

# Estadística Espacial en Ecología del Paisaje

## Introducción

H. Jaime Hernández P.

Facultad de Ciencias Forestales - U. de Chile

# Tipos de datos en análisis espacial

---

- Patrones espaciales puntuales
- Muestras geoestadísticas
- *Lattices*
  - Mapas de polígonos
  - Grillas raster
  - Gráficos (nodos y arcos)

Variable Regionalizada: Variable numérica que se distribuye en el espacio

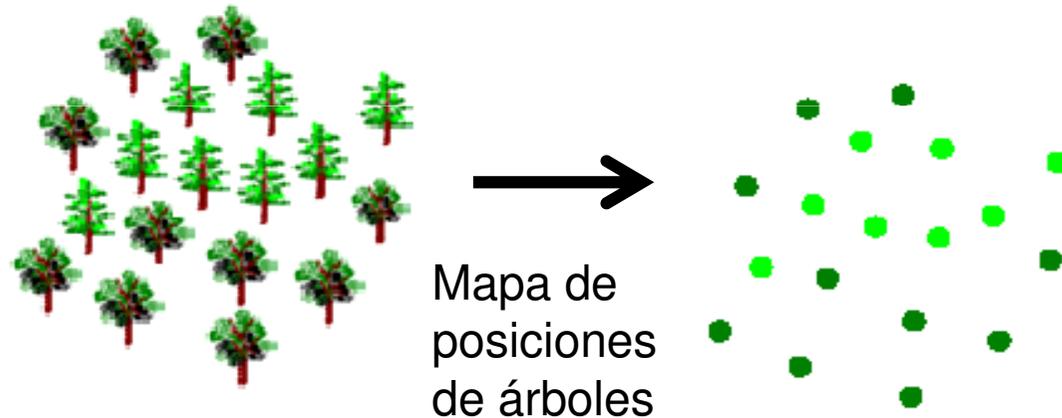
# Parque vs. Procesos Puntuales

- Parches (vectoriales o rásters)
- Patrones puntuales
- Covariable (usualmente grillas)

# Patrones espaciales puntuales

---

Colección de posiciones espaciales de entidades de interés.  
Ejemplo: conjunto de árboles referenciados con sus posiciones  $(x,y)$ . Cada punto se etiqueta con variables asociadas como especies, tamaños, estado, etc.



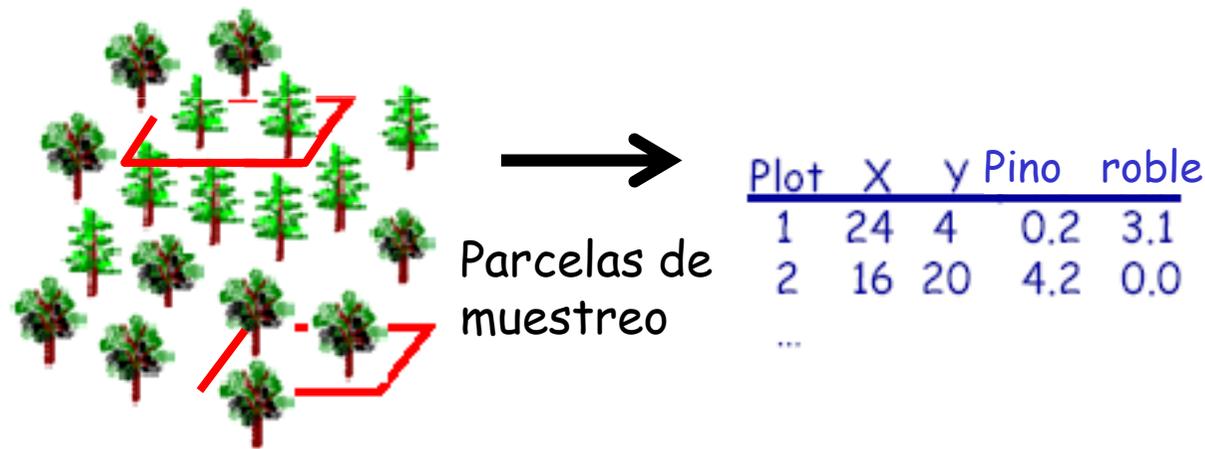
Pregunta primaria: ¿Están agrupados los árboles?

Objetivo del análisis: encontrar la escala espacial a la cuál los puntos tienden a estar más o menos agrupados que al azar.

# Muestras geoestadísticas

---

Mediciones sobre muestras espaciales discretas, separadas a una distancia dada. Ejemplo: parcelas de muestreo florístico. Interesa la variable asociada (ejem: diversidad) más que la posición de cada individuo en la parcela (SOPORTE).



Pregunta primaria: ¿Parcelas más cercanas son más parecidas?

Objetivo del análisis: encontrar la escala espacial a la cuál los valores agregados por parcela tienden a ser similares.

# Lattices

---

Dato tipo *Lattice* (enrejado) representan regiones de algún tamaño y pueden ser intuitivamente definidos usando el concepto de borde. Una región es un área que tiene bordes adyacentes a otra región. Ejemplos: usos de suelo, píxeles de una imagen satelital, o nodos unidos con arcos (uniones funcionales).



Pregunta primaria: ¿A qué escala regiones adyacentes tienden a tener valores similares?

# Autocorrelación espacial

---

La autocorrelación espacial se refiere al comportamiento espacial de una variable (puntual, geoestadística o tipo *lattice*) en la cual datos "más cercanos" tienden a ser "más" o "menos" similares que debido sólo al azar.

Ejemplo 1: Calidad del hábitat en zonas silvestres.

Ejemplo 2: Tamaño de la copa en un bosque de *Araucarias* de estructura irregular.

Ejemplo 3: Promedio acumulado de notas de alumnos de pregrado, cuando están sentados en el curso de Cálculo II.

Primera Ley de la Geografía (Waldo Tobler): Todo está relacionado con todo, sin embargo objetos más cercanos están más relacionados. Posibilidad de Predicción espacial.

# Métodos ordenados por objetivo de investigación

---

## OBJETIVO

## ESTADÍSTICA ESPACIAL

### Exploración

Vecino más cercano, Índice de Morán, Índice de Geary, semivarianza  $\gamma$  (variograma), Correlogramas, Ripley's K (uni and bivariada), *joint count*, índices agregados (ejem.: varianza/media), etc.

### Inferencia

Ripley's K (uni and bivariada), Índice de Morán, Índice de Geary, semivarianza  $\gamma$  (variograma), *joint count*, etc.

### Interpolación (*Mapping*)

Análisis de tendencias superficiales (trend), kriging, splines, polígonos de Voroni, etc.

