

# ***GEO ESTADÍSTICA***

*Una presentación*

*Lic. Daniel Horacio Calvo*

# Definición de Geoestadística

- Etimológicamente, el término **Geoestadística** designa el estudio estadístico de fenómenos naturales.
- Matheron fue el primero en utilizar este término extensamente proponiendo la siguiente definición:

**La geoestadística es la aplicación del formalismo de las funciones aleatorias al reconocimiento y estimación de fenómenos naturales**

- Un fenómeno natural puede ser caracterizado por la distribución en el espacio de una o más variables llamadas **variables regionalizadas**.

# Objeto de la Geoestadística

- El objeto de la Geoestadística es: **describir cuantitativamente variables naturales distribuidas en el espacio**

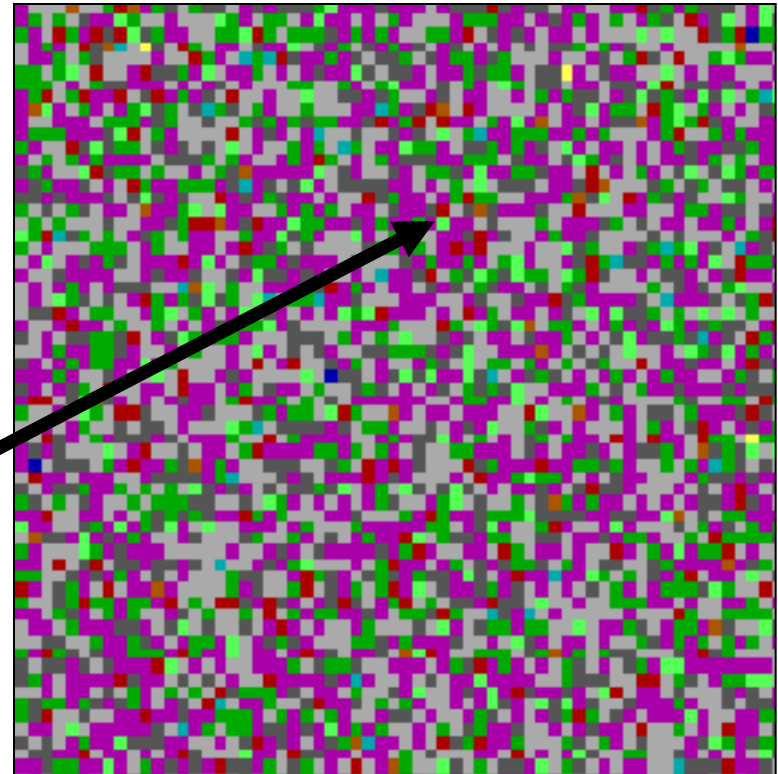
por ejemplo:

- La concentración de un mineral precioso en una mina.
- La profundidad y el espesor de un estrato geológico.
- La porosidad y la permeabilidad de un medio poroso.
- Las propiedades del suelo en una región.
- La lluvia caída sobre una cuenca.
- Las concentraciones de un soluto en una zona contaminada.
- ...

# Geoestadística: concepto



**Los datos experimentales son una realización de un campo aleatorio**



**Realización de una variable aleatoria**

# Pasos de un estudio geoestadístico

- Descripción de la información.
- Análisis de la continuidad espacial.
- Predicción espacial.
- Evaluación de la incertidumbre.

# Descripción de la información

- Los datos hablan más claramente cuando ellos están organizados.
- **Objetivos:**
  - Explorar y describir los datos.
  - Condensar la información.
  - Obtener un vehículo para comunicar la información.

# Tipos de descripciones

- Según el número de variables a describir:
  - Descripción univariada.
  - Descripción bivariada.
- Si los datos están asociados a una localización en el espacio:
  - Descripción espacial.
- Los métodos para realizar la descripción de un conjunto de datos pueden clasificarse según las técnicas utilizadas en:
  - Métodos gráficos.
  - Métodos numéricos.

