

# Curso de Geoestadística

## 3. Análisis Estructural

Ramón Giraldo H.

PhD. Estadística

Profesor Departamento de Estadística  
Universidad Nacional de Colombia

# Variable Regionalizada

- Es un proceso estocástico con dominio definido en un espacio euclidiano m-dimensional  $R^m$

$$\left\{ Z(x) : x \in D \subset R^m \right\}$$

- Si  $m = 2$ ,  $Z(x)$ , a una variable aleatoria medida en un punto  $x$  del plano.
- $D$ : Fijo y Continuo.

# Estacionariedad Fuerte

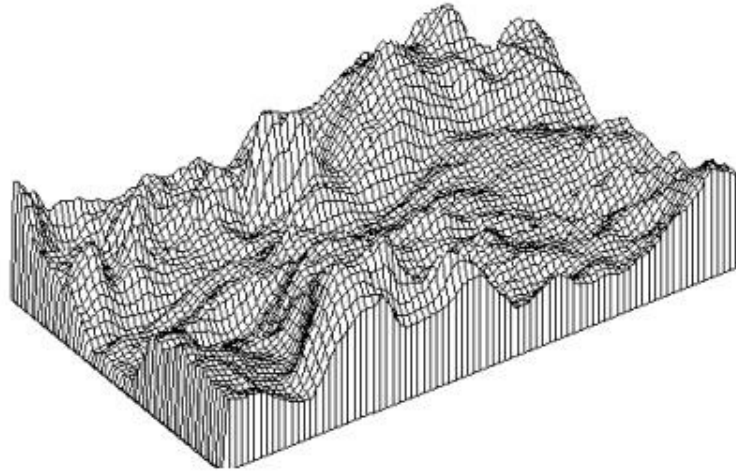
$$E[Z(x)] = m \quad \forall x \in D \subset R^m$$

$$\text{Cov}[Z(x), Z(x+h)] = C(h) < \infty$$

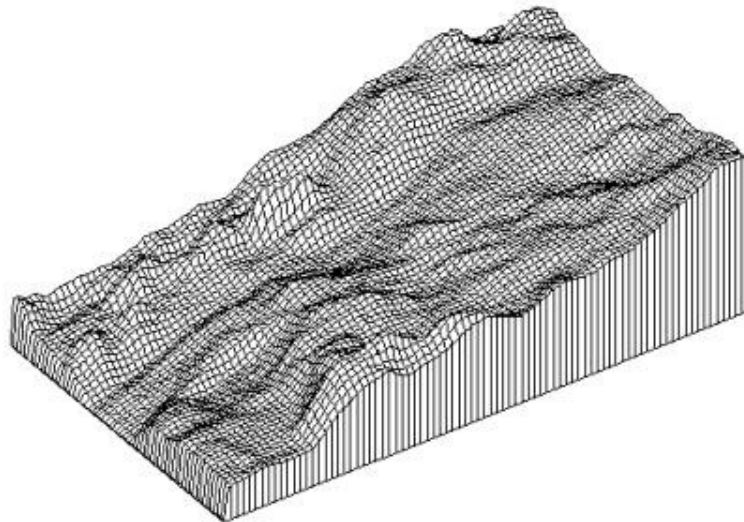
# Estacionariedad Débil

$$E[Z(x) - Z(x+h)] = 0$$

$$V[Z(x) - Z(x+h)] = 2\gamma(h) < \infty$$



estacionario



No estacionario

# Variograma

Esta función permite medir la relación que existe entre los datos de acuerdo con la cercanía ( $h$ ) entre los sitios

$$\begin{aligned} V[Z(x) - Z(x+h)] &= E[Z(x) - Z(x+h)]^2 - (E[Z(x) - Z(x+h)])^2 \\ &= E[Z(x) - Z(x+h)]^2 \end{aligned}$$

Cálculo con datos de una muestra de sitios (estimador de momentos)

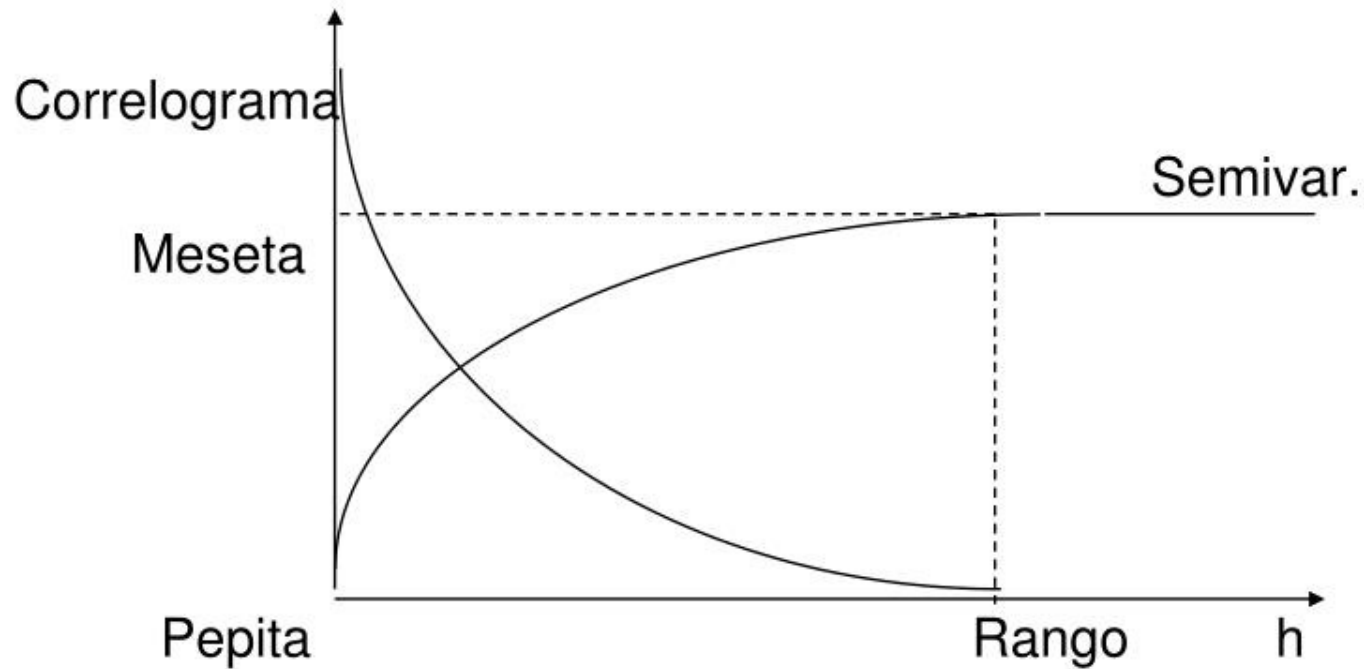
$$2\hat{\gamma}(h) = \frac{\sum (z(x+h) - z(x))^2}{n(h)}$$