# Vecino más cercano (NNI)

---

Se analiza como el radio entre 2 estadísticas:

i. Distancia al vecino más cercano  $d(NN)$

Donde  $Min(d_{ij})$  es la distancia de cada punto a su vecino más cercano, y  $N$  es el número de puntos en la distribución.

$$d(NN) = \sum_{i=1}^N \left[ \frac{Min(d_{ij})}{N} \right]$$

ii.  $d(NN)$  esperada basada en una distribución al azar:

$$d(ran) = 0.5 \sqrt{\frac{A}{N}}$$

Índice Vecino más Cercano:

$$NNI = \frac{d(NN)}{d(ran)}$$

Interpretación: si  $NNI < 1 \rightarrow$  distribución agrupada

# Índices de Moran (I) y Geary (c)

---

Índice de Moran (I) indica el grado de autocorrelación en función de la distancia (d). Comparable a "r" de Pearson.

$$-1 \leftarrow I \rightarrow +1$$

Autocorr: -- ++

Índice de Geary (G) indica el grado de autocorrelación entre posiciones cercanas (Pearson).

$$0 \leftarrow G \rightarrow 1$$

Autocorr: ++ --

$$I = \frac{\sum_i \sum_j w_{ij} \cdot (z_i - \bar{z}) \cdot (z_j - \bar{z})}{\sum_i \sum_j w_{ij} \cdot s_z^2}$$

$$\text{donde } s_z^2 = \sum_j (z_j - \bar{z})^2 / n$$

$$G = \frac{\sum_i \sum_j w_{ij} \cdot (z_i - z_j)^2}{2 \cdot \sum_i \sum_j w_{ij} \cdot \sigma_z}$$

$$\text{donde } \sigma_z = \sum_i (z_i - \bar{z})^2 / (N - 1)$$

$w_{ij}$  = elementos de la matriz binaria. 1 si  $x_i$  e  $y_j$  están en la clase de distancia d. 0 en otro caso.  $W(d)$  es la suma de los  $w_{ij}$ .

# *Joint Count*

---

Útil para imágenes binarias (lattice de 0 y 1). La hipótesis nula es que una región vecina tiene mayor probabilidad de tener la misma categoría que la región focal.

$J_{BB}$  = número de regiones conjuntas que tienen la misma categoría.

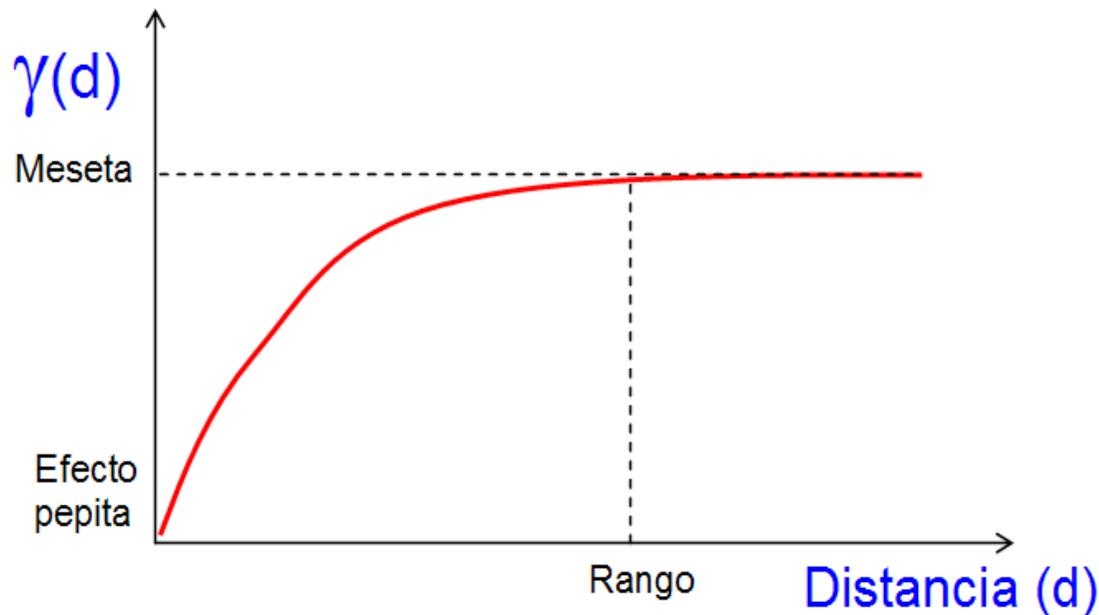
$$J_{BB} = \sum \sum w_{ij} Y_{ij}, \quad \text{where } Y_{ij} = x_i x_j$$

$w_{ij} = 1$  si  $x_i$  e  $y_i$  son adyacentes; es cero es otro caso.

# Semivarianza (variograma)

Modelo de dependencia espacial: **Variograma**.

$$\hat{\gamma}(d) = \frac{\sum \sum w_{ij}(d)(x_i - x_j)^2}{2W(d)}$$



$\gamma(d)$ : Indica la diferencia al cuadrado promedio de puntos separados a una distancia igual a  $d$ . *para la variable de interés*

$w_{ij}$ : es 1 si el par  $x_i, x_j$  están en la categoría de distancia  $d$ ; es 0 en otro caso.

Variograma Experimental  
vs.  
Variograma Teórico

# Variograma experimental

## Definición

Denotemos como  $\{\mathbf{x}_\alpha, \alpha = 1 \dots n\}$  los sitios con datos y como  $\{z(\mathbf{x}), \mathbf{x} \in D\}$  la variable regionalizada. El **variograma experimental** mide la desviación cuadrática promedio entre dos datos en función de su separación:

$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2|N(\mathbf{h})|} \sum_{(\alpha, \beta) \in N(\mathbf{h})} [z(\mathbf{x}_\alpha) - z(\mathbf{x}_\beta)]^2$$

donde  $N(\mathbf{h}) = \{(\alpha, \beta) \text{ tales que } \mathbf{x}_\alpha - \mathbf{x}_\beta = \mathbf{h}\}$   
 $|N(\mathbf{h})|$  es el cardinal de  $N(\mathbf{h})$