# Comparación

Descripción  
espacial

Datos Localizados

Modelación de la  
continuidad espacial

Predicciones  
localizadas

Geoestadística

Descripción

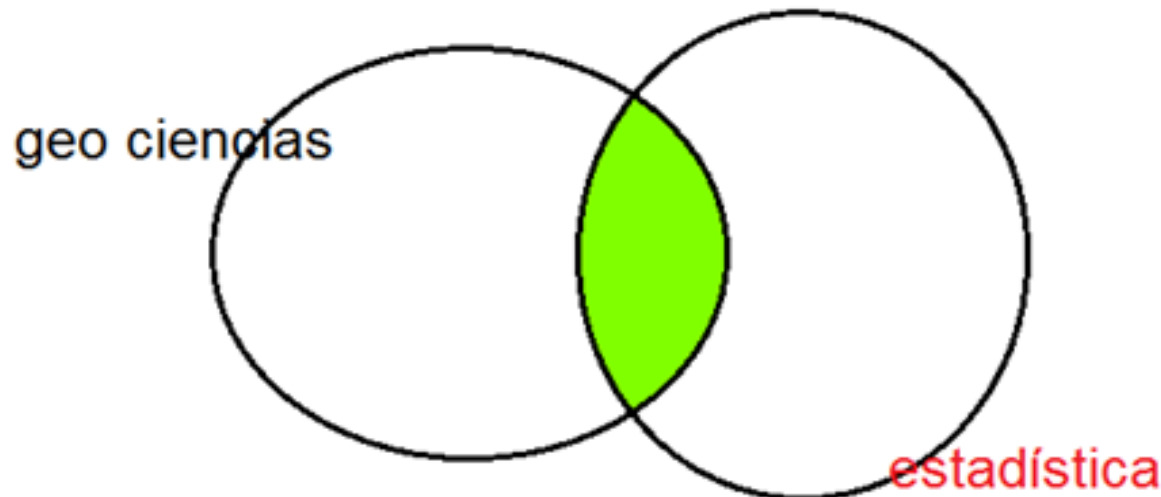
univariada y bivariada

Datos no Localizados

Modelación de los Datos

Estimación de Parámetros  
y Predicción

Estadística



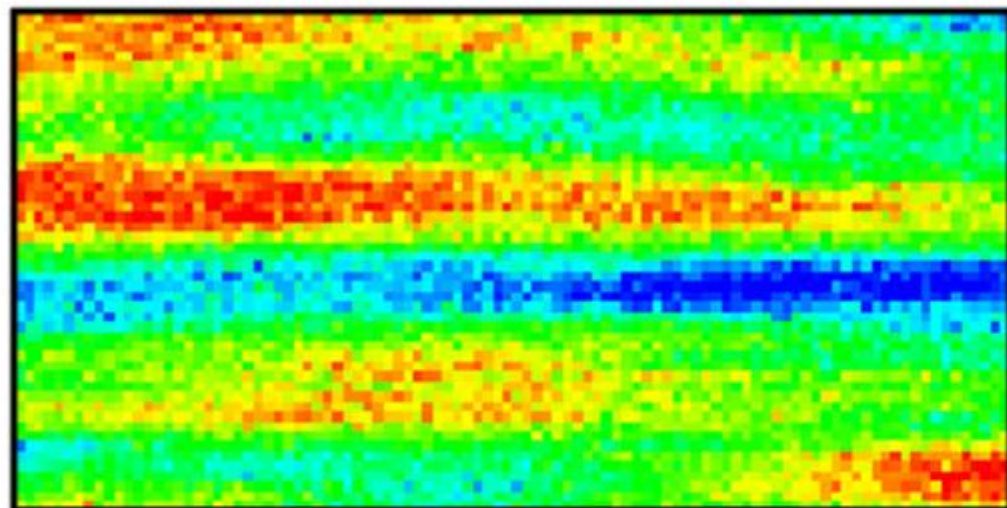


# Análisis de la continuidad espacial

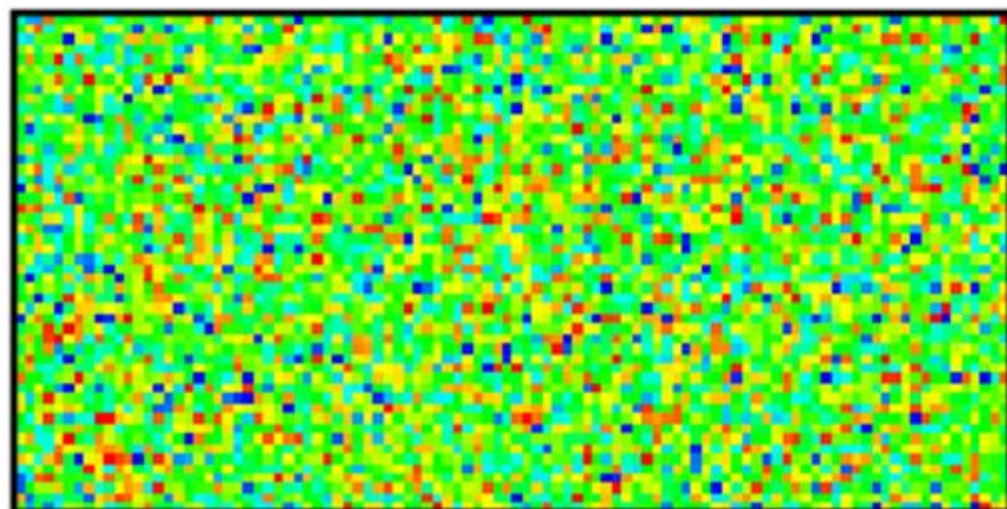
- Una particularidad de los datos utilizados en las ciencias del terreno, es que la información está asociada a una **localización en el espacio**.
- Las características espaciales de los datos tales como la localización de los **valores extremos**, la **tendencia** general, el grado de **continuidad**, etc., son de considerable interés.
- El objetivo de la descripción espacial es describir y cuantificar la **relación entre medidas del mismo atributo y/o de atributos distintos en dos localizaciones**.

# Continuidad espacial

Datos estandarizados:  $[x - \text{prom}(x)] / \text{desv.std.}$



Datos	5000
Media	0.18
Varianza	0.80
Máximo	2.32
Cuartil 3	0.77
Mediana	0.23
Cuartil 1	-0.36
Mínimo	-2.43



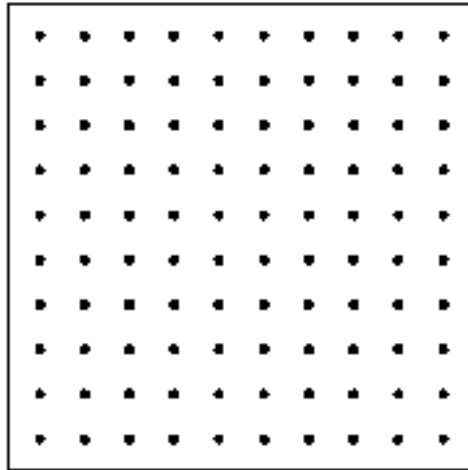
Datos	5000
Media	0.18
Varianza	0.80
Máximo	2.32
Cuartil 3	0.77
Mediana	0.23
Cuartil 1	-0.36
Mínimo	-2.43

# Continuidad espacial

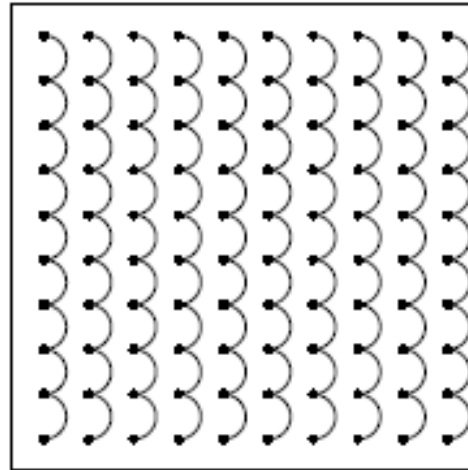
- Las herramientas para realizar la **descripción univariada** de la información no son capaces de capturar las características espaciales de un atributo.
- Las herramientas utilizadas para describir la **relación entre dos variables** pueden utilizarse para describir la relación entre el valor de una variable y los valores de la misma variable en localizaciones cercanas.
- El **diagrama de dispersión** es utilizado para mostrar la continuidad espacial.
- Los **estadísticos bivariados** clásicos sirven para sumarizar la información contenida en un diagrama de dispersión.

# Diagramas de dispersión tipo h

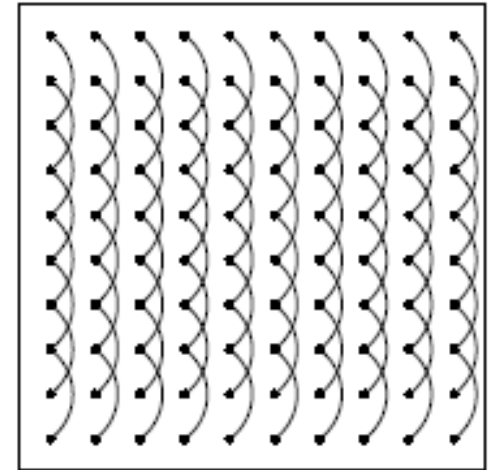
$h = (0,0)$



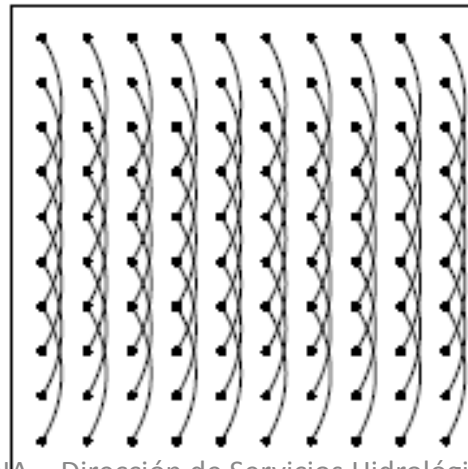
$h = (0,1)$



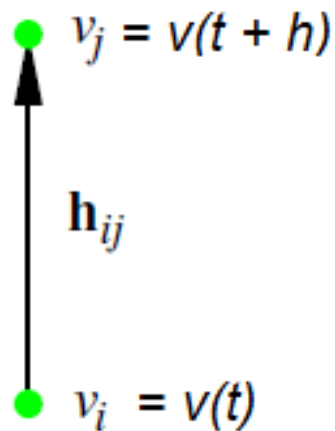
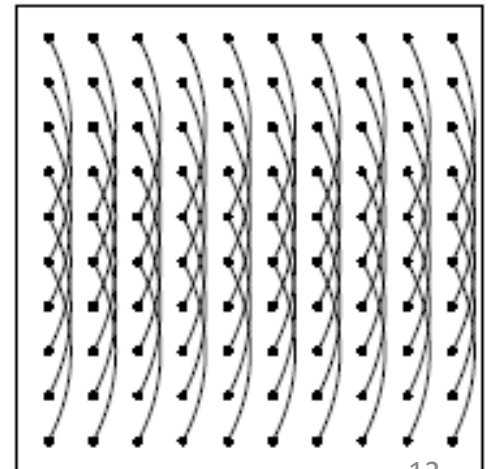
$h = (0,2)$



