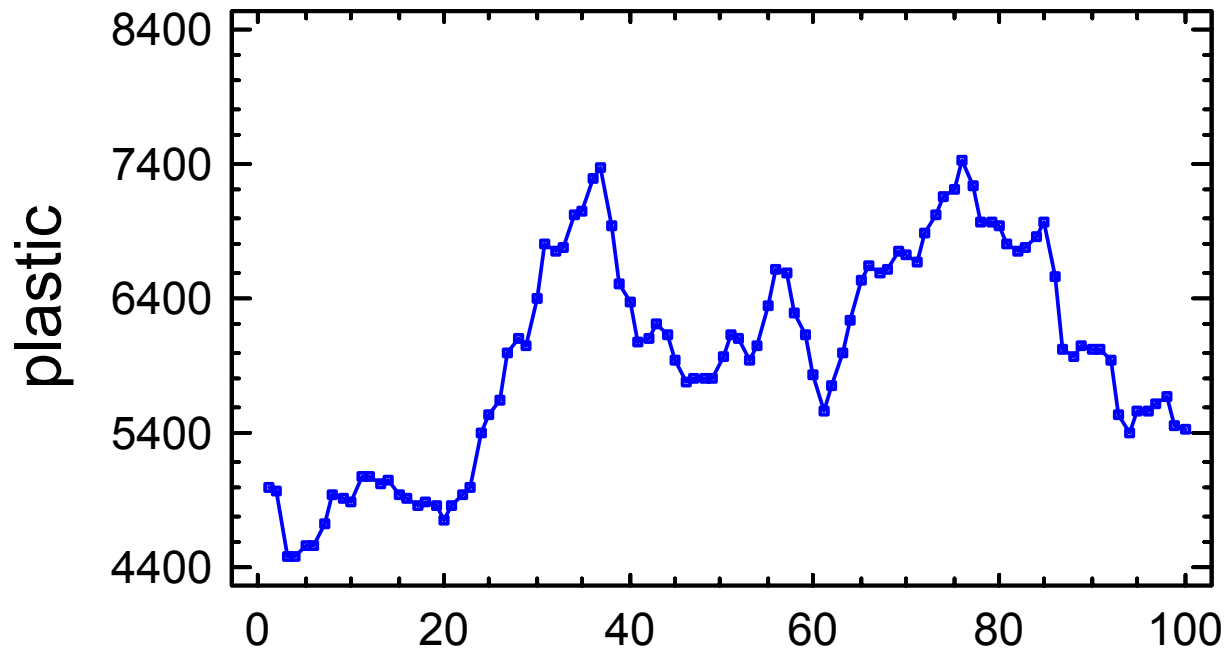
Cuatro realizaciones de un proceso estocástico



Concepto de Estacionaridad: Ejemplos gráficos en R (ii)

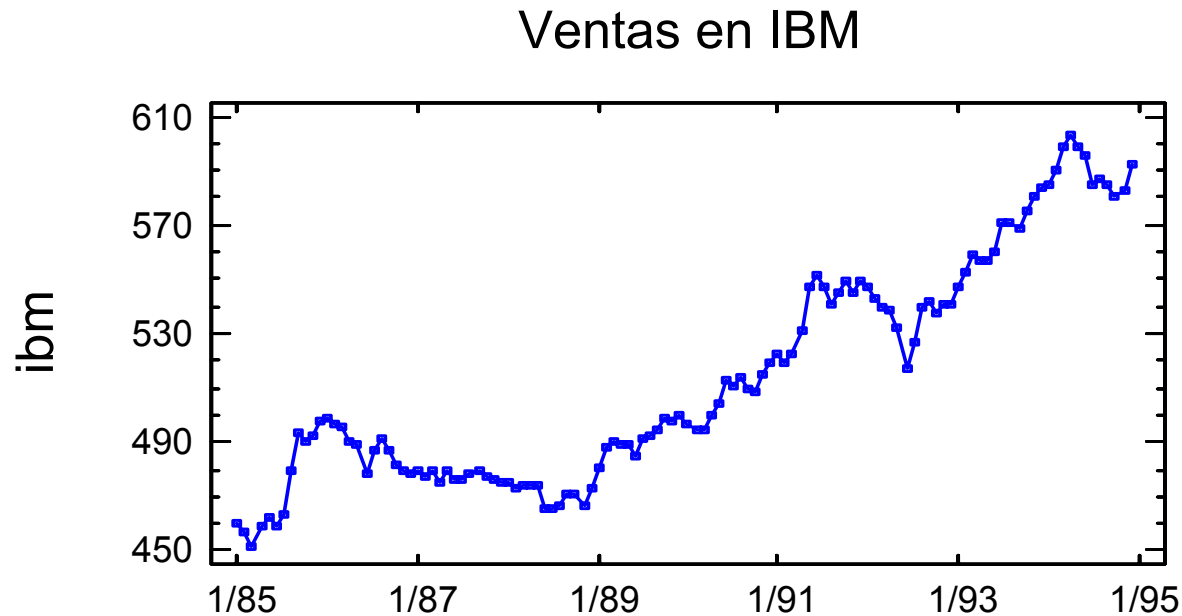
Serie *no estacionaria*:

Serie de demanda semanal de plástico



Concepto de Estacionaridad: Ejemplos gráficos en R (iii)

Serie no estacionaria:

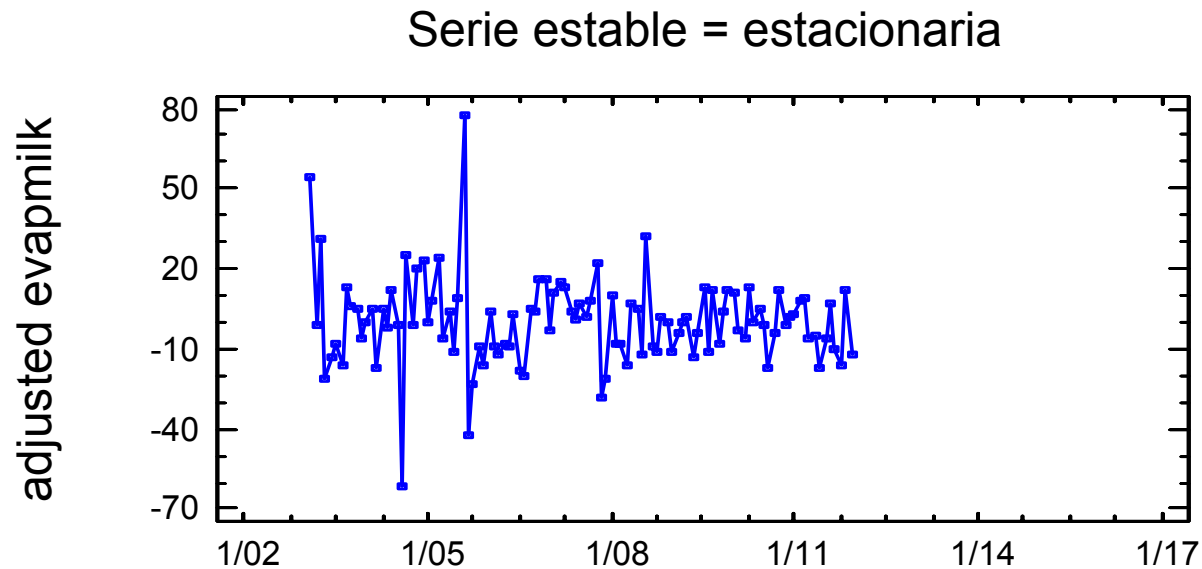


Serie no estable: Con tendencia (positiva)

Concepto de Estacionaridad: Ejemplos gráficos en R (iv)

Series estacionarias:

Los valores de la serie oscilan alrededor de un valor constante



Hipótesis de la Geoestadística (iii)

Procesos Cuasiestacionarios:

En la práctica la función estructural, covarianza o semivariograma, es sólo usada por límites $|h| \leq b$

El límite b representa la extensión de la región en la que el fenómeno estudiado conserva cierta homogeneidad del comportamiento de $Z(x_i)$.

En otros casos, b pudiera ser la magnitud de una zona homogénea y dos variables $Z(x)$ y $Z(x+h)$ no pueden ser consideradas en la misma homogeneización de su variabilidad si $|h| > b$.

Hipótesis de la Geoestadística (iv)

Estas condiciones de estacionaridad se asumen en el desarrollo teórico,

en la práctica deben ser verificadas en los datos antes de comenzar un estudio geoestadístico

para lo que se puede realizar un análisis estadístico de la información, de modo que se pueda tener cierto grado de confiabilidad en la aplicación de estos métodos.