# Variograma experimental

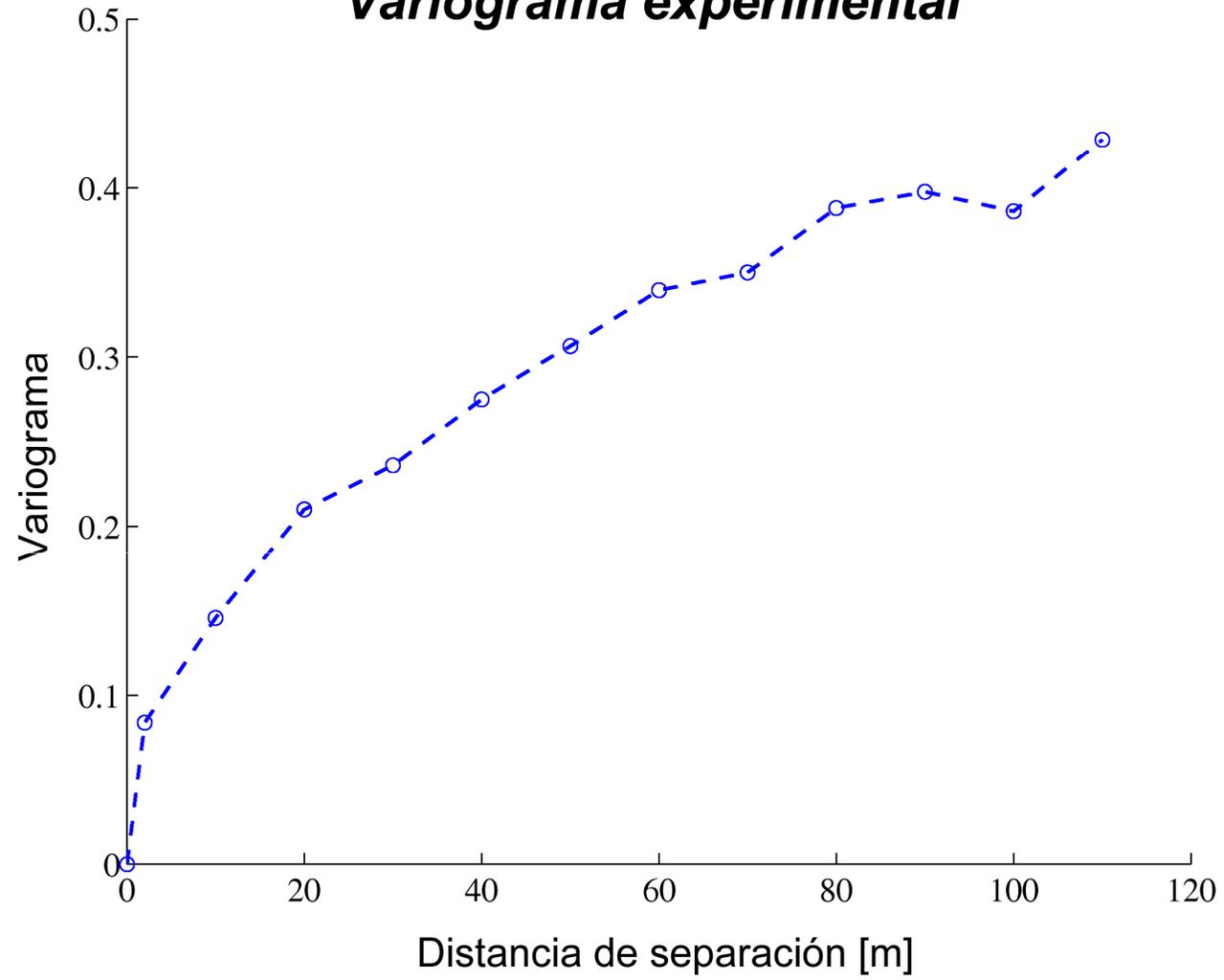
## Ejercicio

Consideremos los siguientes datos espaciados cada 100 m

5 3 6 4 2 1 1 2 4 3 2

Aplicar la fórmula anterior para calcular el variograma experimental para distancias de 100m, 200m... hasta 1000m.

## *Variograma experimental*



# Herramientas alternativas

## Covarianza experimental centrada

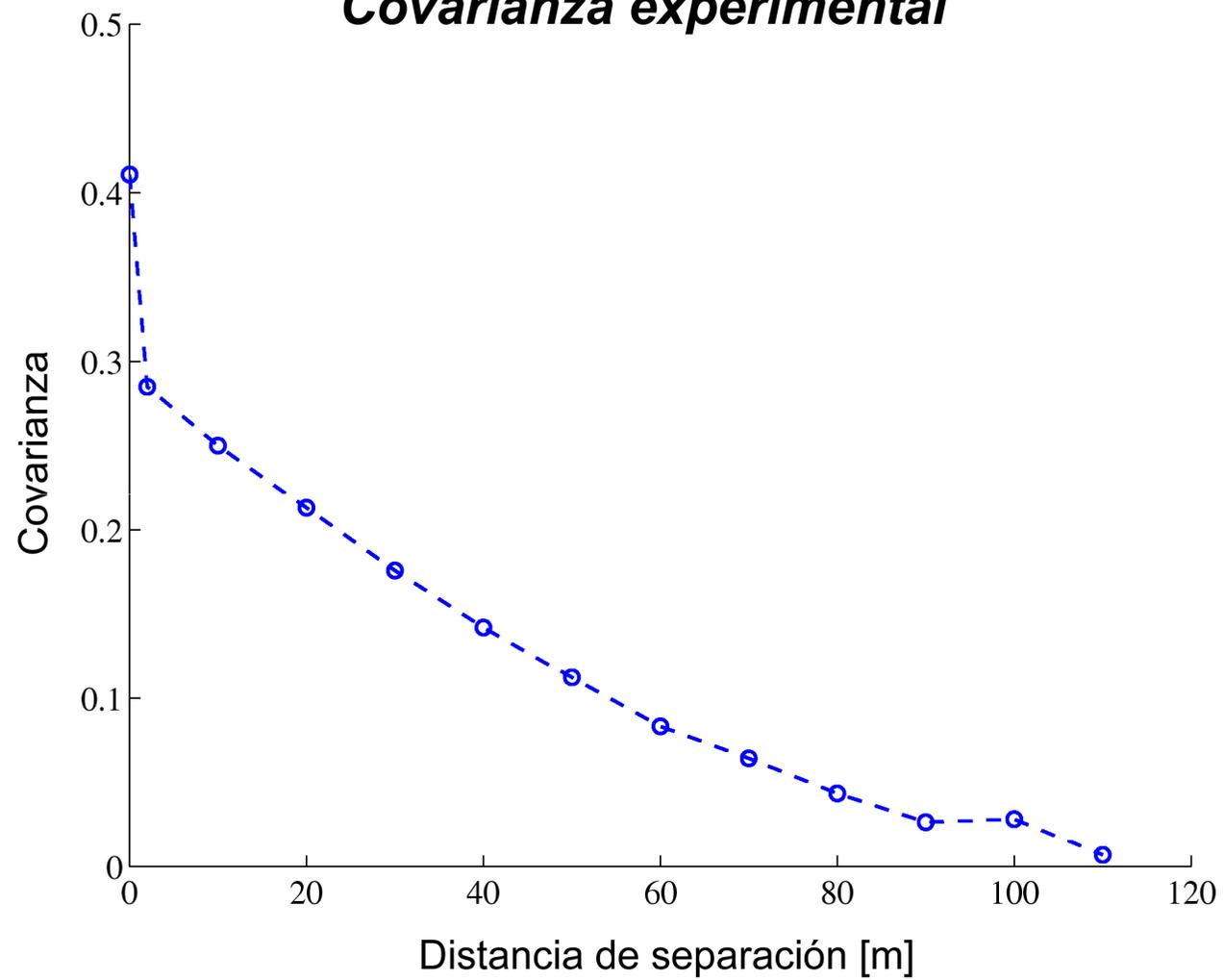
$$\hat{C}(\mathbf{h}) = \frac{1}{|N(\mathbf{h})|} \sum_{(\alpha, \beta) \in N(\mathbf{h})} [z(\mathbf{x}_\alpha) - \bar{z}][z(\mathbf{x}_\beta) - \bar{z}]$$