$h = (0,3)$



$h = (0,4)$

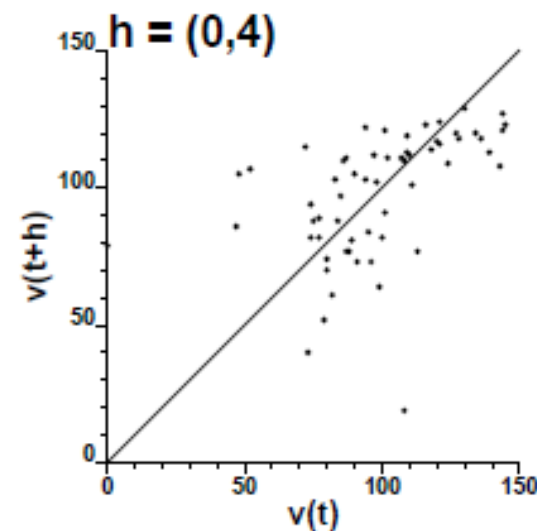
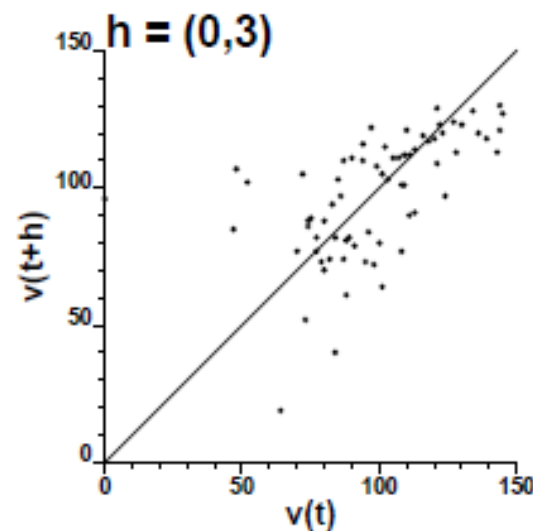
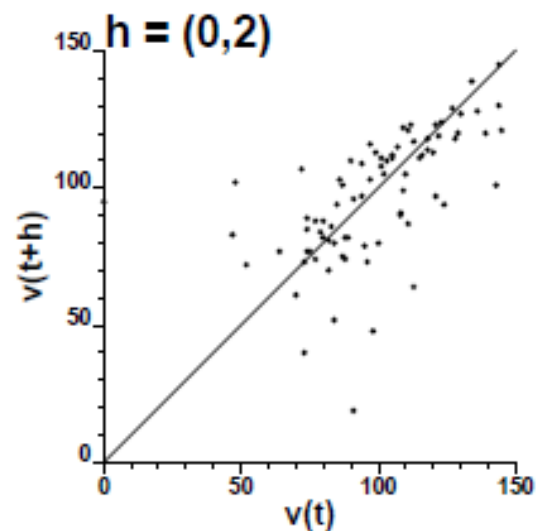
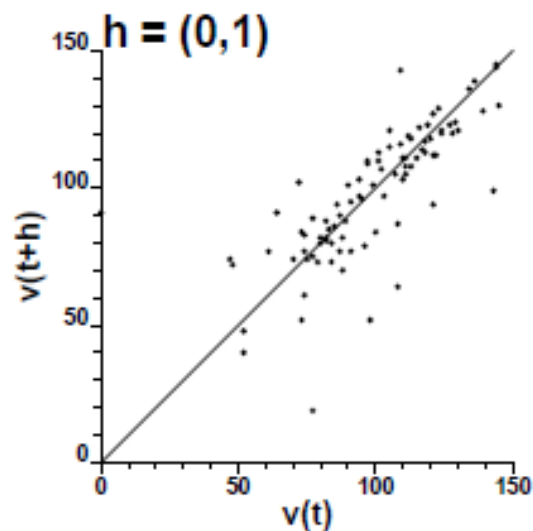
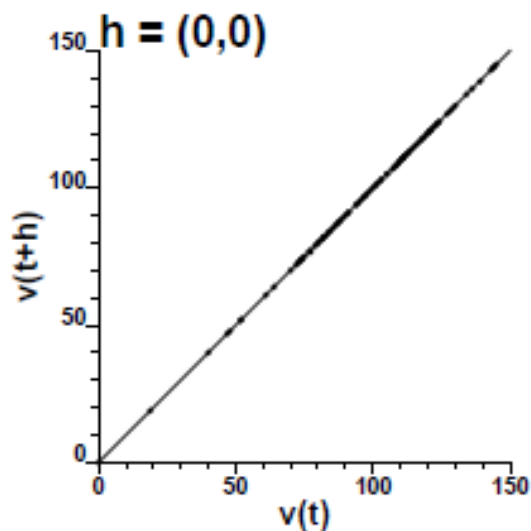


# Mapa con la localización de los Datos

81	77	103	112	123	19	40	111	114	120
•	•	•	•	•	•	•	•	•	•
82	61	110	121	119	77	52	111	117	124
•	•	•	•	•	•	•	•	•	•
82	74	97	105	112	91	73	115	118	129
•	•	•	•	•	•	•	•	•	•
88	70	103	111	122	64	84	105	113	123
•	•	•	•	•	•	•	•	•	•
89	88	94	27	116	108	73	107	118	127
•	•	•	•	•	•	•	•	•	•
77	82	86	101	109	113	79	102	120	121
•	•	•	•	•	•	•	•	•	•
74	80	85	90	97	101	96	72	128	130
•	•	•	•	•	•	•	•	•	•
75	80	83	87	94	99	95	48	139	145
•	•	•	•	•	•	•	•	•	•
77	84	74	108	121	143	91	52	136	144
•	•	•	•	•	•	•	•	•	•
87	100	47	111	124	109	0	98	134	144
•	•	•	•	•	•	•	•	•	•



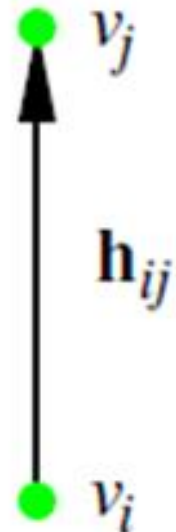
## Diagramas de dispersión tipo h



# Medidas de continuidad espacial

Al considerar el vector separación  $\mathbf{h}$ .

- Función de covarianza  $C(\mathbf{h})$ .
- Función de correlación o correlograma  $\rho(\mathbf{h})$ .
- Variograma  $\gamma(\mathbf{h})$ .



# Función de covarianza

$$C(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{(i,j) | \mathbf{h}_{ij} = \mathbf{h}} v_i \cdot v_j - m_{-\mathbf{h}} \cdot m_{+\mathbf{h}}$$

$$m_{-\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{i | \mathbf{h}_{ij} = \mathbf{h}} v_i \quad m_{+\mathbf{h}} = \frac{1}{N(\mathbf{h})} \sum_{j | \mathbf{h}_{ij} = \mathbf{h}} v_j$$

Recordar que:  $C_{XY} = \frac{1}{n} \sum_{i=1}^n x_i \cdot y_i - m_X \cdot m_Y$



# Función de correlación o correlograma

$$\rho(\mathbf{h}) = \frac{C(\mathbf{h})}{\sigma_{-\mathbf{h}} \cdot \sigma_{+\mathbf{h}}}$$

$$\sigma_{-\mathbf{h}}^2 = \frac{1}{N(\mathbf{h})} \sum_{i|\mathbf{h}_{ij}=\mathbf{h}} v_i^2 - m_{-\mathbf{h}}^2 \quad \sigma_{+\mathbf{h}}^2 = \frac{1}{N(\mathbf{h})} \sum_{j|\mathbf{h}_{ij}=\mathbf{h}} v_j^2 - m_{+\mathbf{h}}^2$$