## Otras Medidas de Correlación Espacial

$$C(h) = \sigma^2 - \gamma(h)$$

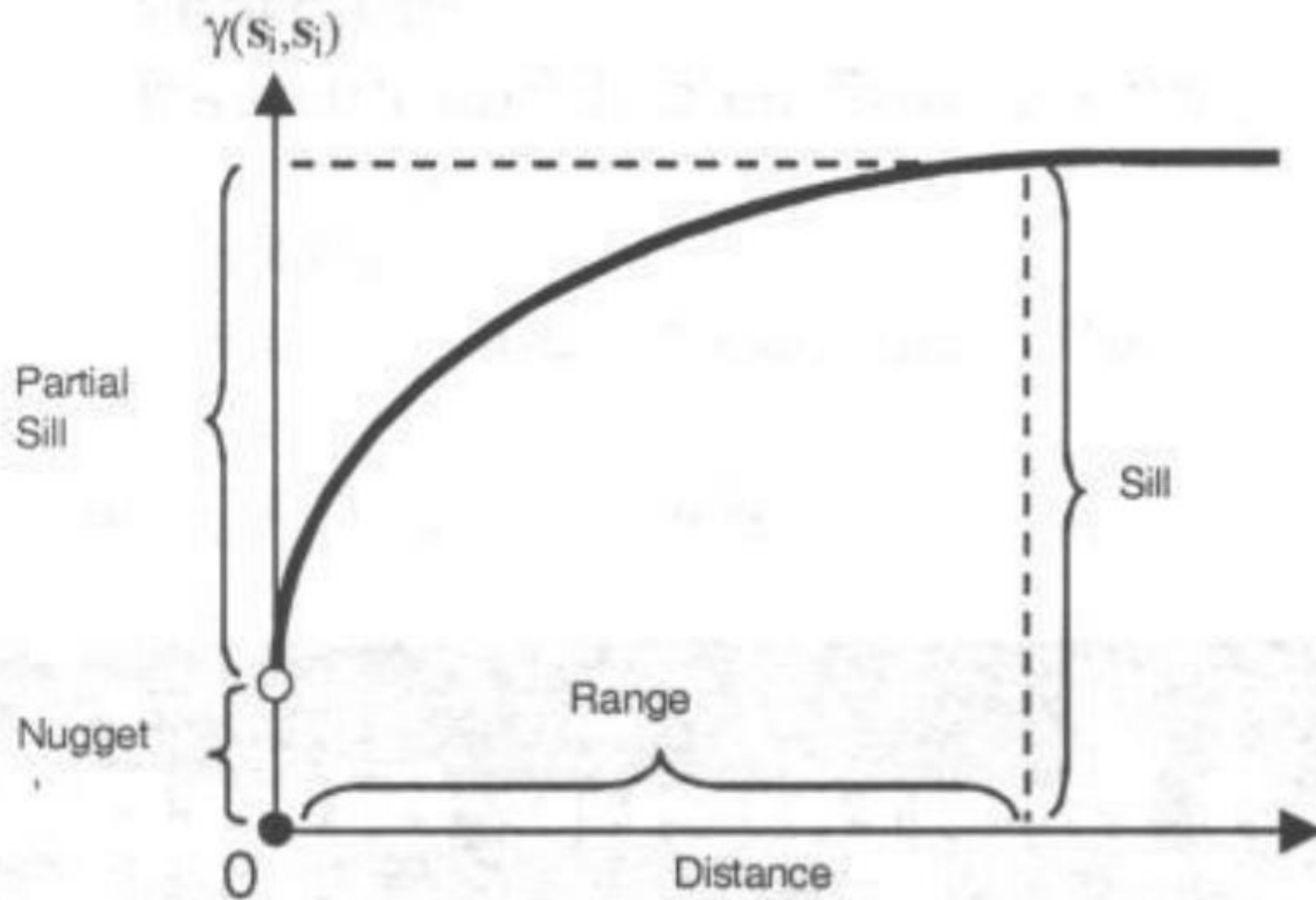
$$\rho(h) = 1 - \frac{\gamma(h)}{\sigma^2}$$

# Semivariograma y Correlograma

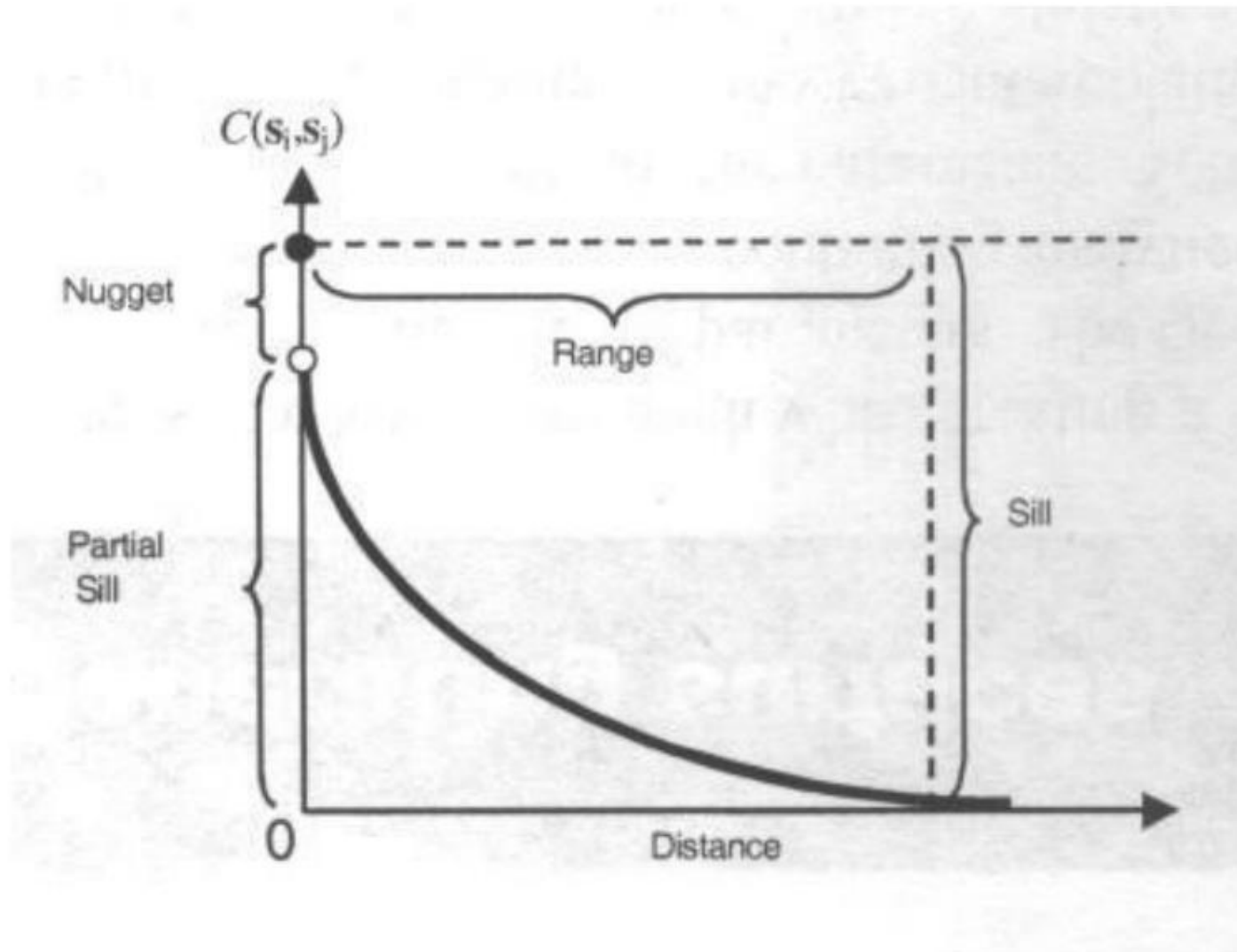




# Variograma



# Función de Covarianza



# Ilustración

44		40	42	40	39	37	36	
42		43	42	39	39	41	40	38
37	37	37	35	38	37	37	33	34
35	38		35	37	36	36	35	200
36	35	36	35	34	33	32	29	28
38	37	35		30		29	30	32
←100→								

## Cálculo del Semivariograma

$$\gamma(100) = (38 - 37)^2 + (37 - 35)^2 + \dots + (37 - 36)^2$$

$$\gamma(200) = (38 - 35)^2 + (35 - 30)^2 + \dots + (39 - 36)^2$$

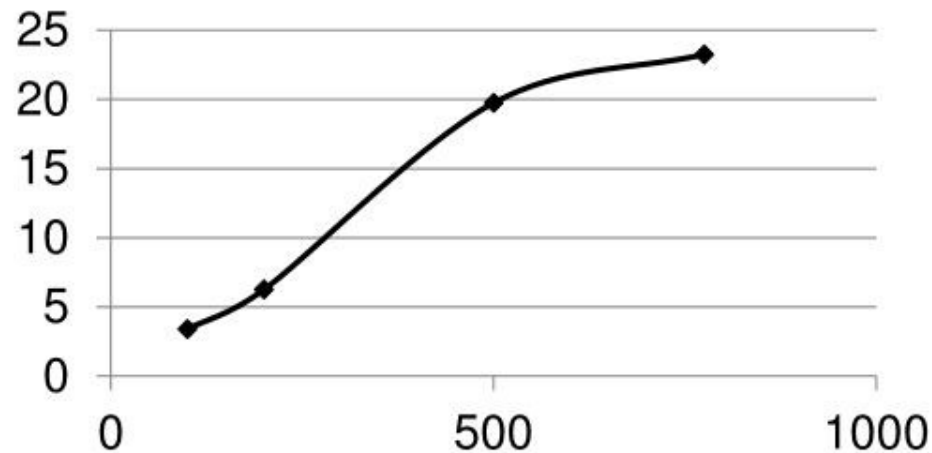
El semivariograma experimental se calcula mediante la suma de los cuadrados de las diferencias entre observaciones que se encuentran a una distancia  $h$  (en el ejemplo 100, 200, 300, etc)