Se puede invertir en variograma al plantear

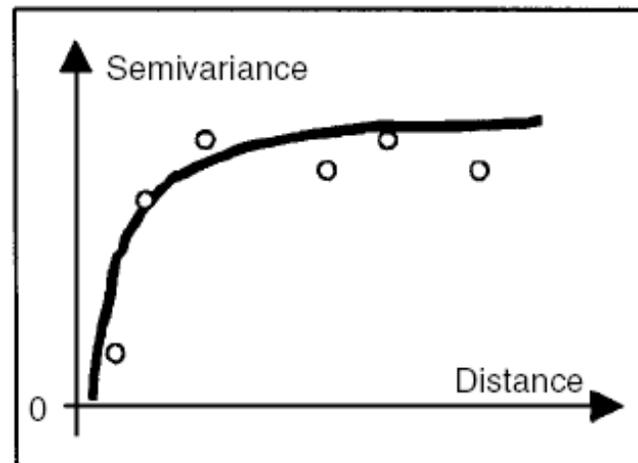
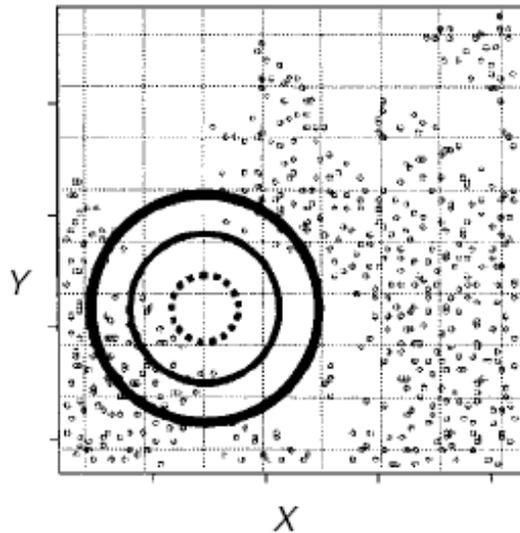
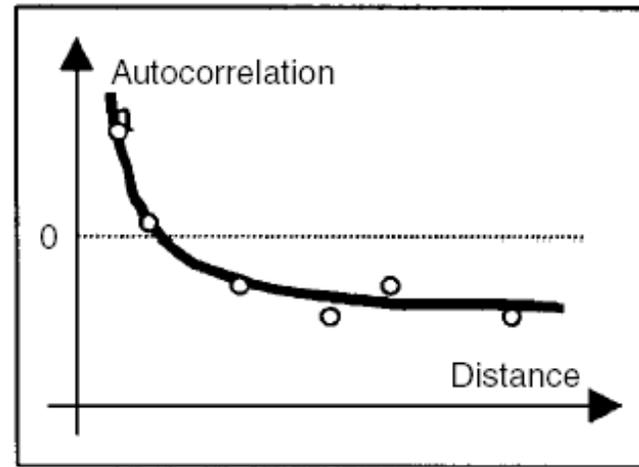
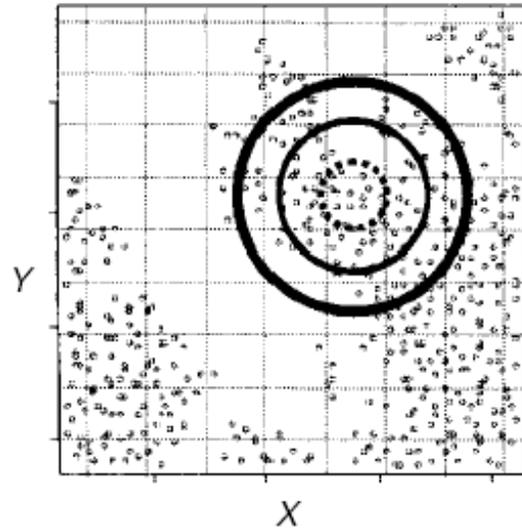
$$\hat{C}(\mathbf{h}) - \hat{C}(\mathbf{0})$$

Este estimador puede ser más robusto que el variograma experimental tradicional, pero está sesgado