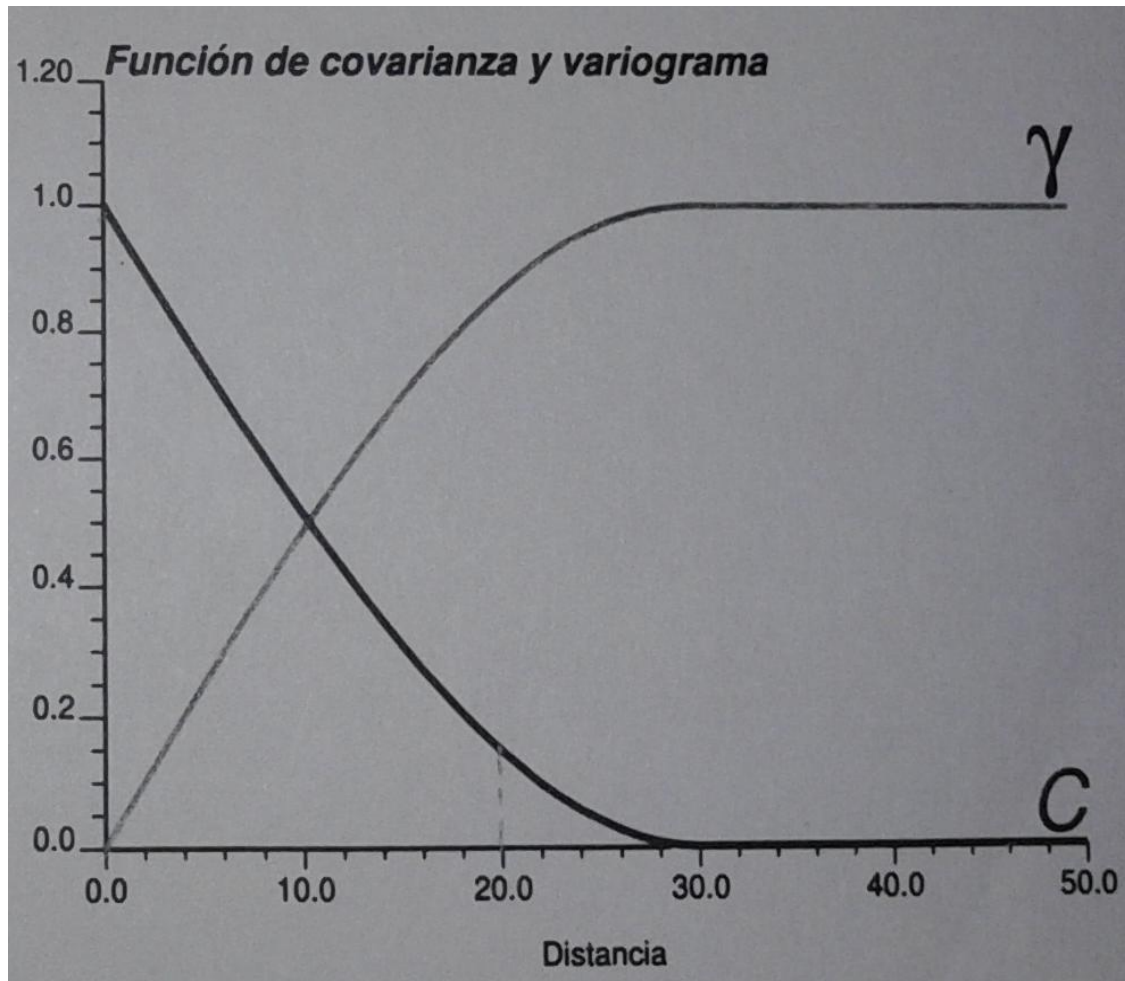
Recordar que:  $\rho_{XY} = \frac{C_{XY}}{\sigma_X \cdot \sigma_Y}$

# Variograma

$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{(i,j) | \mathbf{h}_{ij} = \mathbf{h}} (v_i - v_j)^2$$

Recordar que:  $M_{XY} = \frac{1}{2n} \sum_{i=1}^n (x_i - y_i)^2$

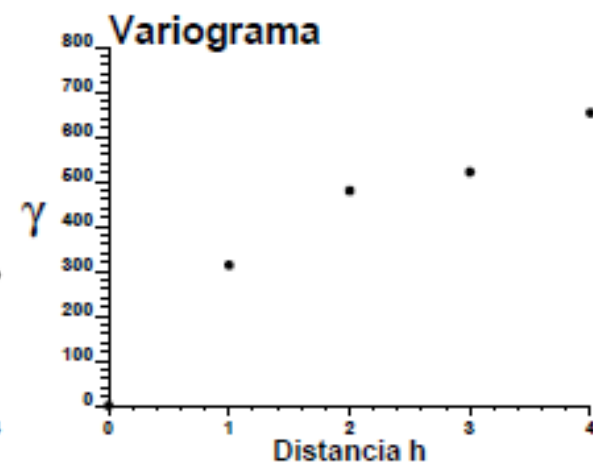
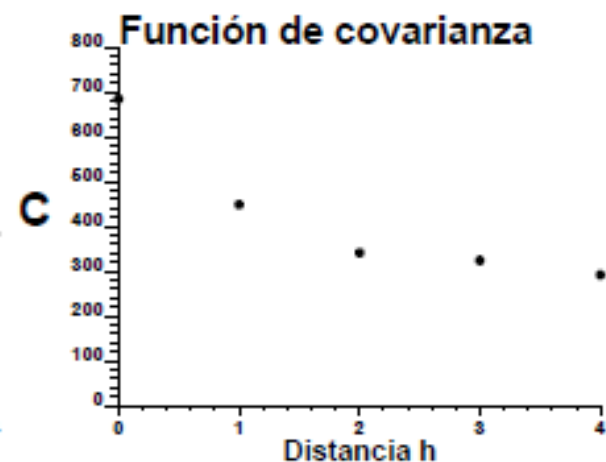
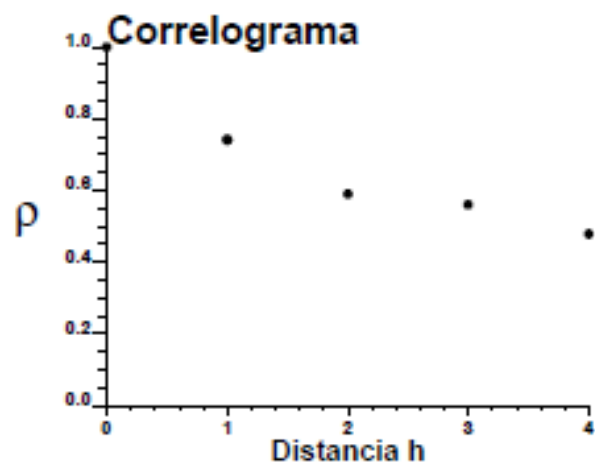
# Relación variograma - covarianza



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

# Medidas de continuidad espacial

Distancia h	Coefficiente de correlación $\rho$	Covarianza C	Momento de inercia $\gamma$
(0,0)	1.000	686.4	0.000
(0,1)	0.742	448.8	312.8
(0,2)	0.590	341.0	479.2
(0,3)	0.560	323.8	521.4
(0,4)	0.478	291.5	652.9



## Datos irregularmente espaciados

- En la práctica difícilmente dispongamos de un conjunto de datos **exhaustivamente** conocido como el utilizado hasta ahora en nuestro análisis.
- En su lugar dispondremos de un conjunto de datos que en general es una pequeña fracción del conjunto de datos exhaustivo y que se encuentran casi siempre **irregularmente distribuidos en el espacio**.
- Las herramientas para realizar la descripción espacial de datos exhaustivamente conocidos deben ser adaptadas para que sean útiles a la hora de analizar la continuidad espacial de datos escasamente muestreados e irregularmente distribuidos en el espacio.