El análisis estructural

Está compuesto por:

- El cálculo del semivariograma experimental.
- El ajuste a este de un modelo teórico conocido.

El cálculo del **semivariograma experimental** es la herramienta geoestadística más importante en la determinación de las características de variabilidad y correlación espacial del fenómeno estudiado.

El semivariograma experimental (i)

El variograma se define como la media aritmética de todos los cuadrados de las diferencias entre pares de valores experimentales separados una distancia h

o lo que es lo mismo,

$$\text{Var}\{Z(x+h)-Z(x)\} = 2\gamma(h)$$

La varianza de los incrementos de la variable regionalizada en las localizaciones separadas una distancia h .

El semivariograma experimental (ii)

La función $\gamma^*(h)$ se denomina **semivariograma empírico**, puede obtenerse por la expresión.

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

donde:

$N(h)$ es el número de pares a la distancia h .

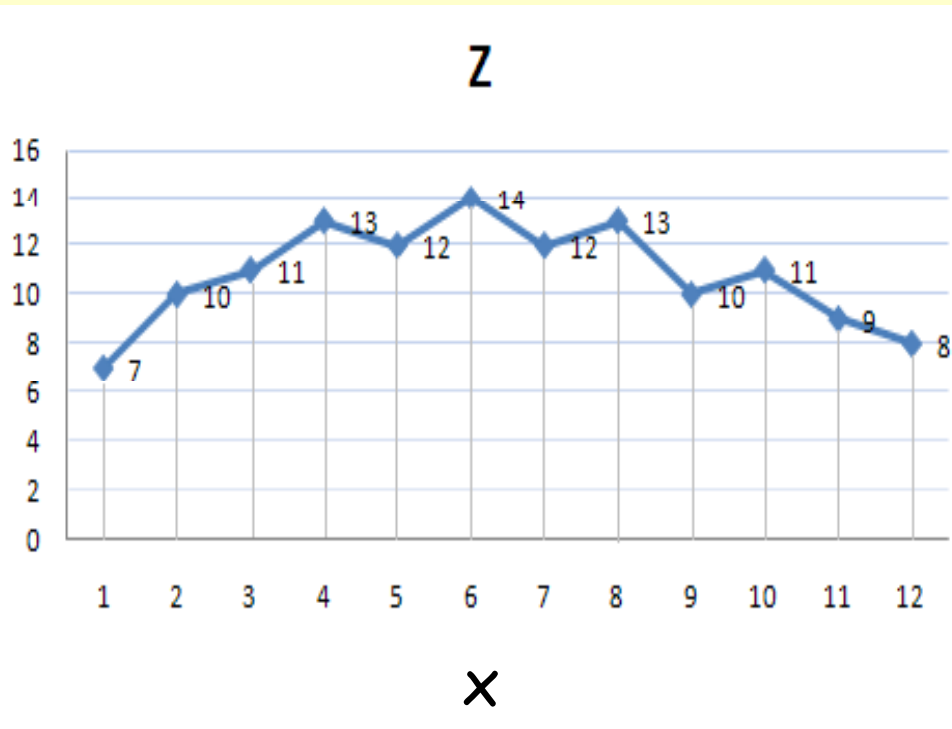
h es el incremento (lag).

$Z(x_i)$ son los valores observados

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

El semivariograma experimental: Ejemplo

Para ilustrar el proceso de cálculo se considera un ejemplo sencillo (de Samper y Carrera, 1990): Sea Z la variable definida sobre un dominio unidimensional que toma los valores :



x	Z(x)	Z(x+1)	$[Z(x+1)-Z(x)]^2$
1	7	10	9
2	10	11	1
3	11	13	4
4	13	12	1
5	12	14	4
6	14	12	4
7	12	13	1
8	13	10	9
9	10	11	1
10	11	9	4
11	9	8	1
12	8		
N(1)=11			39