## ***Covarianza experimental***

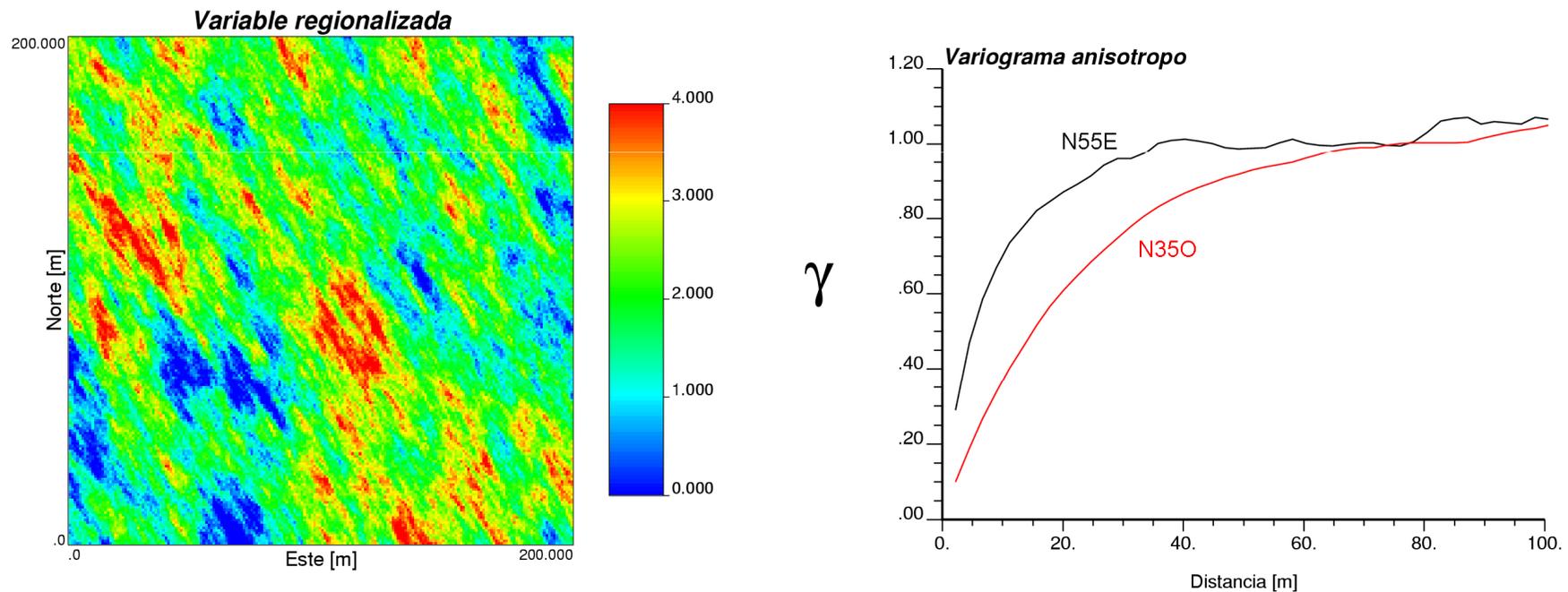


# Semivarianza (variograma)



# Comportamiento direccional

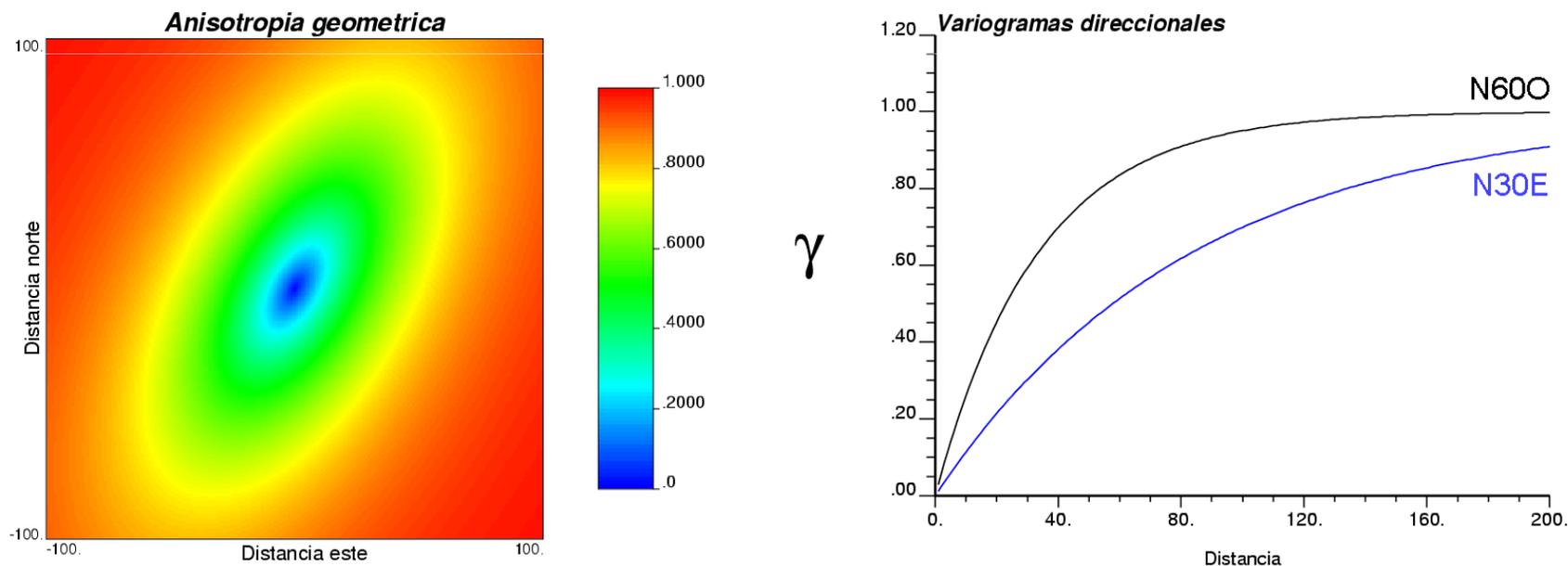
El estudio de los variogramas direccionales permite identificar las **anisotropías** de la variable regionalizada.



# Anisotropía

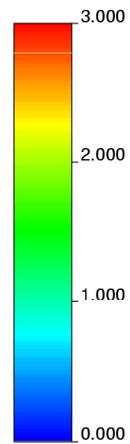
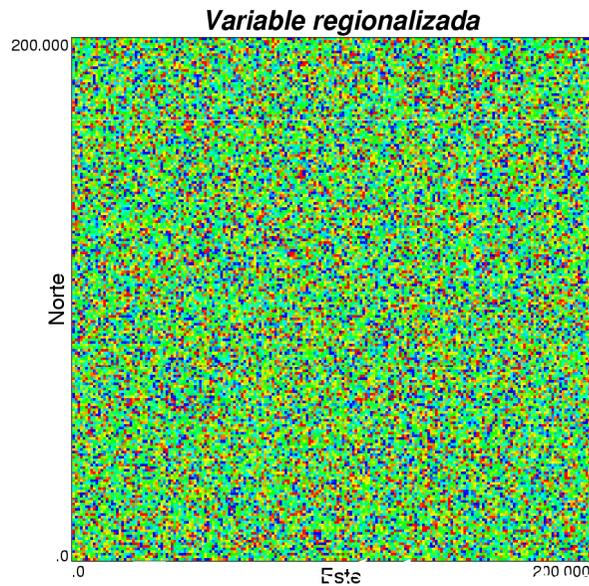
## Anisotropía geométrica

El mapa variográfico dibuja elipses (2D) o elipsoides (3D). Sólo se requiere especificar las direcciones principales (ortogonales) y los alcances correspondientes.

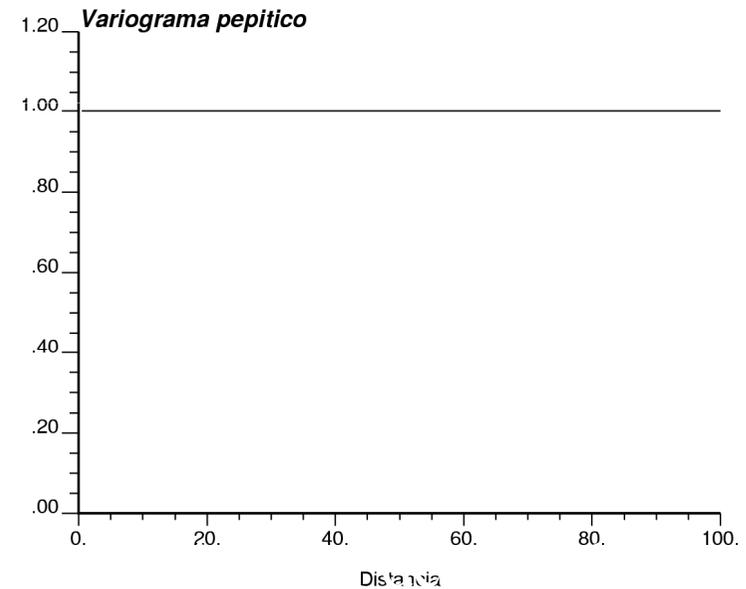


# Modelos elementales

**Efecto pepita**  $\gamma(\mathbf{h}) = \begin{cases} 0 & \text{si } \mathbf{h} = \mathbf{0} \\ C & \text{en caso contrario} \end{cases}$