Ejercicio: calcular el valor del variograma empírico a distancia 100

## Semivarianza Datos de la Ilustración

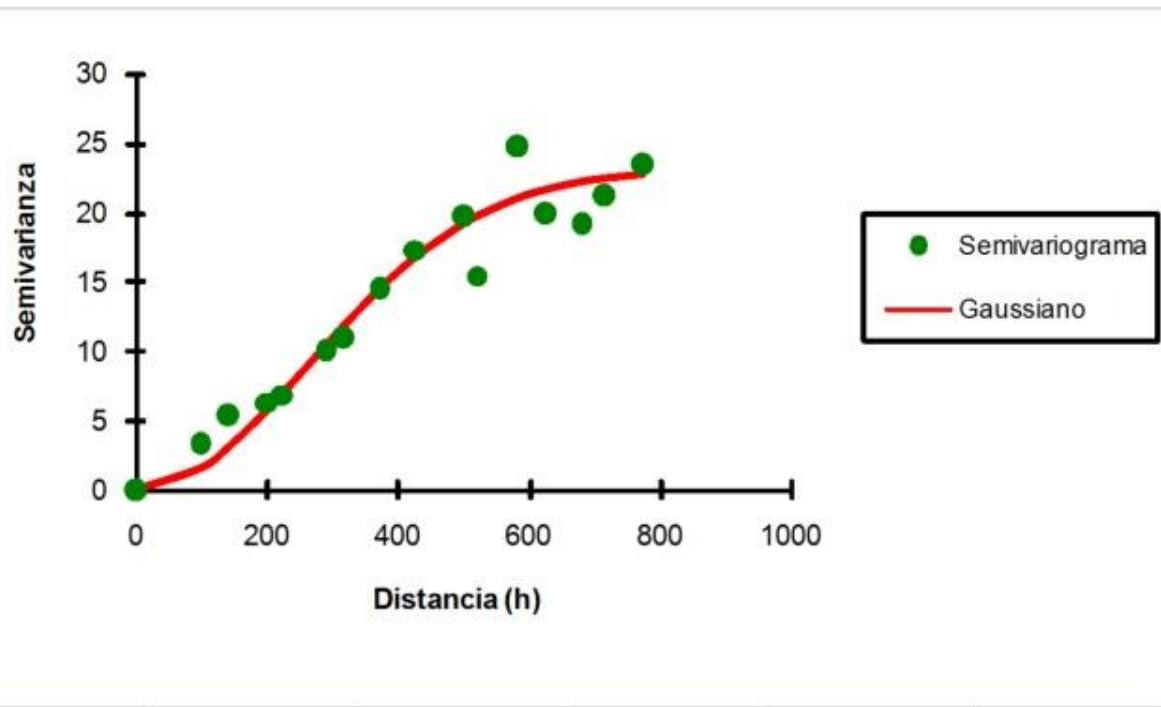
Distancia	Semivariograma
100	3.403
200	6.258
500	19.750
775	23.259

**Semivariograma**



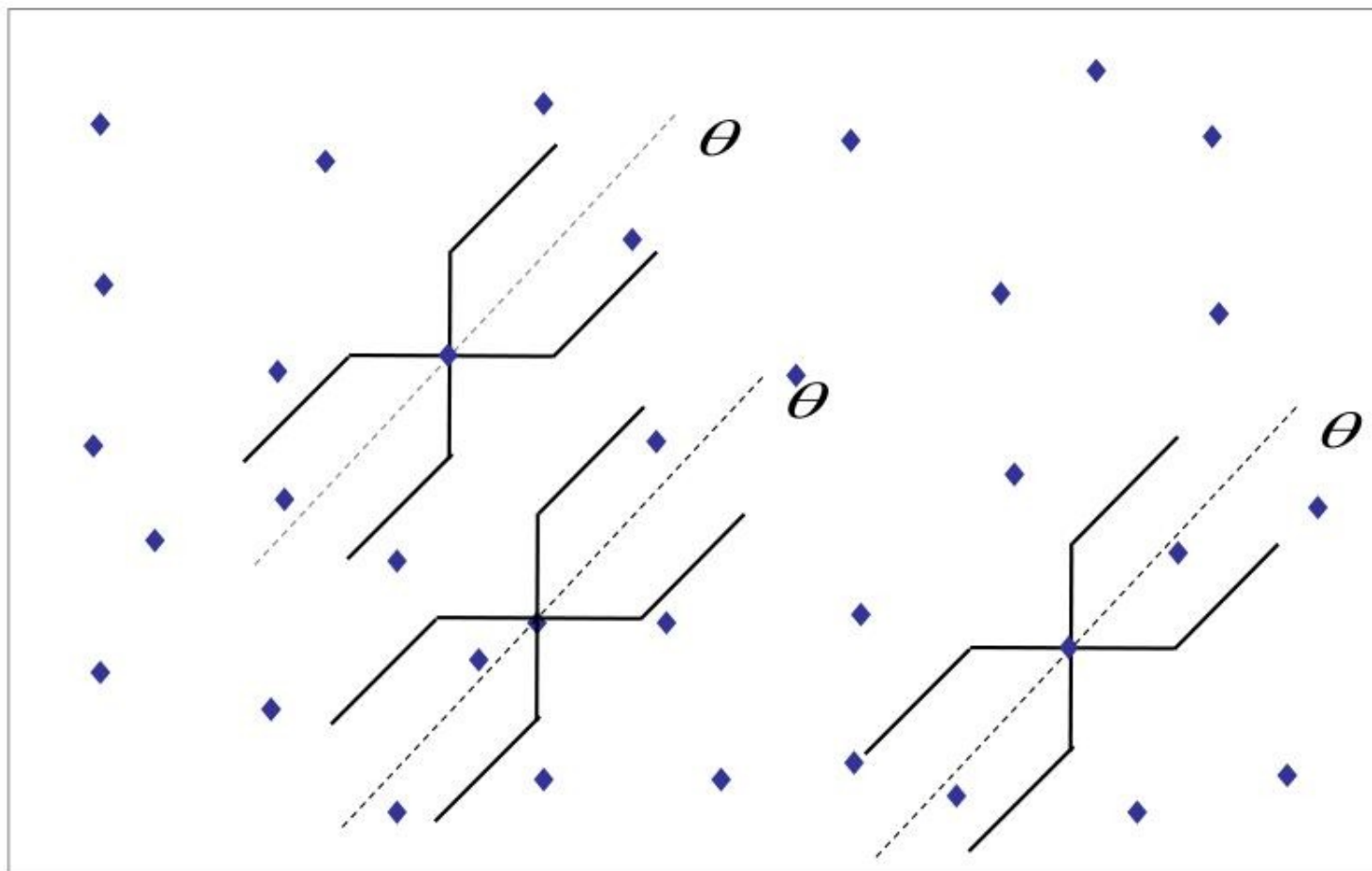
# Semivariograma Empírico y Ajuste de un modelo teórico