$$h=1;$$

$$\sum [Z(x+1)-Z(x)]^2=39$$

El semivariograma experimental: Ejemplo

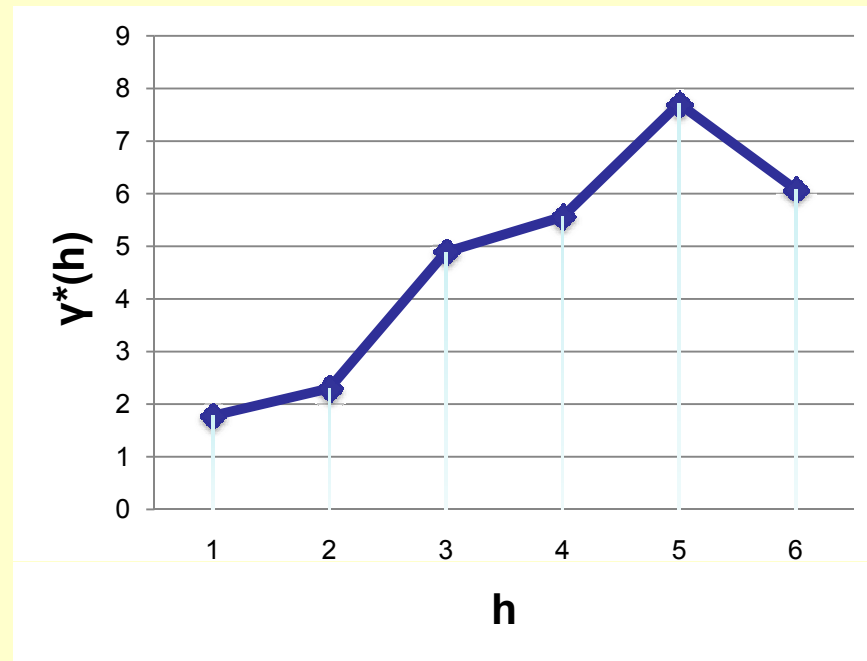
El valor estimado del semivariograma, para $h=1$, resulta ser:

$$1/2(39/11) = 1,78.$$

El proceso se repite para las parejas distantes dos unidades, etc.

El cálculo se resume en la siguiente tabla:

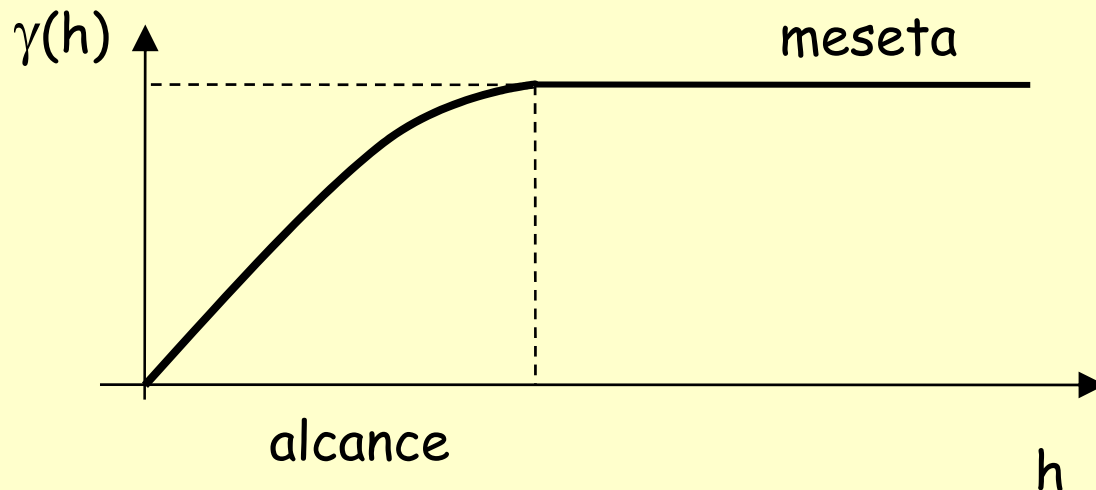
h	$N(h)$	$\sum_1^{N(h)} [Z(x_i + h) - Z(x_i)]^2$	$\gamma^*(h)$
1	11	39	1,78
2	10	46	2,30
3	9	88	4,89
4	8	89	5,56
5	7	108	7,71
6	6	73	6,08



El semivariograma experimental (iii)

El gráfico de $\gamma(h)$ tiene las siguientes características

- Pasa por el origen (para $h=0$, $\gamma(h)=0$)
- Es en general una función creciente de h .



Forma clásica del semivariograma

En la mayor parte de los casos $\gamma(h)$ crece hasta cierto límite llamado meseta, en otros casos puede crecer indefinidamente.