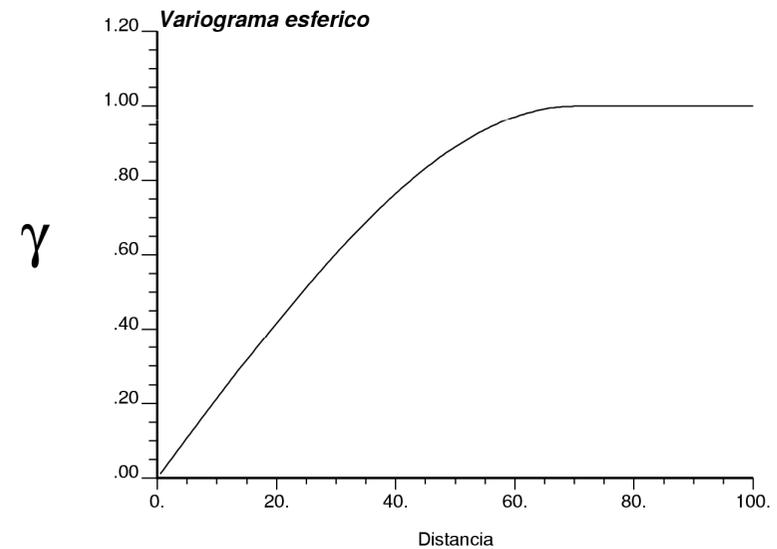
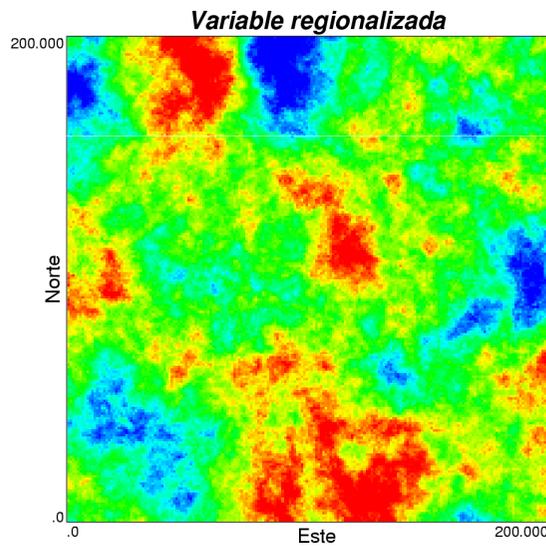
$\gamma$



# Modelos elementales

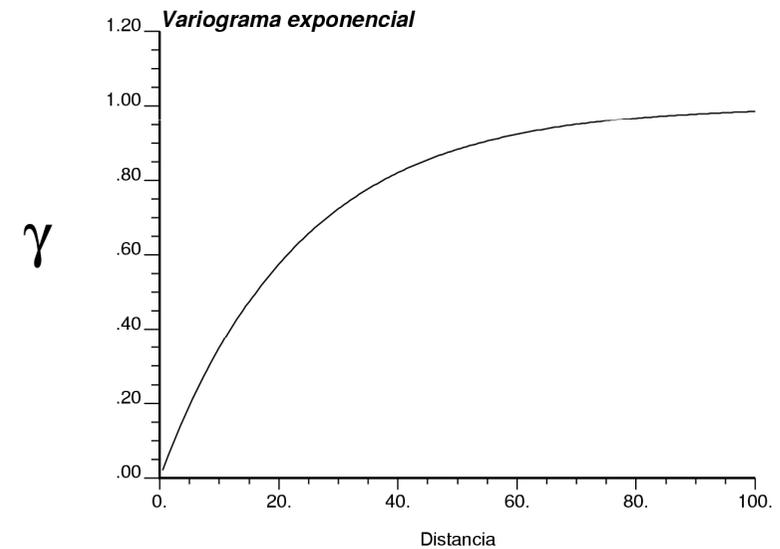
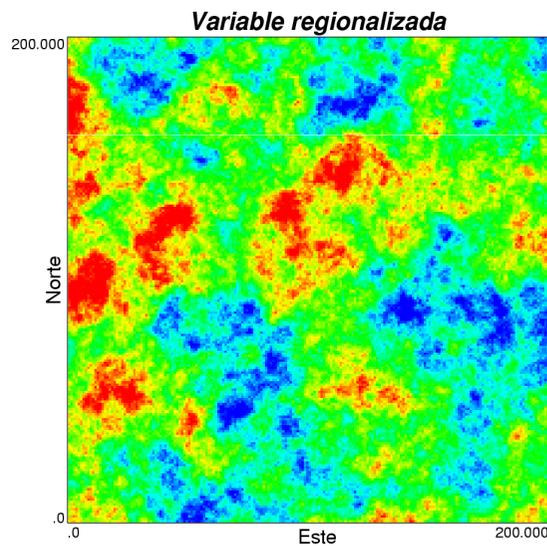
## Modelo esférico

$$\gamma(\mathbf{h}) = \begin{cases} C \left\{ \frac{3|\mathbf{h}|}{2a} - \frac{1}{2} \left( \frac{|\mathbf{h}|}{a} \right)^3 \right\} & \text{si } |\mathbf{h}| \leq a \\ C & \text{en caso contrario} \end{cases}$$



# Modelos elementales

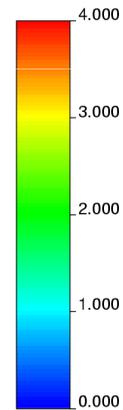
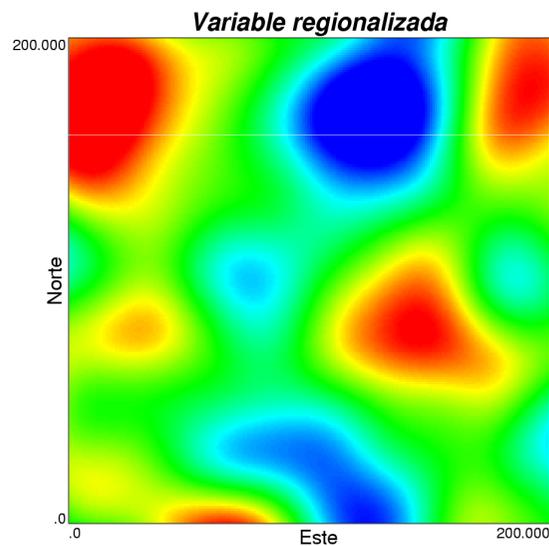
**Modelo exponencial**  $\gamma(\mathbf{h}) = C \left\{ 1 - \exp\left(-\frac{3|\mathbf{h}|}{a}\right) \right\}$