## Datos irregularmente espaciados

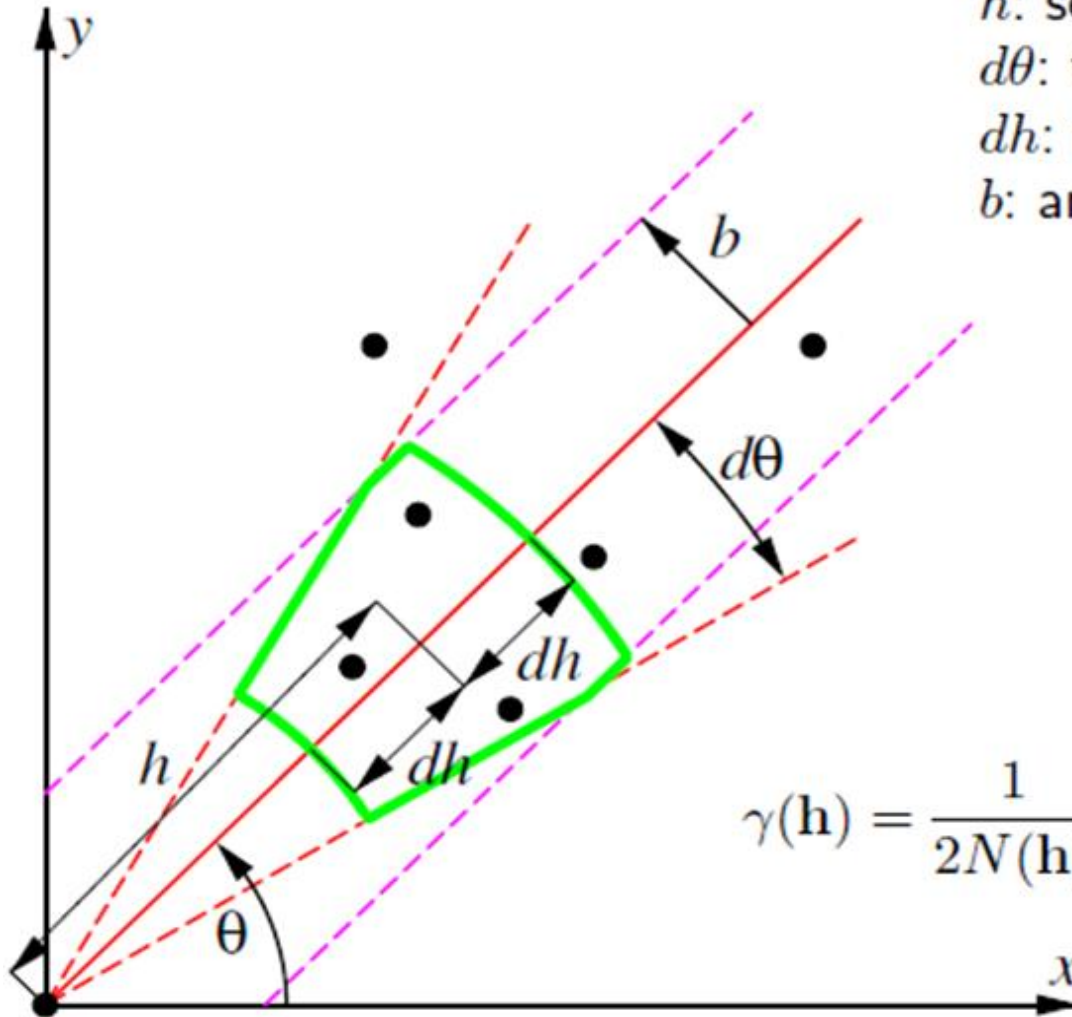
$\theta$ : dirección

$h$ : separación

$d\theta$ : tolerancia angular

$dh$ : tolerancia lineal

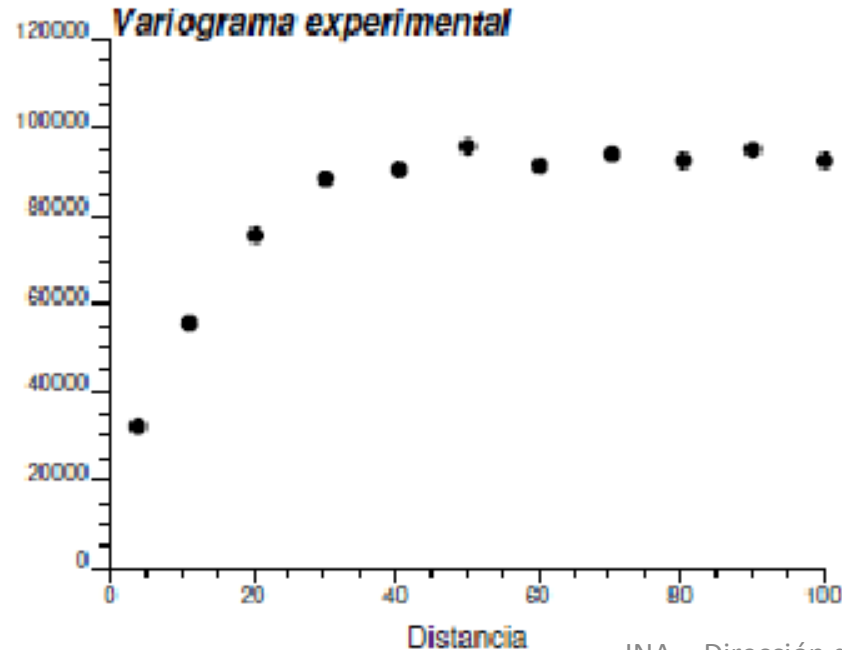
$b$ : ancho de banda



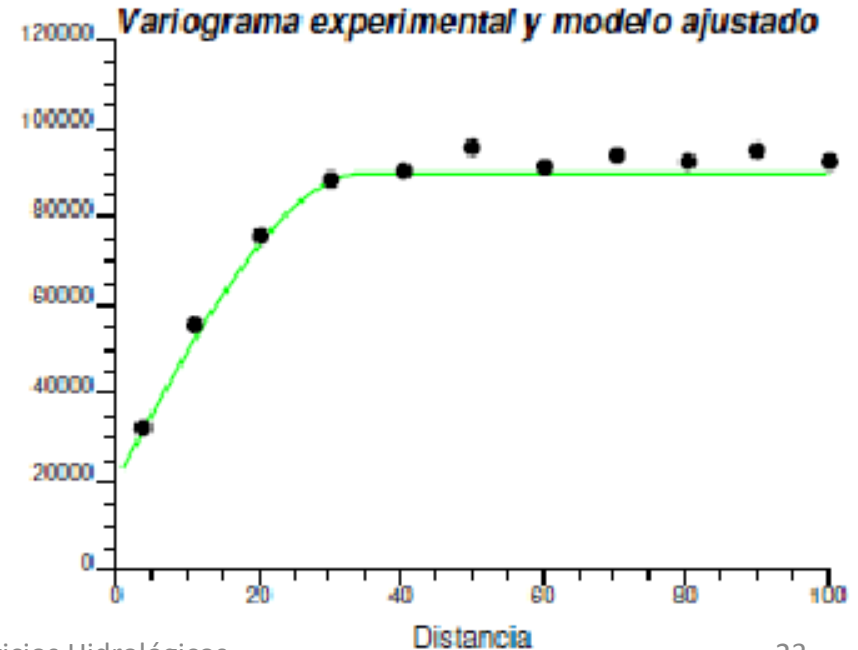
$$\gamma(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{(i,j)|\mathbf{h}_{ij} \approx \mathbf{h}} (v_i - v_j)^2$$

# Práctica de la descripción espacial

La expresión del variograma nos permite cuantificar la continuidad espacial para una serie de distancias de separación y direcciones determinando así el **variograma experimental**.



Los métodos geoestadísticos requieren conocer el variograma para cualquier distancia de separación y cualquier dirección en la cual es considerada dicha separación, por lo que es necesario ajustar un **modelo** al variograma experimental.