Modelo de Variograma Gaussiano



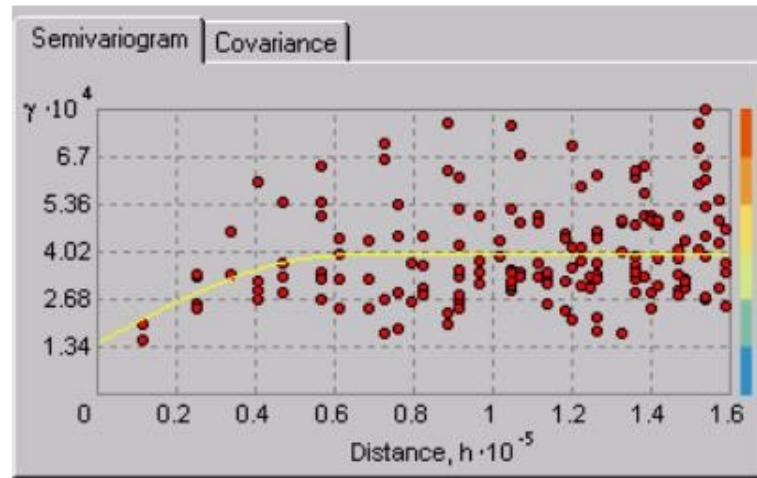
$$\gamma(h) = C_1 \left( 1 - \exp\left(\frac{-h^2}{a^2}\right) \right)$$

# Cálculo del variograma en dirección $\theta$



# Variograma “Cloud”

This is an example of a variogram produced using ArcGIS's Geostatistical Analyst.



$$\gamma_{ij} = \frac{[Z(s_i) - Z(s_j)]^2}{2}$$



# Modelos teóricos de variograma

Esférico

$$\gamma(h) = \begin{cases} C_1 \left( \frac{3}{2} \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right) & h \leq a \\ C_1 & h > a \end{cases}$$

Exponencial

$$\gamma(h) = C_1 \left( 1 - \exp\left(\frac{-3h}{a}\right) \right)$$

Gaussiano

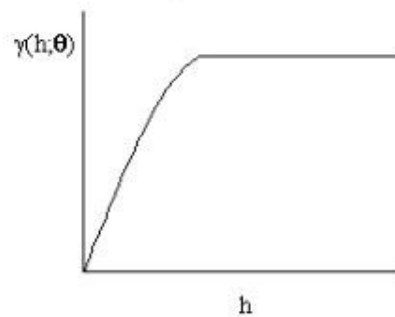
$$\gamma(h) = C_1 \left( 1 - \exp\left(\frac{-h^2}{a^2}\right) \right)$$

Efecto Nugget

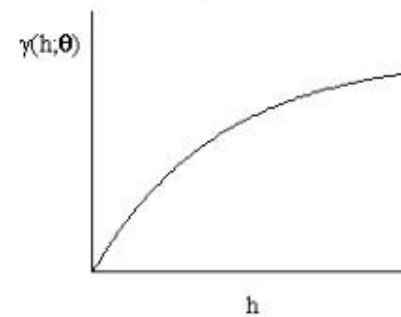
$$\gamma(h) = \begin{cases} 0 & \text{if } h = 0 \\ 1 & \text{otro caso} \end{cases}$$

# Modelos de Semivariograma

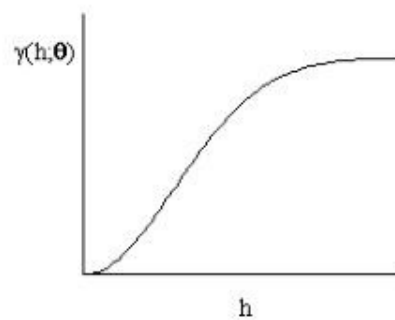
Spherical



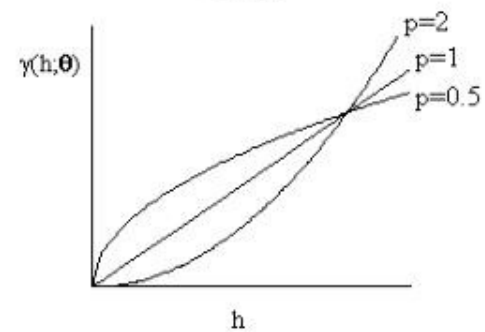
Exponential



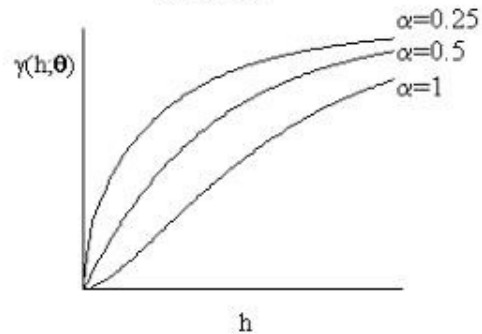
Gaussian



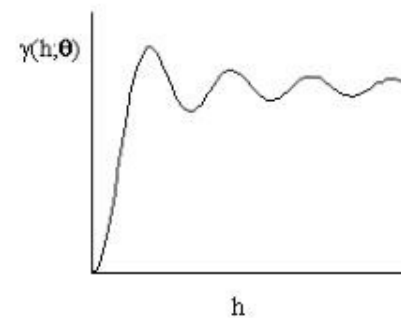
Power



K-Bessel



Cardinal-Sine



# Parámetros del Variograma Teórico

## **Pepita:**

Se denota por  $C_0$  y representa una discontinuidad puntual del semivariograma en el origen. Puede ser debido a errores de medición en la variable o a la escala de la misma. En algunas ocasiones puede ser indicativo de que parte de la estructura espacial se concentra a distancias inferiores a las observadas.