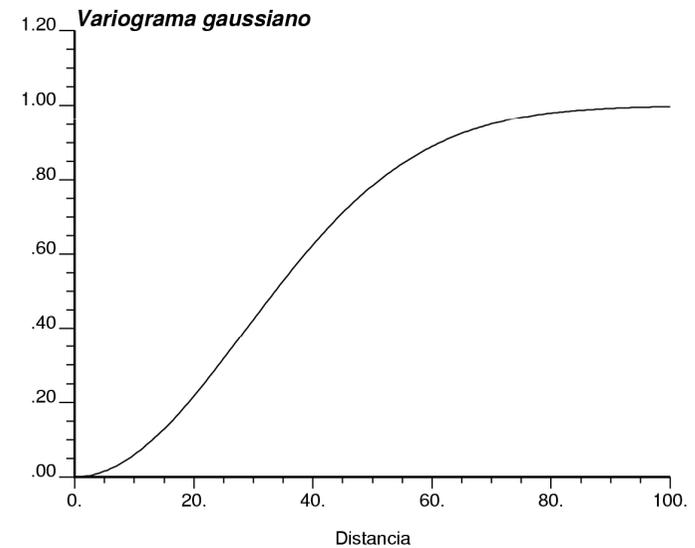
# Modelos elementales

## Modelo Gaussiano

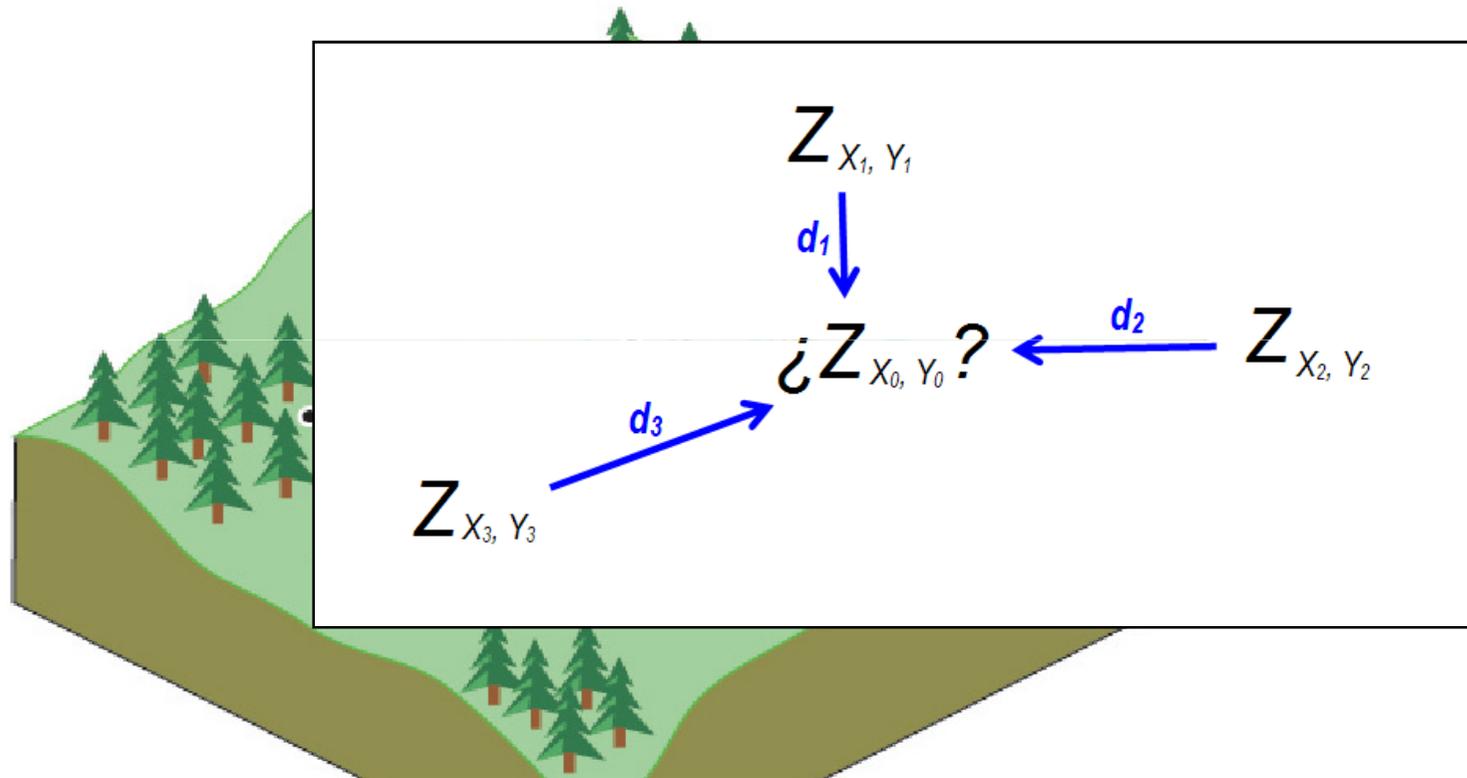
$$\gamma(\mathbf{h}) = C \left\{ 1 - \exp\left(-\frac{3|\mathbf{h}|^2}{a^2}\right) \right\}$$



$\gamma$



# Interpolación espacial



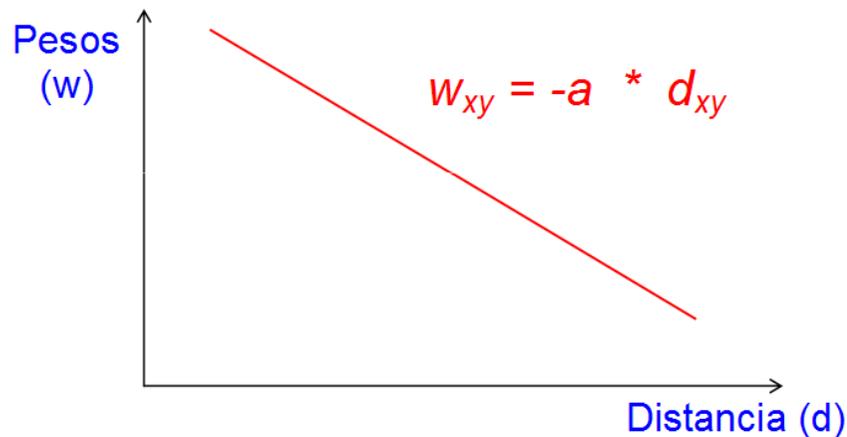
¿Cómo predecir el valor de la variable  $Z$  en la posición  $(x_0, y_0)$  si conozco su valor en las posiciones  $(x_1, y_1)$ ,  $(x_2, y_2)$ ,  $(x_3, y_3)$ ?