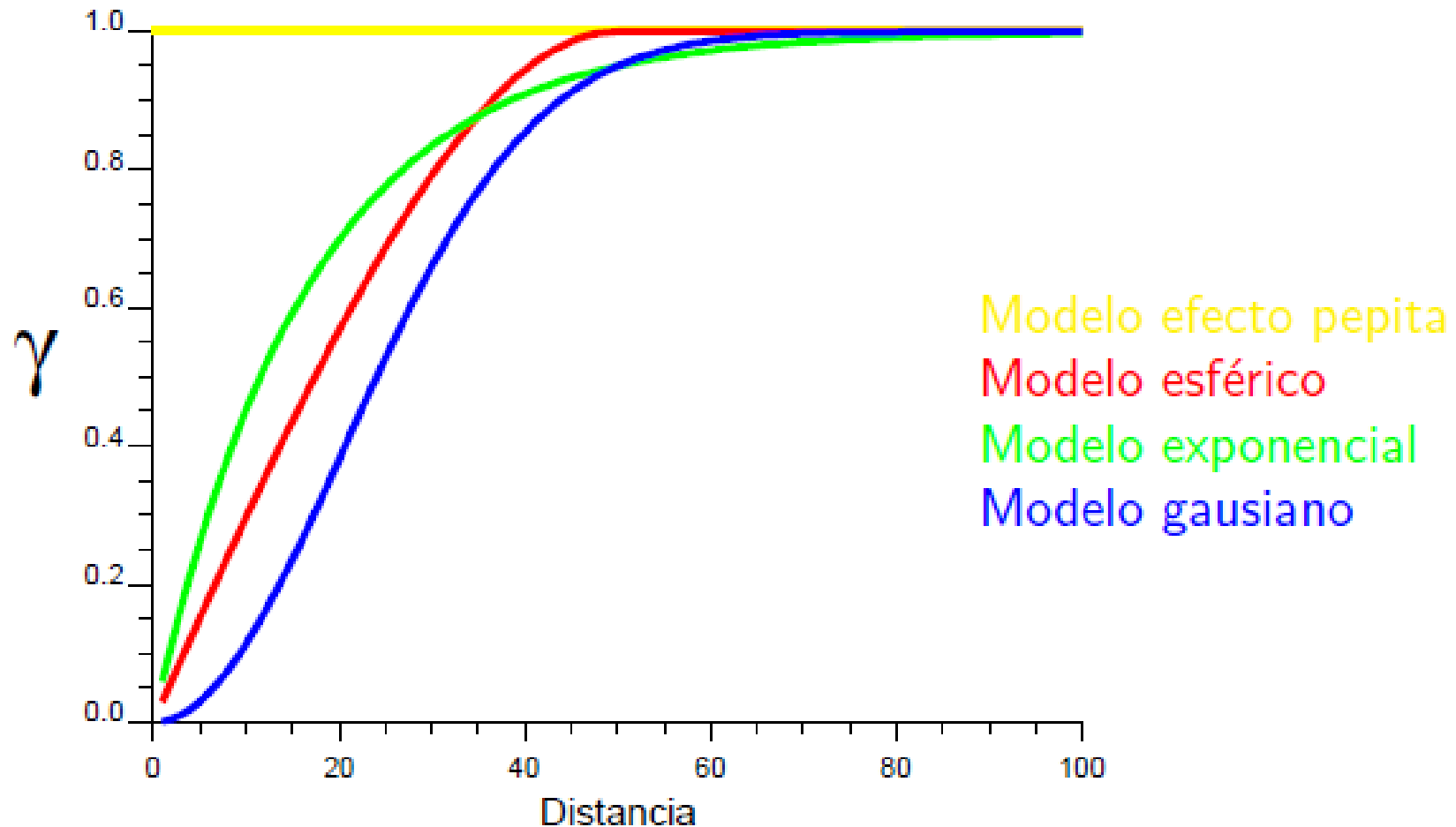


# Metodología

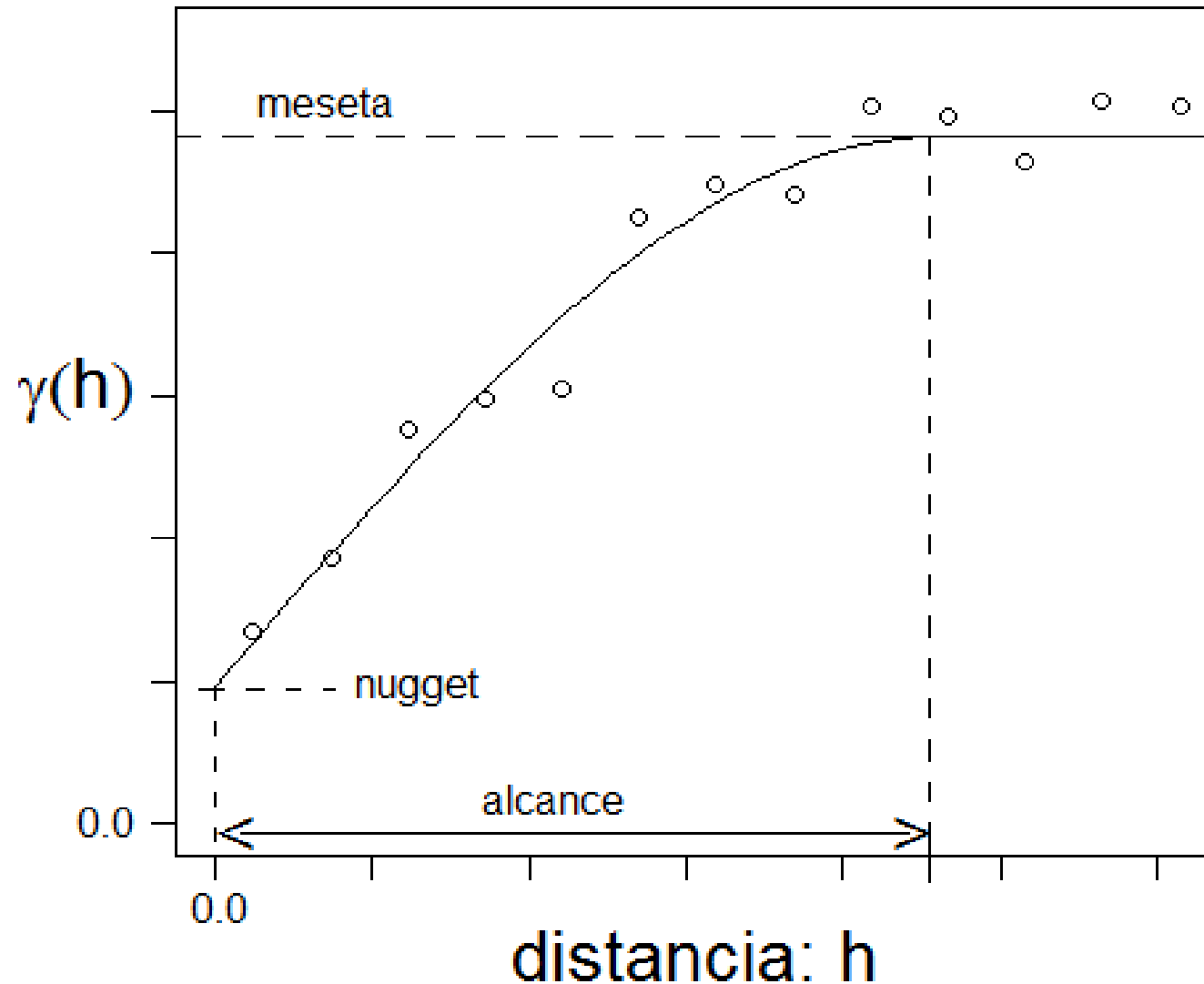
- Cálculo del **variograma omnidireccional**: permite tener una idea inicial acerca de los parámetros relativos a distancia (alcance e incremento de la distancia de separación).
- Cálculo de los **variogramas direccionales**: permiten determinar la existencia de anisotropía y sus direcciones principales.
- Ajuste de un **modelo** a los variogramas experimentales: permite conocer el valor del variograma para cualquier distancia de separación y dirección.



# Modelos básicos



# Variograma experimental + Modelo



- Omnidireccional
- Variograma experimental
- Modelo esférico
- Meseta
- Nugget
- Alcance

# Isotropía - Anisotropía

- **Isotropía**

la variación en superficie de la variable  $v$  es igual para todas las direcciones.