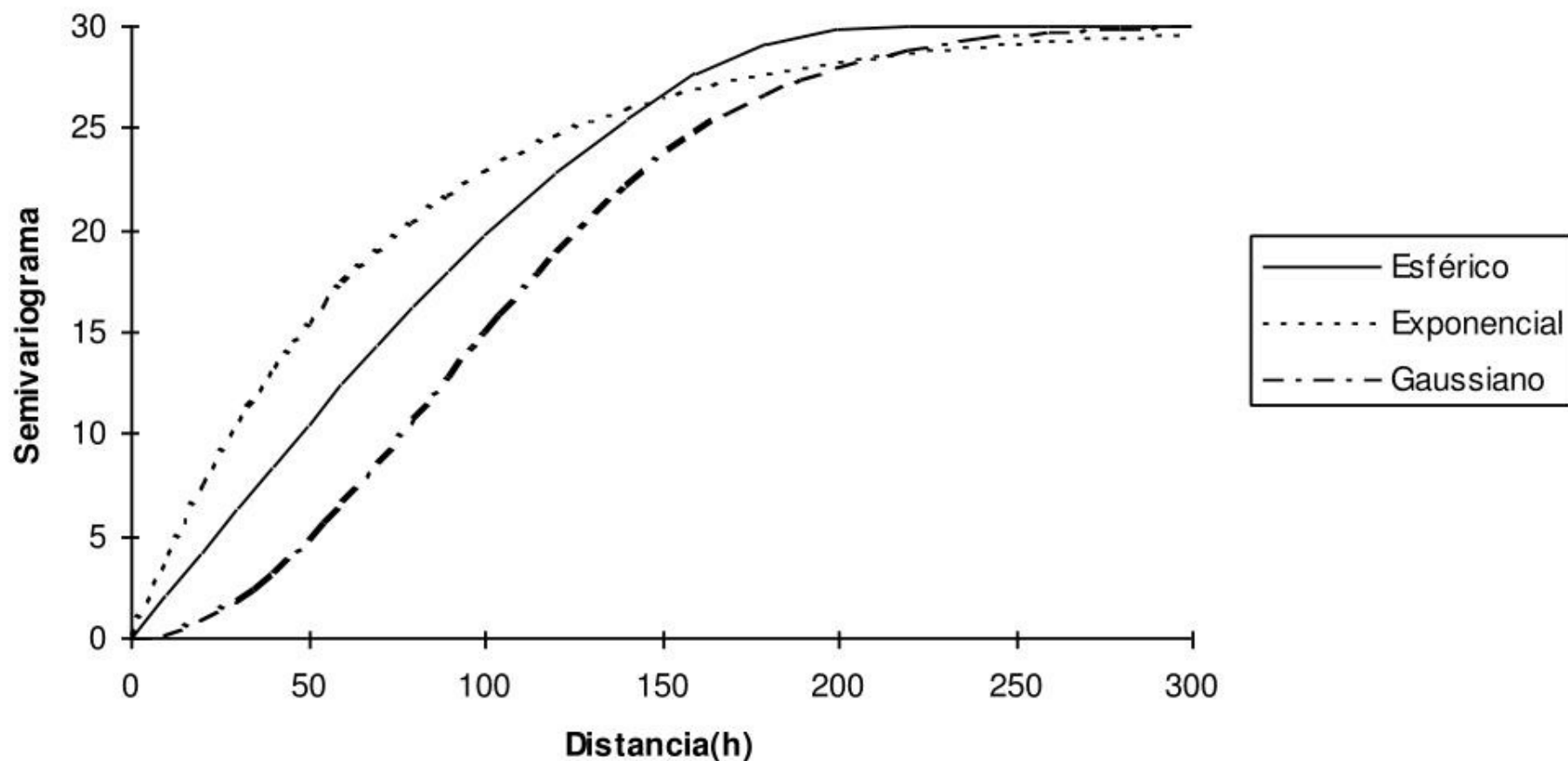
## **Meseta**

Varianza de los datos. Se denota por  $C_1$  o por  $(C_0 + C_1)$  cuando la pepita es diferente de cero. Se sugiere que en un modelo que explique bien la realidad, la pepita no debe representar mas del 50% de la meseta. Si el ruido espacial en las mediciones explica en mayor proporción la variabilidad que la correlación del fenómeno, las predicciones que se obtengan pueden ser muy imprecisas.

## **Rango**

- Es la distancia hasta la cual hay correlación entre los datos. Entre más pequeño sea el rango, más cerca se esta del modelo de independencia espacial. El rango no siempre aparece de manera explícita en la fórmula del semivariograma.