# Interpolación espacial

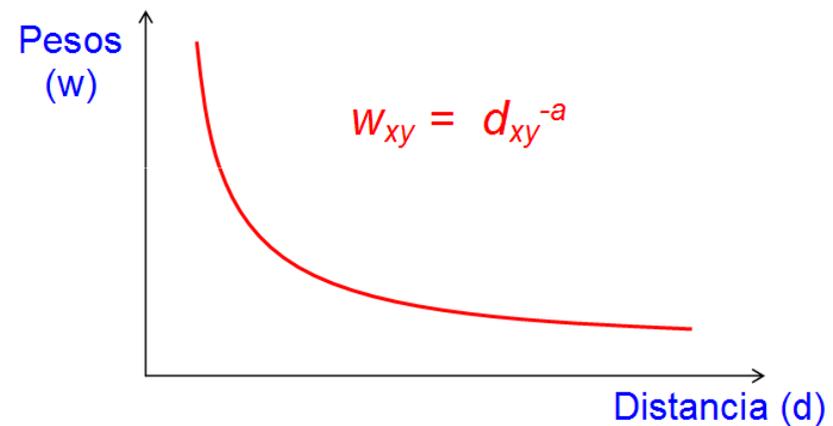
---

Modelo para describir el efecto de la distancia

Lineal inverso a la distancia

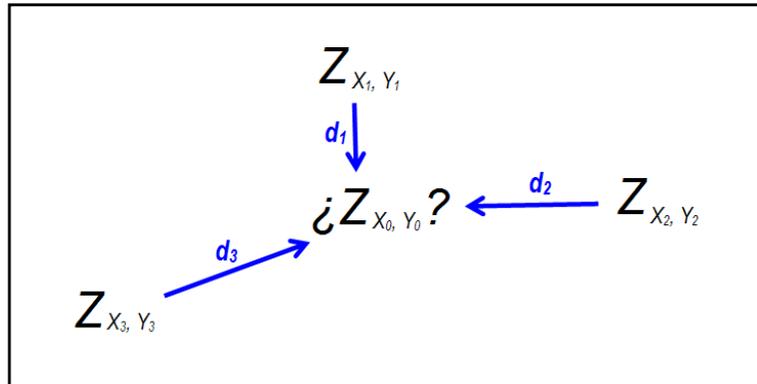


Exponencial negativo respecto a la distancia



$$Z_0 = w_1 * Z_1 + w_2 * Z_2 + w_3 * Z_3$$

# Kriging



$$Z_0 = w_1 * Z_1 + w_2 * Z_2 + w_3 * Z_3$$

$$z^*(\mathbf{x}_0) = a + \sum_{\alpha=1}^n \lambda_{\alpha} z(\mathbf{x}_{\alpha})$$

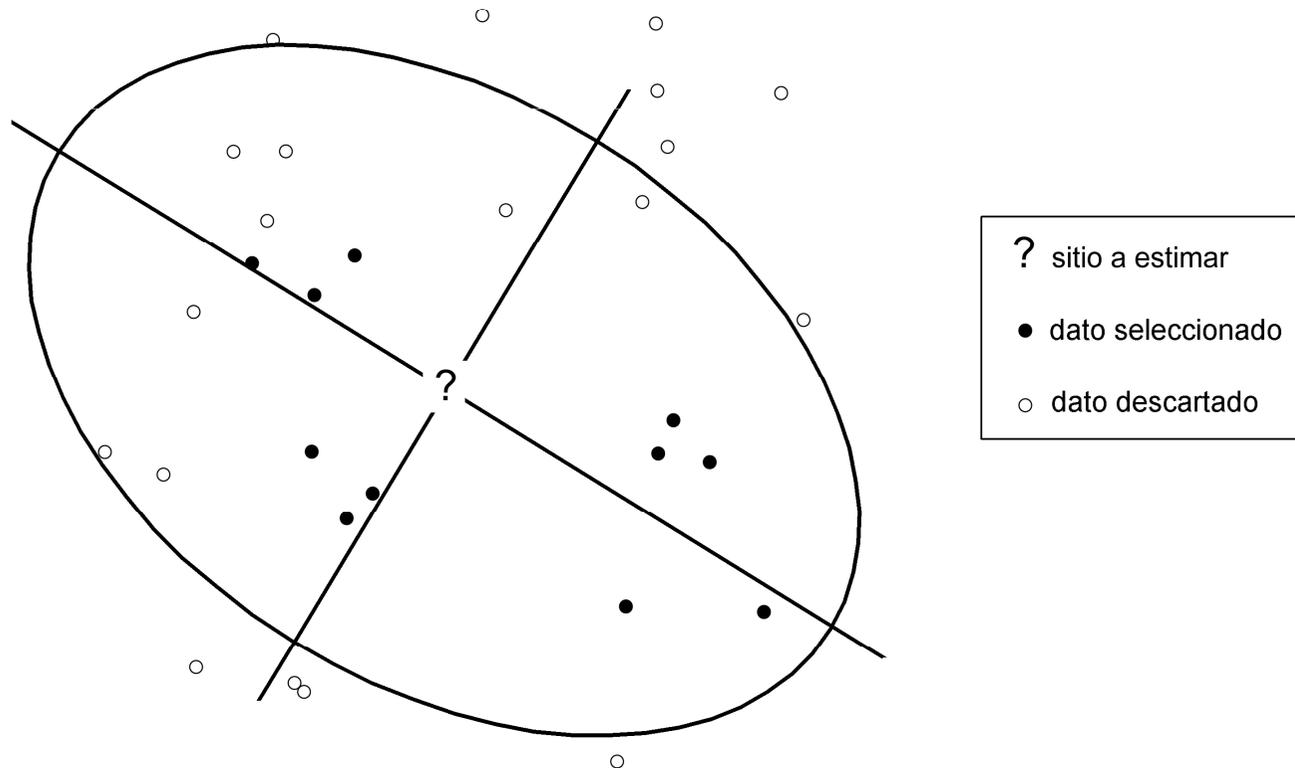
Los pesos son calculados a partir del variograma de modo que se cumplen las siguientes condiciones:

1. Linealidad (  $z_0(x_0) = a + \sum w_i z(x_i)$  )
2. Modelo insesgado (  $E [Z^*(x_0) - Z(x_0)] = 0$  )
3. Optimo (  $\text{var} [Z^*(x_0) - Z(x_0)]$  es mínima )

## ¿Cuáles son los datos a utilizar en la estimación?

Se puede utilizar todos los datos disponibles (*vecindad única*) o sólo una parte de ellos (*vecindad móvil*). La palabra “vecindad” se refiere a la zona del espacio, centrada en el sitio a estimar, donde se busca los datos que servirán en la estimación.

### ***Búsqueda de tres datos por cuadrante***



Diapositiva: Xavier Emery, FCFM, U. de Chile