El semivariograma experimental (iv)

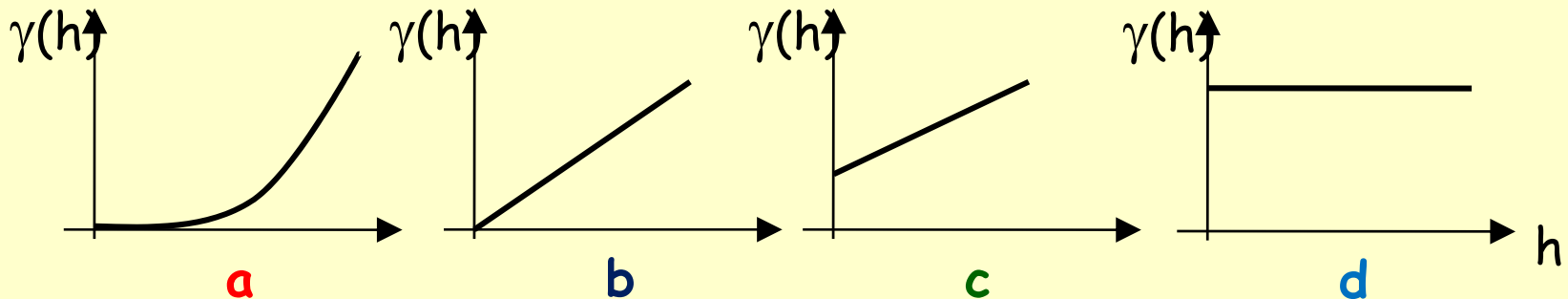
El comportamiento en el origen puede tener diferentes formas, las cuales son según distintos autores:

Parabólico: Caracteriza a una variable muy regular, siendo continua y diferenciable (a).

Lineal: Caracteriza a una variable continua, pero no diferenciable, es decir menos regular (b).

Discontinuidad en el origen: "Efecto de pepita" (c), es el caso en que $\gamma(h)$ no tiende a cero cuando h tiene a cero. Representa a una variable muy irregular.

Discontinuo puro: o "ruido blanco" (d), representa el caso de mayor discontinuidad, es el caso límite de ausencia de estructura, donde los valores de dos puntos cualesquiera no tienen correlación alguna.



El semivariograma experimental (v)

En la práctica, y especialmente cuando se trabaja en dos o tres dimensiones, las distancias entre los puntos de cada pareja son distintos y puede no haber dos parejas de puntos situados a la misma distancia. Por lo tanto, no es posible aplicar con fiabilidad la fórmula anterior.

Lo que se hace es definir una serie de valores h_j y asignar para el cálculo de $\gamma^*(h_j)$ todas las parejas cuya distancia esté contenida en el intervalo $(h_j - \Delta h_j, h_j + \Delta h_j)$. Lo más común es repetir el cálculo con varias amplitudes y elegir aquélla que permita una buena definición del semivariograma, de forma que los h_j no estén tan separados que no se pueda discernir, ni tan juntos que el número de parejas sea pequeño y las oscilaciones del semivariograma experimental resulten excesivas.

Construcción del semivariograma experimental en 2D (i)

Para datos distribuidos en dos dimensiones.

$Z(x)$ una función aleatoria con N variables aleatorias regionalizadas $Z(x_i)$
donde $x = \{x, y\}$ es la localización y $Z(x_i)$ es el valor medido correspondiente.

Aunque el programa lo va a calcular automáticamente, los primeros pasos del proceso son:

- 1.- Cálculo de la cantidad de pares de datos posibles por: $N_p = N(N-1)/2$
- 2.- Para cada par, cálculo de la distancia entre las localizaciones correspondientes

$$d_i = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2} \quad i = 1, \dots, N_p$$