# Comparación de modelos Teóricos de Variograma



## Otras Medidas de Correlación Espacial

$$C(h) = \sigma^2 - \gamma(h)$$

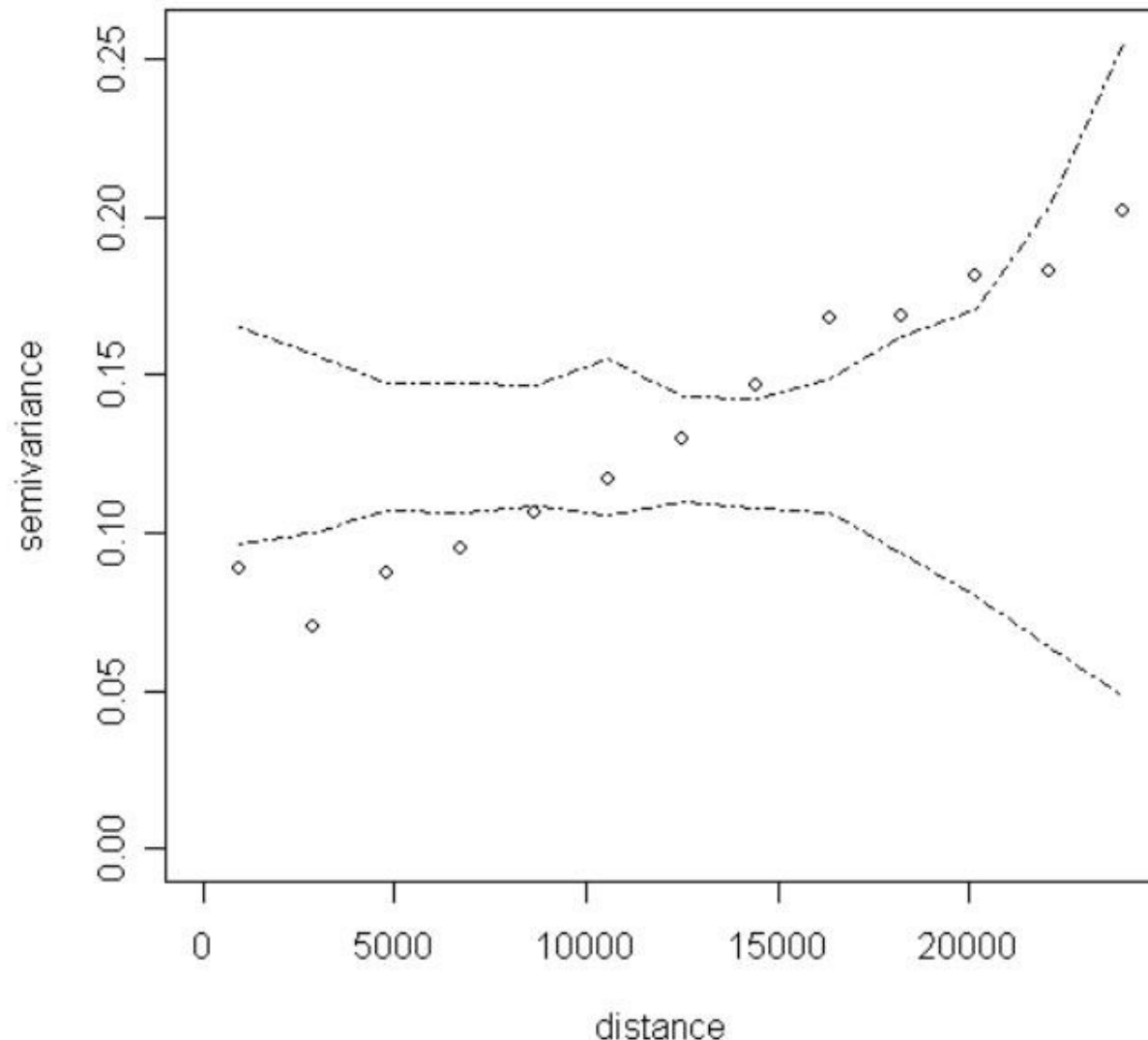
Covariograma

$$\rho(h) = 1 - \frac{\gamma(h)}{\sigma^2}$$

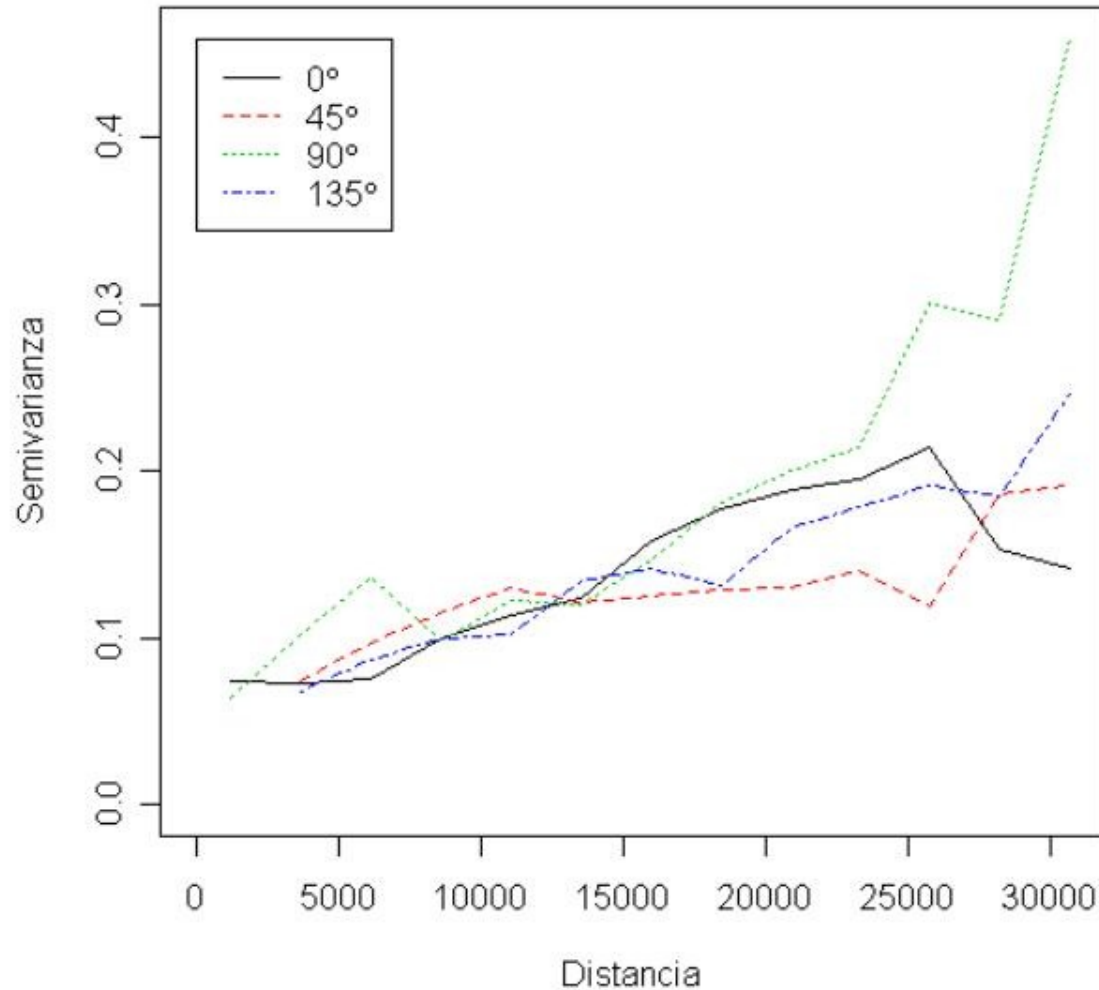
Correlograma

Es más fácil trabajar con el variograma

# Test de Montecarlo de Correlación Espacial



# Isotropía: Igual Correlación en todas las direcciones



## Anisotropías :

Generalmente cuando el variograma experimental es calculado en distintas direcciones presenta distintos comportamientos con la variación de la distancia.

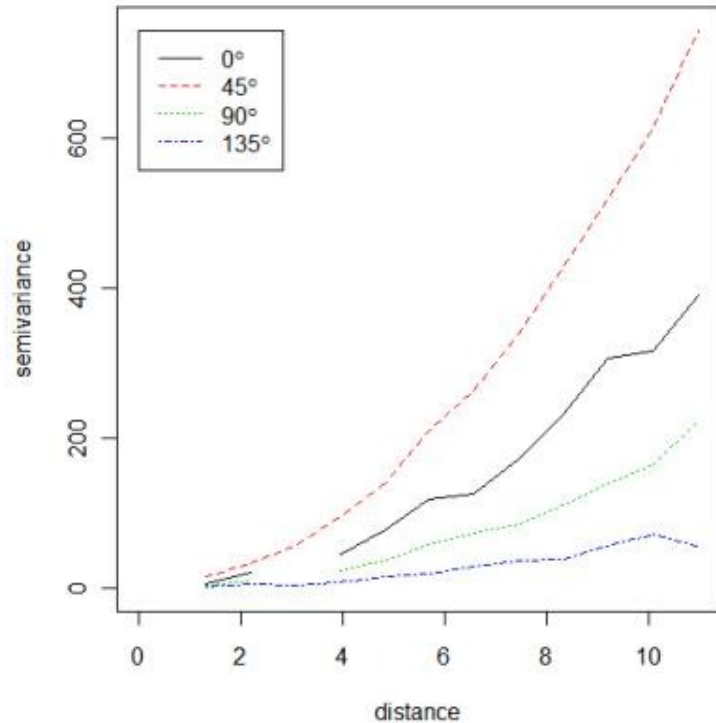
Anisotropía Geométrica

Anisotropía Zonal

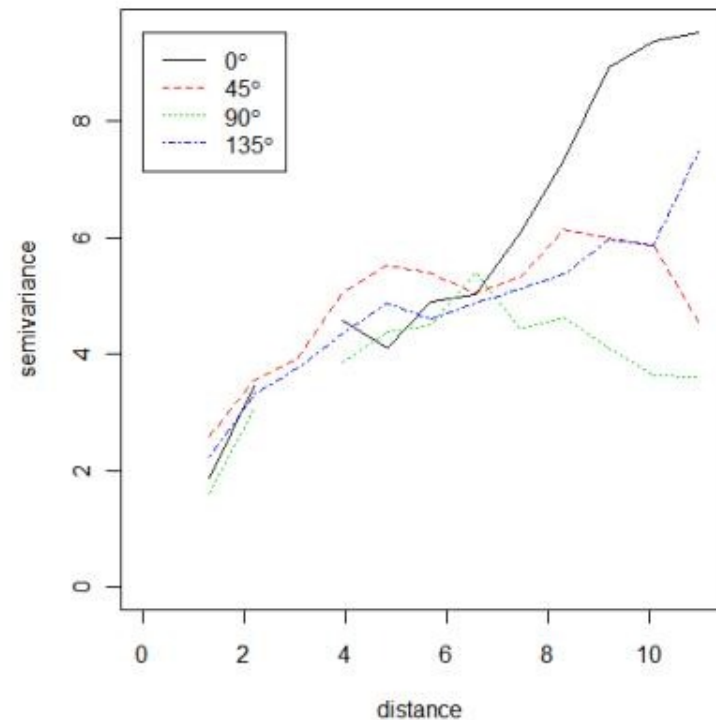
Anisotropía Híbrida



Anisotropía: hay más correlación en una dirección que en otra.



Anisotropía y tendencia.  
Mayor Correlación en dir 90 grados  
(línea verde). Difícil detección!!



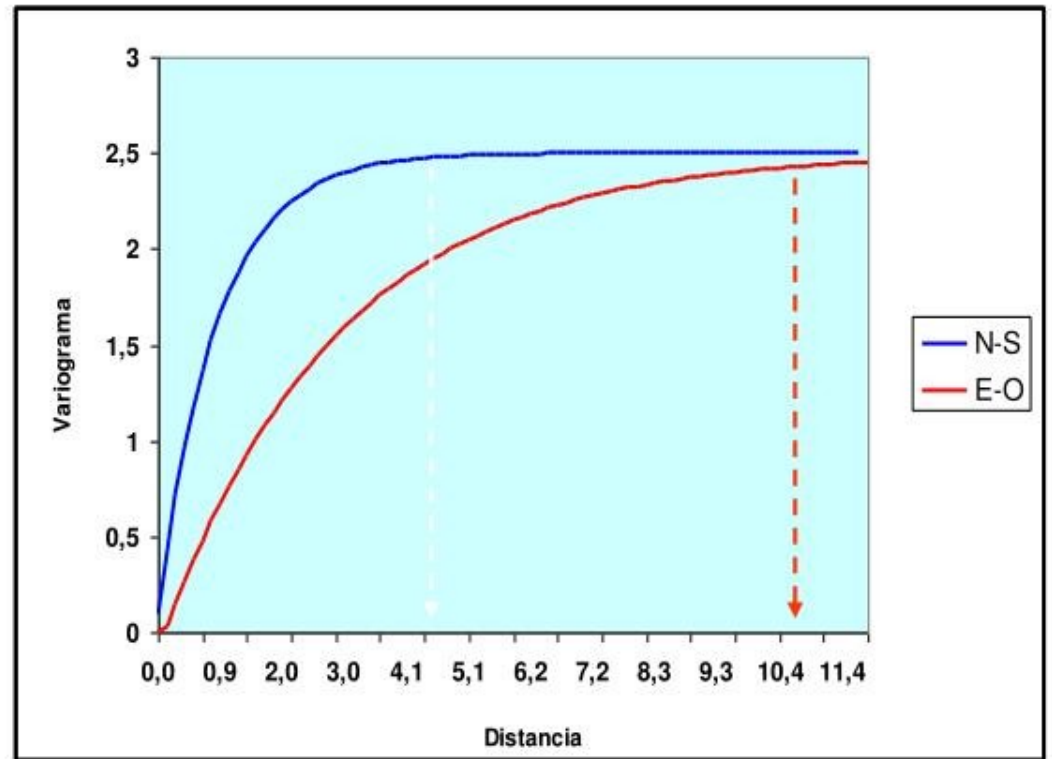
Anisotropía. Mayor correlación en  
dirección 0 grados. Es más fácil  
detectarla bajo estacionariedad.