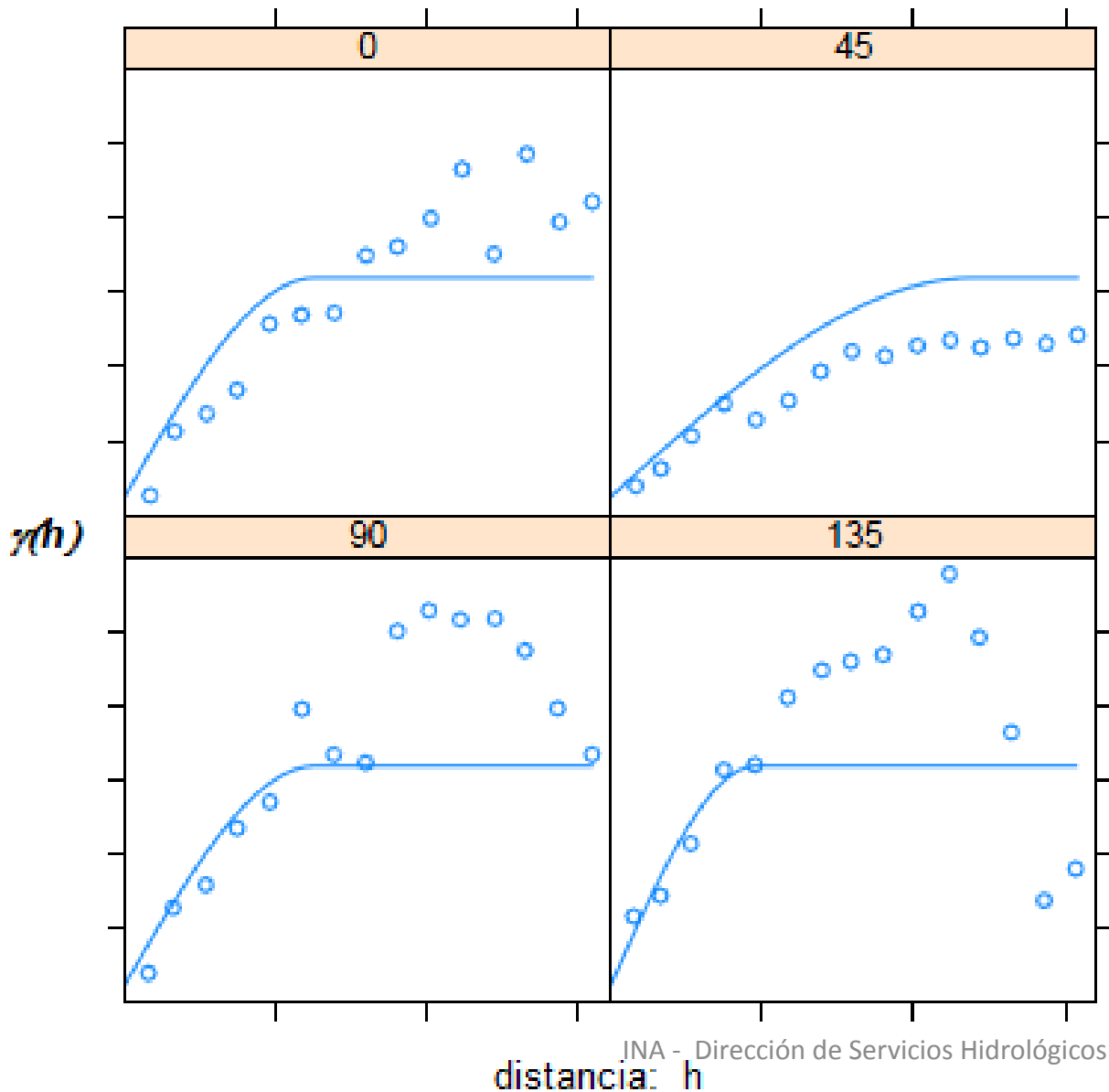
- **Anisotropía**

Existen direcciones de mayor (y menor) variabilidad superficial de  $v$ .

- **Detección**

Mediante variogramas direccionales.

# Anisotropía: Variogramas Direccionales



- 45 grados (N-S)  
máxima  
continuidad y  
alcance
- 135 grados(N-S)  
mínima  
continuidad y  
alcance

# Estimación

- El objetivo de un estudio geoestadístico no es solamente **describir la información** y **construir un modelo de continuidad espacial**.
- La información acerca de los datos y el modelo de continuidad espacial nos permiten realizar la **estimación** de los valores de una o varias variables donde ellas no fueron muestreadas.
- Todos los métodos de estimación geoestadísticos construyen cada estimador como una **combinación lineal ponderada de la información disponible**, esto es:

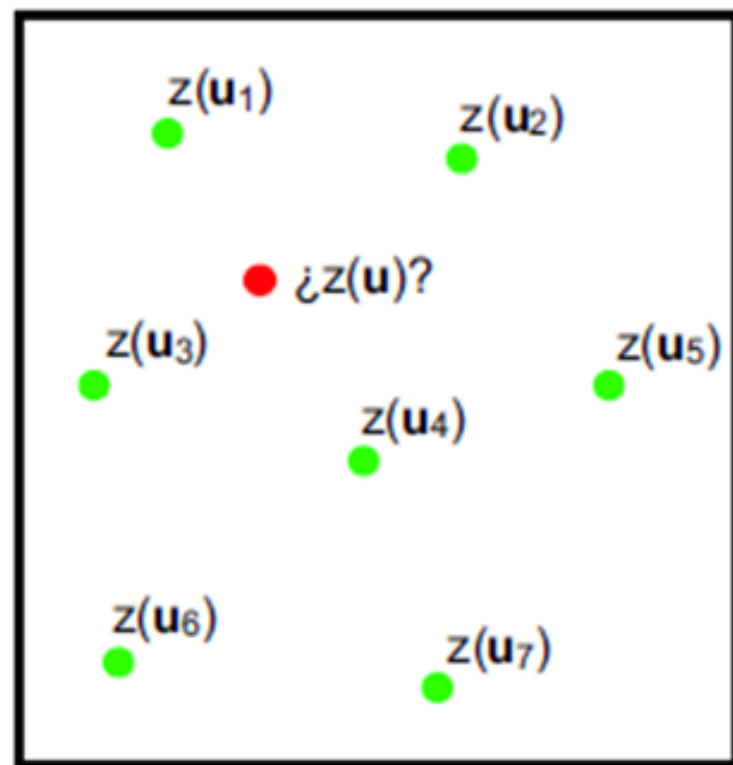
$$z^* = \sum_{i=1}^n \lambda_i \cdot z_i$$

# El kriging

- La idea básica del **kriging** es estimar el valor desconocido de un atributo  $z$  en el punto de coordenadas  $u$ , a partir de  $n$  valores conocidos de  $z$ , cuyas coordenadas son  $u_\alpha$ , con  $\alpha = 1, \dots, n$ .
- Por ejemplo, el **estimador por kriging ordinario** tiene la forma siguiente:

$$z_{KO}^*(u) = \sum_{\alpha=1}^n \lambda_\alpha z(u_\alpha)$$

donde el \* denota que se trata de un estimador y  $\lambda_\alpha(u)$  son coeficientes de ponderación desconocidos a priori.



# Krigeado Ordinario

- **Objetivo:** obtener los coeficientes  $\lambda_\alpha$  (para cada punto  $u$ ), satisfaciendo:
  1. Estimación **insesgada**: el error esperado sea 0.
  2. Varianza del error  $\sigma_R^2$  **mínima**, minimizando:

$$\sigma_R^2 = \sigma^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_\alpha \lambda_\beta C(u_\alpha - u_\beta) - 2 \sum_{\alpha=1}^n \lambda_\alpha C(u_\alpha - u)$$

- Se buscan los  $\lambda_\alpha$  que minimizan  $\sigma_R^2$  con la **restricción**:

$$\sum_{\beta=1}^n \lambda_\beta = 1$$

# Krigeado Ordinario

- **Derivando** la expresión:

$$\sigma_R^2 = \sigma^2 + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(u_{\alpha} - u_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(u_{\alpha} - u) + 2\mu \left( \sum_{\beta=1}^n \lambda_{\beta} - 1 \right)$$

(con multiplicador de Lagrange  $2\mu$ ), respecto de los  $\lambda_{\alpha}$  e **igualando a 0** queda el sistema:

$$\frac{\partial \sigma_R^2}{\partial \lambda_{\alpha}} = 0 \Rightarrow \sum_{\beta=1}^n \lambda_{\beta} C(u_{\alpha} - u_{\beta}) + \mu = C(u_{\alpha} - u); \quad \alpha = 1 \dots n$$

- ... más la restricción: 
$$\sum_{\beta=1}^n \lambda_{\beta} = 1$$



# Krigeado Ordinario

- Lo anterior supone resolver un sistema de  $n + 1$  ecuaciones lineales denominado **sistema de krigeado ordinario**, que se escribe:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu = C(\mathbf{u}_{\alpha} - \mathbf{u}), & \alpha = 1, \dots, n \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \end{cases}$$

donde  $C(\cdot)$  es la función de covarianza que se obtiene de la información disponible.