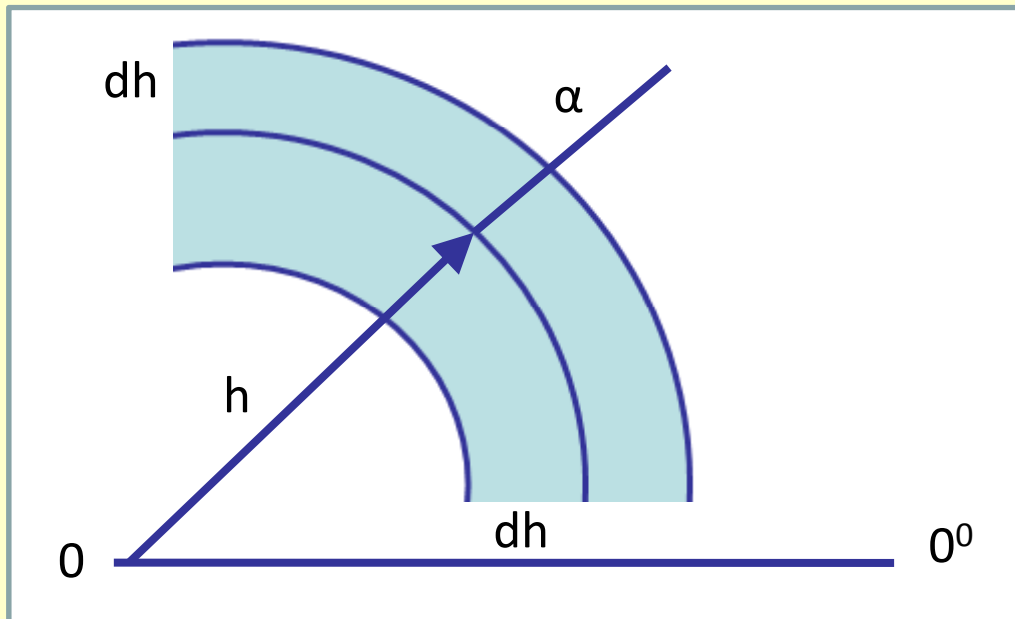
almacenando para cada i :

- P_1 : Número del primer punto del par,
- P_2 : Número del segundo punto del par,
- d : Valor de la distancia entre los dos puntos del par.
- Angulo α' que fija la dirección de la recta que pasa por los dos puntos del par.

Construcción del semivariograma experimental en 2D (ii)

Se evalúa la expresión del semivariograma para todos los pares de localizaciones separadas a la distancia h y que cumplan las condiciones siguientes :

1.- La distancia entre las localizaciones x_i y x_i+h sea mayor que $h-dh$ y menor que $h+dh$, o lo que es lo mismo, el segundo punto del par esté incluido en el espacio definido por $h-dh$ y $h+dh$ encontrándose el primer punto del par en el origen O este origen se mueve entre las muestras a analizar.