## Anisotropía Geométrica :

Es aquella en la que el variograma en distintas direcciones presenta el mismo *sill* pero rangos distintos

Mayor continuidad espacial en la dirección de mayor rango

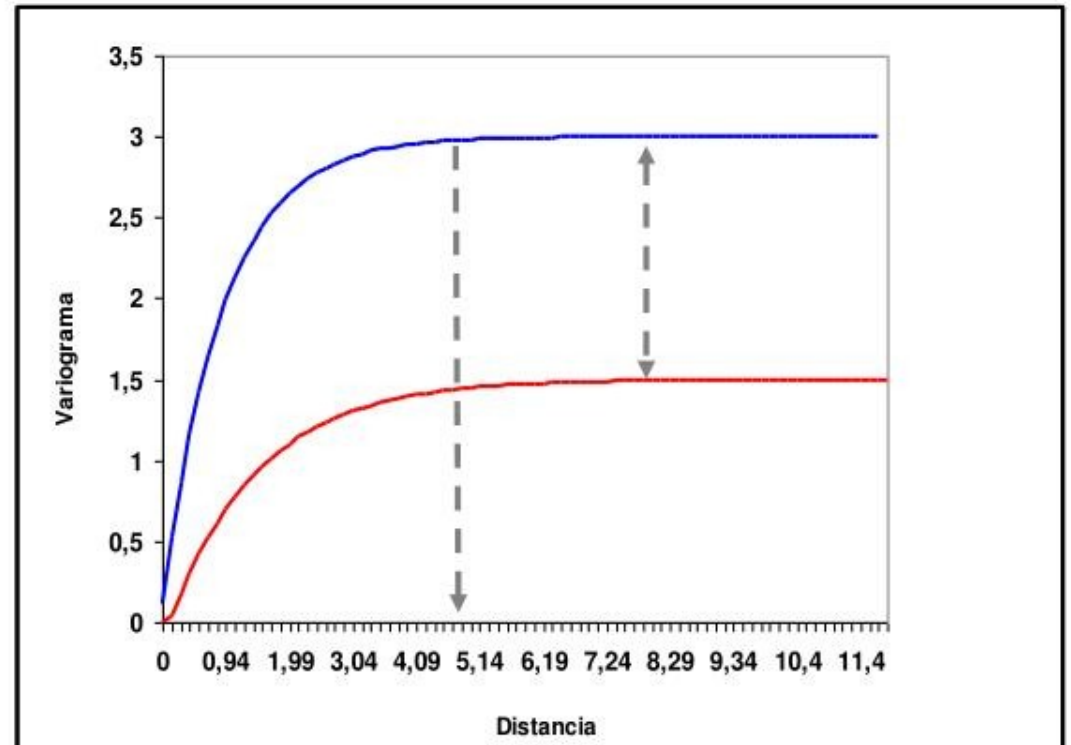
Menor continuidad espacial en la dirección de menor rango



## Anisotropía Zonal :

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

Presencia de diferentes estructuras

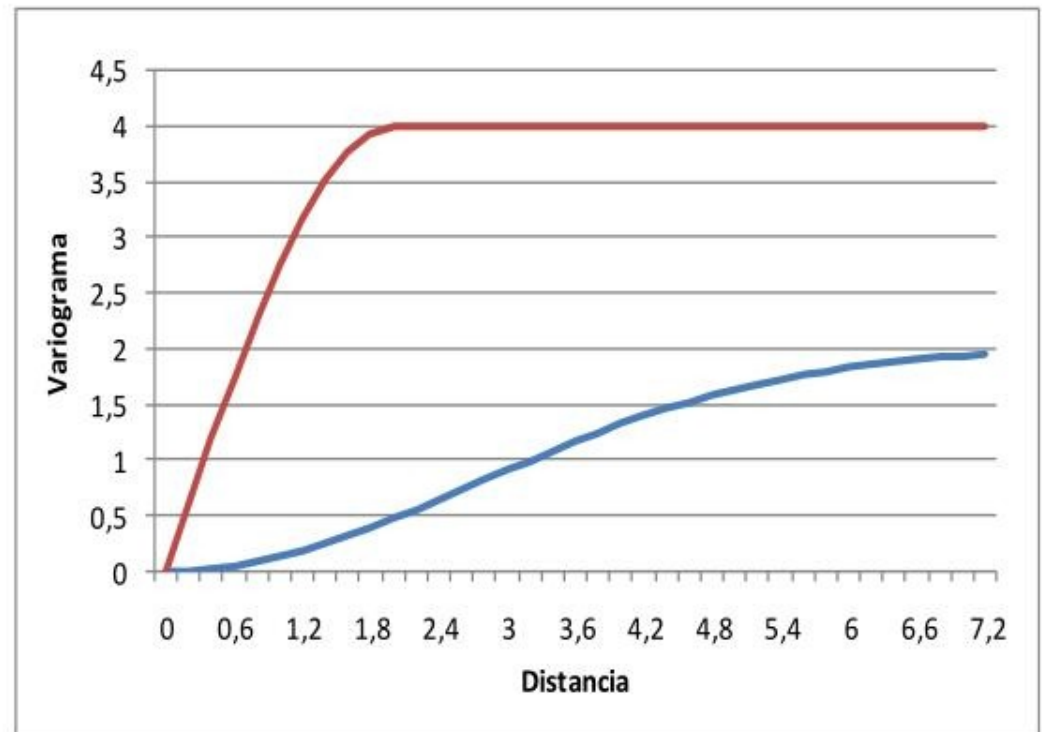


## Anisotropía Híbrida :

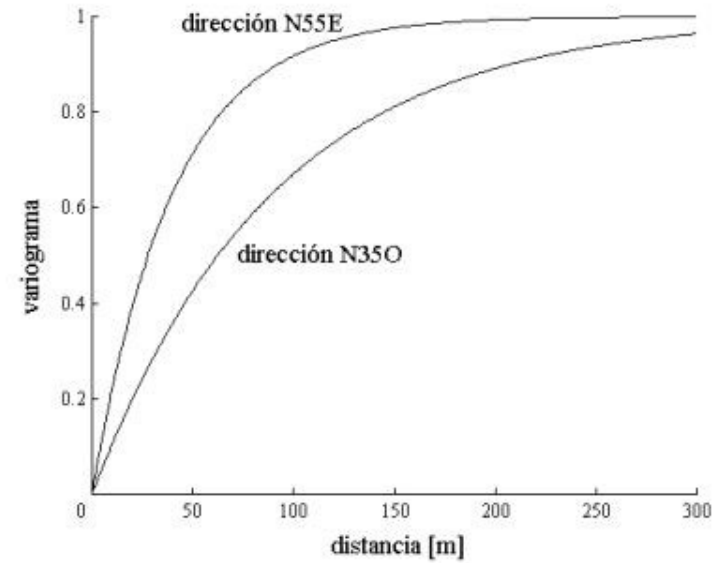
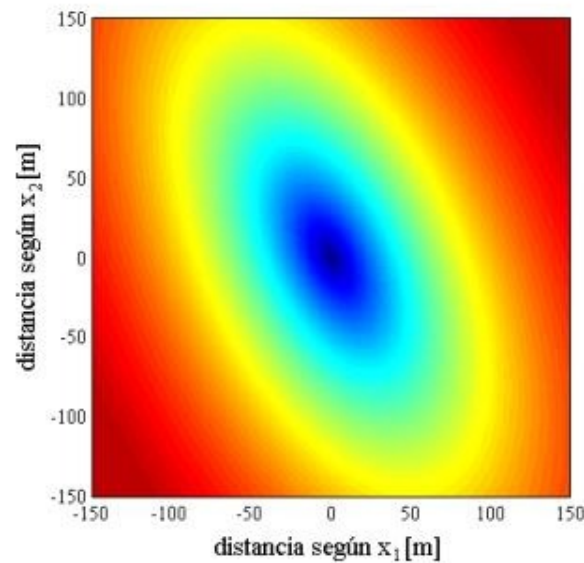
Es aquella en la que el variograma en distintas direcciones presenta rangos diferentes y distintos *sill*.

Presencia de diferentes estructuras

Característico de variogramas horizontales y verticales



Mapa de Variograma: isovalores del variograma experimental en función de la separación (distancia y orientación). Caso de anisotropía geométrica (en cualquier dirección el valor máximo es 1 = sill)



## Modelos de variograma anisotrópicos

-Anisotropía geométrica:

Puede ser corregida por una transformación lineal de coordenadas que permita reducir una elipse a un círculo.

-Anisotropía zonal:

puede ser corregido separando el semivariograma en sus componentes isotrópicos horizontal y anisotrópico vertical.

# Estimación de la Estructura de Correlación

Mínimos Cuadrados  
(Usando el variograma muestral)

- 1.1. Ordinarios
- 1.2. Ponderados
- 1.3. Cuasi-Verosimilitud

2. Geoestadística Basada en Modelo  
(No se usa el variograma muestral)

- 2.1. Máxima-Verosimilitud
- 2.2. MV Restringida
- 2.3. Bayesiana

## Métodos para ajustar el variograma

1) choose the most likely candidate model

2) Methods for estimating the parameters of the model :

- ✓ non-linear least squares estimation – allows for the estimation of parameters that enter the equation non-linearly but ignores any dependences among the empirical variogram values
- ✓ non-linear weighted least-squares – generalized least squares in which the variance-covariance of the variogram data points is accounted for in the estimation procedure
- ✓ maximum likelihood assuming the data are Normally distributed but the estimators are likely to be highly biased, especially in small samples (the usual remedy is jackknifing)
- ✓ restricted maximum likelihood – maximize a slightly altered likelihood function which reduces the bias of the MLEs



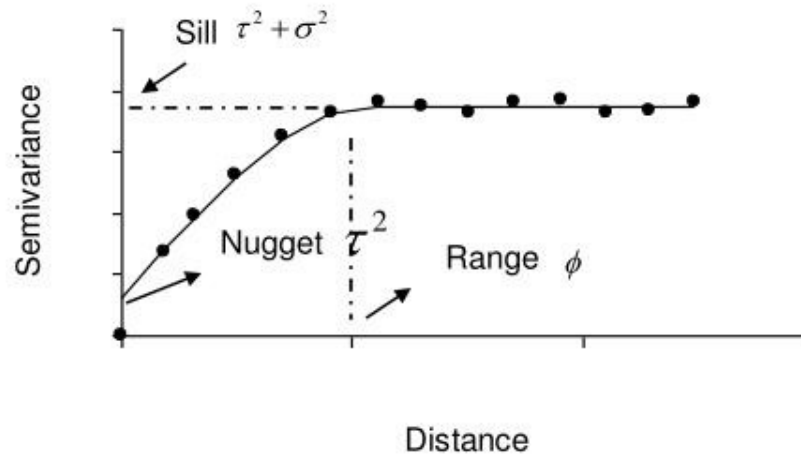
# Estimación por Mínimos Cuadrados

$$OLS(\theta) = \min \sum (\hat{\gamma}(u) - \gamma(u; \theta))^2$$

$$WLS(\theta) = \min \sum n(u) (\hat{\gamma}(u) - \gamma(u; \theta))^2$$

$$QL(\theta) = \min \sum \frac{1}{\gamma(u; \theta)} (\hat{\gamma}(u) - \gamma(u; \theta))^2$$

# Estimación por Mínimos Cuadrados



$$\gamma(u; \theta) = \begin{cases} \tau^2 + \sigma^2 \left( \frac{3}{2} \left( \frac{u}{\phi} \right) - \frac{1}{2} \left( \frac{u}{\phi} \right)^3 \right) & u \leq \phi \\ \tau^2 + \sigma^2 & u > \phi \end{cases} \rightarrow \text{Esferical}$$

$$\gamma(u; \theta) = \tau^2 + \sigma^2 \left( 1 - \exp \left( \frac{-3u}{\phi} \right) \right) \rightarrow \begin{matrix} \text{Exponential} \\ \text{Matérn} \\ k = .5 \end{matrix}$$

$$\gamma(u; \theta) = \tau^2 + \sigma^2 \left( 1 - \exp \left( \frac{-u^2}{\phi^2} \right) \right) \rightarrow \begin{matrix} \text{Gaussian} \\ \text{Matérn} \\ k \rightarrow \infty \end{matrix}$$

## Distribución Normal Multivariada

$$\vec{Y}(s) = \begin{pmatrix} Y(s_1) \\ \vdots \\ Y(s_n) \end{pmatrix} \sim N \left( \vec{\mu}(s) = \begin{pmatrix} \mu(s_1) \\ \vdots \\ \mu(s_n) \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma^2 & \cdots & C(s_1, s_n) \\ \vdots & \ddots & \vdots \\ C(s_n, s_1) & \cdots & \sigma^2 \end{pmatrix} \right)$$

$$f(\vec{Y}(s)) = \frac{1}{(2\pi)^2 |\Sigma|^{1/2}} \exp \left( (\vec{Y}(s) - \vec{\mu}(s))^T \Sigma^{-1} (\vec{Y}(s) - \vec{\mu}(s)) \right)$$

# Geoestadística Basada en Modelo

Modelo

$$\underset{\text{Observaciones}}{\underset{\swarrow}{Y}}(x) = \underset{\text{Señal}}{\underset{\downarrow}{S}}(x) + \underset{\text{Ruido}}{\underset{\downarrow}{Z}}(x)$$

$$\underset{\sim}{S}(x) = \begin{pmatrix} S(x_1) \\ \vdots \\ S(x_n) \end{pmatrix} \sim MVN\left(\underset{\sim}{\mu}, \sigma^2 R\right) \quad \underset{\sim}{Z}(x) = \begin{pmatrix} Z(x_1) \\ \vdots \\ Z(x_n) \end{pmatrix} \sim MVN\left(\underset{\sim}{0}, \tau^2 I\right)$$

$$R = \begin{pmatrix} 1 & \rho_{12} & \cdots & \rho_{1n} \\ & 1 & \cdots & \rho_{2n} \\ & & \ddots & \vdots \\ & & & 1 \end{pmatrix}, \rho(u) = \rho_{ij} = \text{Corr}(S(x_i), S(x_j)), u = \|x_i - x_j\|$$

## Estimación Máximo Verosímil

$$\theta = (\tau^2, \sigma^2, \phi) \quad \text{Parámetros de } G(\theta) = (\tau^2 I + \sigma^2 R), \quad \mu = X\beta$$

$$l(\beta, \theta) = -\frac{1}{2} \left\{ \log |G(\theta)| + (y - X\beta)^T (G(\theta))^{-1} (y - X\beta) \right\} \quad \text{Log-Verosimilitud}$$

$$\hat{\beta}(\theta) = (X'G(\theta)^{-1}X)^{-1} (X'G(\theta)^{-1}Y)$$

$$l(\hat{\beta}, \theta) = -\frac{1}{2} \left\{ \log |G(\theta)| + (y - X\hat{\beta})^T (G(\theta))^{-1} (y - X\hat{\beta}) \right\}$$

$l(\hat{\beta}, \theta)$  es maximizada numéricamente

Estimación MV

$$L^*(\theta) = -\frac{1}{2} \left\{ \log |G(\theta)| + \log |X'(G(\theta))^{-1}X| (y - X\hat{\beta})' (G(\theta))^{-1} (y - X\hat{\beta}) \right\}$$

$L^*(\theta)$  Es maximizada numéricamente

MV Restringida

# Las estimaciones difieren!!