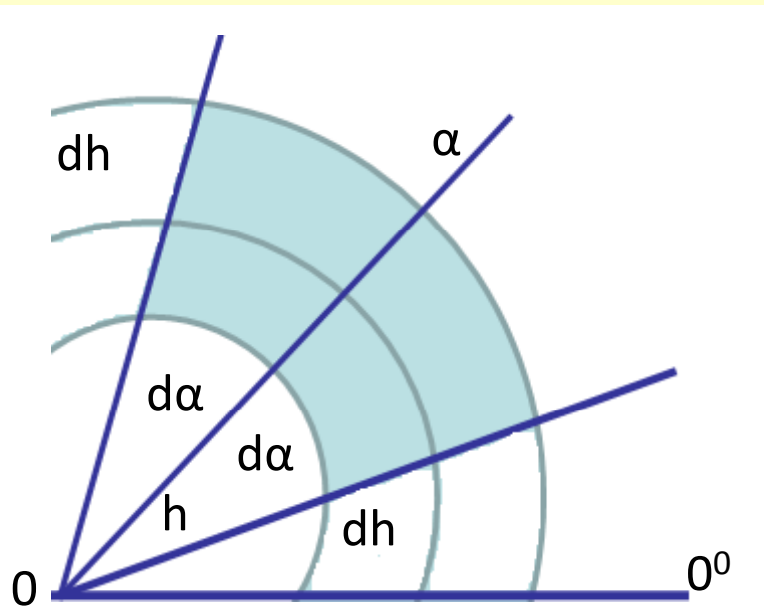


Espacio definido por la tolerancia lineal

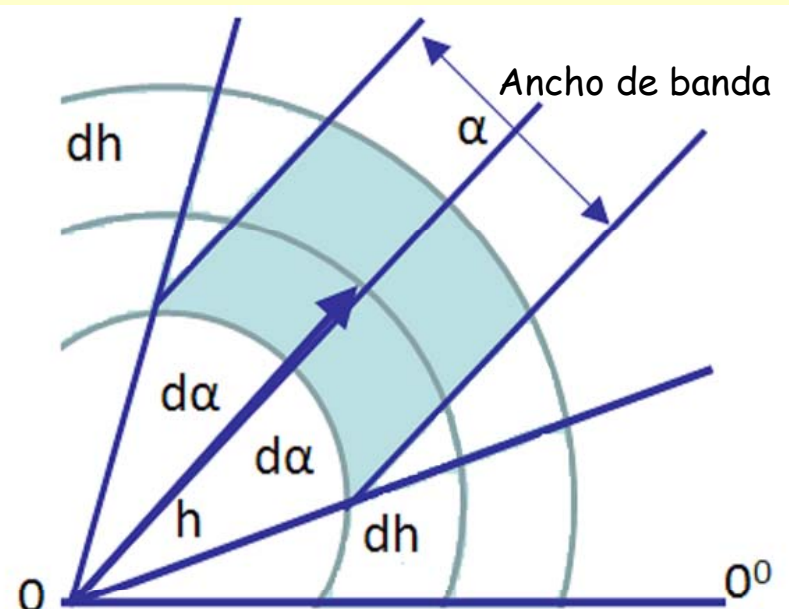
Construcción del semivariograma experimental en 2D (iii)

2.- El ángulo formado entre la línea que une los dos puntos del par y la dirección 0° debe estar incluido entre $\alpha - d\alpha$ y $\alpha + d\alpha$ (figura A).

3.- La distancia entre el segundo punto del par y la línea que define la dirección de cálculo del semivariograma no debe superar el ancho de banda (Deutsch y Journel, 1998) (figura B).



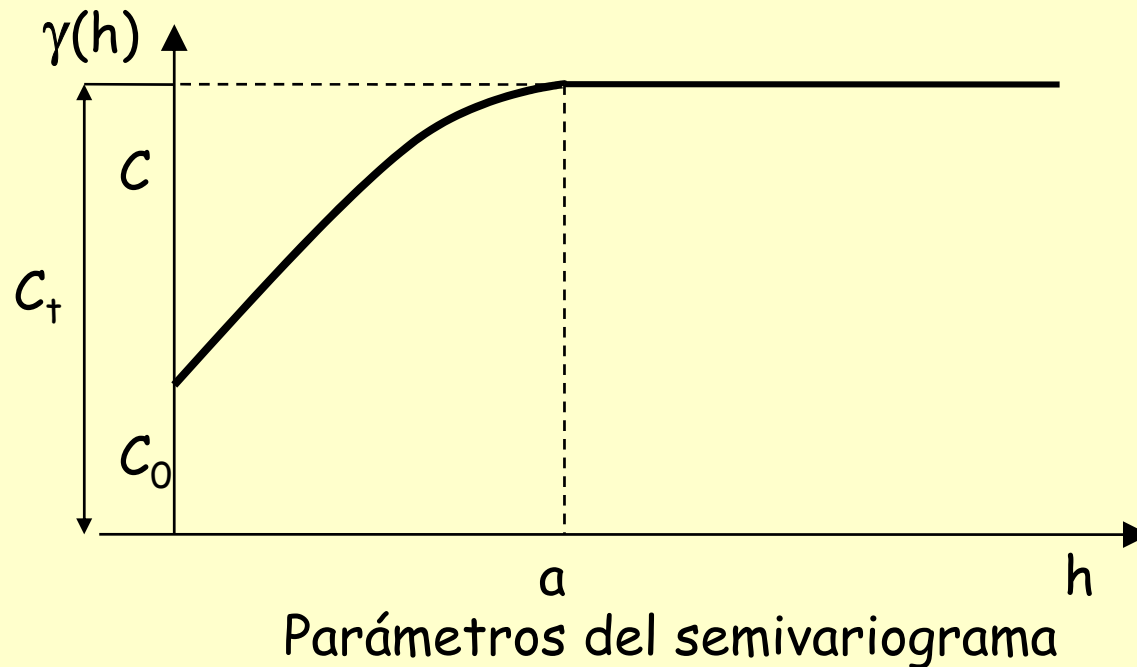
A) Espacio definido por la tolerancia lineal y angular



B) Espacio definido por las tolerancias y el ancho de banda

Construcción del semivariograma experimental en 2D (iv)

Finalmente se representan gráficamente los valores de $\gamma(h)$ en función de h , en un gráfico X-Y donde en la abscisa se representan los valores de h y en la ordenada los de $\gamma(h)$. Obteniendo así el semivariograma experimental o empírico para una dirección, incremento y tolerancias definidas.



Problemas más comunes encontrados en el cálculo de semivariograma

- El valor idóneo del incremento h
- Distribuciones con valores extremos
- La existencia de poblaciones mixtas

Por ello, es necesario el análisis exploratorio descriptivo inicial de la información.

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