# Krigeado Ordinario

- Forma matricial del sistema de KO:

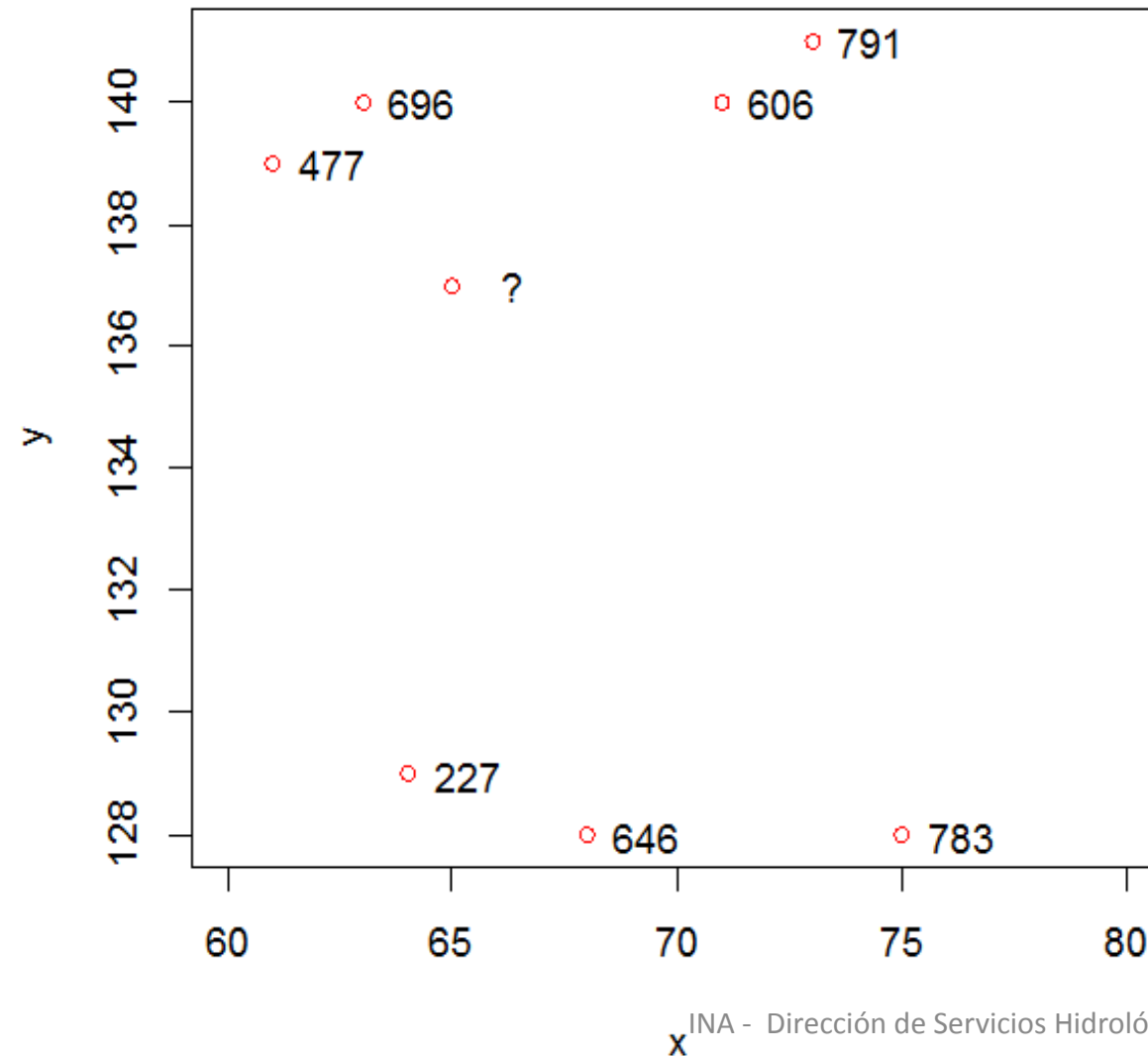
$$\begin{pmatrix} C(u_1 - u_1) & \dots & C(u_1 - u_n) & 1 \\ \dots & \dots & \dots & \dots \\ C(u_n - u_1) & \dots & C(u_n - u_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \dots \\ \lambda_n \\ \mu \end{pmatrix} = \begin{pmatrix} C(u_1 - u) \\ \dots \\ C(u_n - u) \\ 1 \end{pmatrix}$$

$$C.\vec{\lambda} = D$$

- Multiplicando a izquierda por  $C^{-1}$  se despeja:

$$\vec{\lambda} = C^{-1}.D = (\lambda_1, \dots, \lambda_n, 1)^T$$

# Ejemplo



Punto	x	y	valor
0	65	137	?
1	61	139	477
2	63	140	696
3	64	129	227
4	68	128	646
5	71	140	606
6	73	141	791
7	75	128	783

# Ejemplo

**Distancias:**

<b>Punto</b>	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>	<b>5</b>	<b>6</b>	<b>7</b>
<b>0</b>	0.00	4.47	3.61	8.06	9.49	6.71	8.94	13.45
<b>1</b>	4.47	0.00	2.24	10.44	13.04	10.05	12.17	17.80
<b>2</b>	3.61	2.24	0.00	11.05	13.00	8.00	10.05	16.97
<b>3</b>	8.06	10.44	11.05	0.00	4.12	13.04	15.00	11.05
<b>4</b>	9.49	13.04	13.00	4.12	0.00	12.37	13.93	7.00
<b>5</b>	6.71	10.05	8.00	13.04	12.37	0.00	2.24	12.65
<b>6</b>	8.94	12.17	10.05	15.00	13.93	2.24	0.00	13.15
<b>7</b>	13.45	17.80	16.97	11.05	7.00	12.65	13.15	0.00

modelo de covarianza :  $C(h) = 10\exp(-0.3h)$

$$\vec{\lambda} = C^{-1}.D$$

# Ejemplo

## C: Matriz de Covariancias entre pares de puntos con observación

10.00	5.11	0.44	0.20	0.49	0.26	0.05	1
5.11	10.00	0.36	0.20	0.91	0.49	0.06	1
0.44	0.36	10.00	2.90	0.20	0.11	0.36	1
0.20	0.20	2.90	10.00	0.24	0.15	1.22	1
0.49	0.91	0.20	0.24	10.00	5.11	0.22	1
0.26	0.49	0.11	0.15	5.11	10.00	0.19	1
0.05	0.06	0.36	1.22	0.22	0.19	10.00	1
1	1	1	1	1	1	1	0

$$\begin{pmatrix} C(u_1 - u_1) & \dots & C(u_1 - u_n) & 1 \\ \dots & \dots & \dots & \dots \\ C(u_n - u_1) & \dots & C(u_n - u_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix}.$$

# Ejemplo

$C^{-1}$ :

								<b>D</b>
0.127	-0.077	-0.013	-0.009	-0.008	-0.009	-0.012	0.1	2.61
-0.077	0.129	-0.01	-0.008	-0.015	-0.008	-0.011	0.1	3.39
-0.013	-0.01	0.098	-0.042	-0.01	-0.01	-0.014	0.2	0.89
-0.009	-0.008	-0.042	0.102	-0.009	-0.009	-0.024	0.1	0.58
-0.008	-0.015	-0.01	-0.009	0.13	-0.077	-0.012	0.1	1.34
-0.009	-0.008	-0.01	-0.009	-0.077	0.126	-0.013	0.1	0.68
-0.012	-0.011	-0.014	-0.024	-0.012	-0.013	0.085	0.2	0.18
0.136	0.121	0.156	0.139	0.118	0.141	0.189	-2.2	1

$$\begin{pmatrix} C(u_1 - u) \\ \dots \\ C(u_n - u) \\ 1 \end{pmatrix}$$

# Ejemplo

$$\lambda = C^{-1} * D = \begin{array}{c|c} \lambda_1 & 0.173 \\ \lambda_2 & 0.318 \\ \lambda_3 & 0.129 \\ \lambda_4 & 0.086 \\ \lambda_5 & 0.151 \\ \lambda_6 & 0.057 \\ \lambda_7 & 0.086 \\ \lambda_8 & -0.907 \end{array}$$

- Valor  $Z^*(u_0) = 0.173 * 477 + 0.318 * 696 + \dots - 0.907 * 783$   
 $Z^*(u_0) = 592.7$

# Caso de aplicación

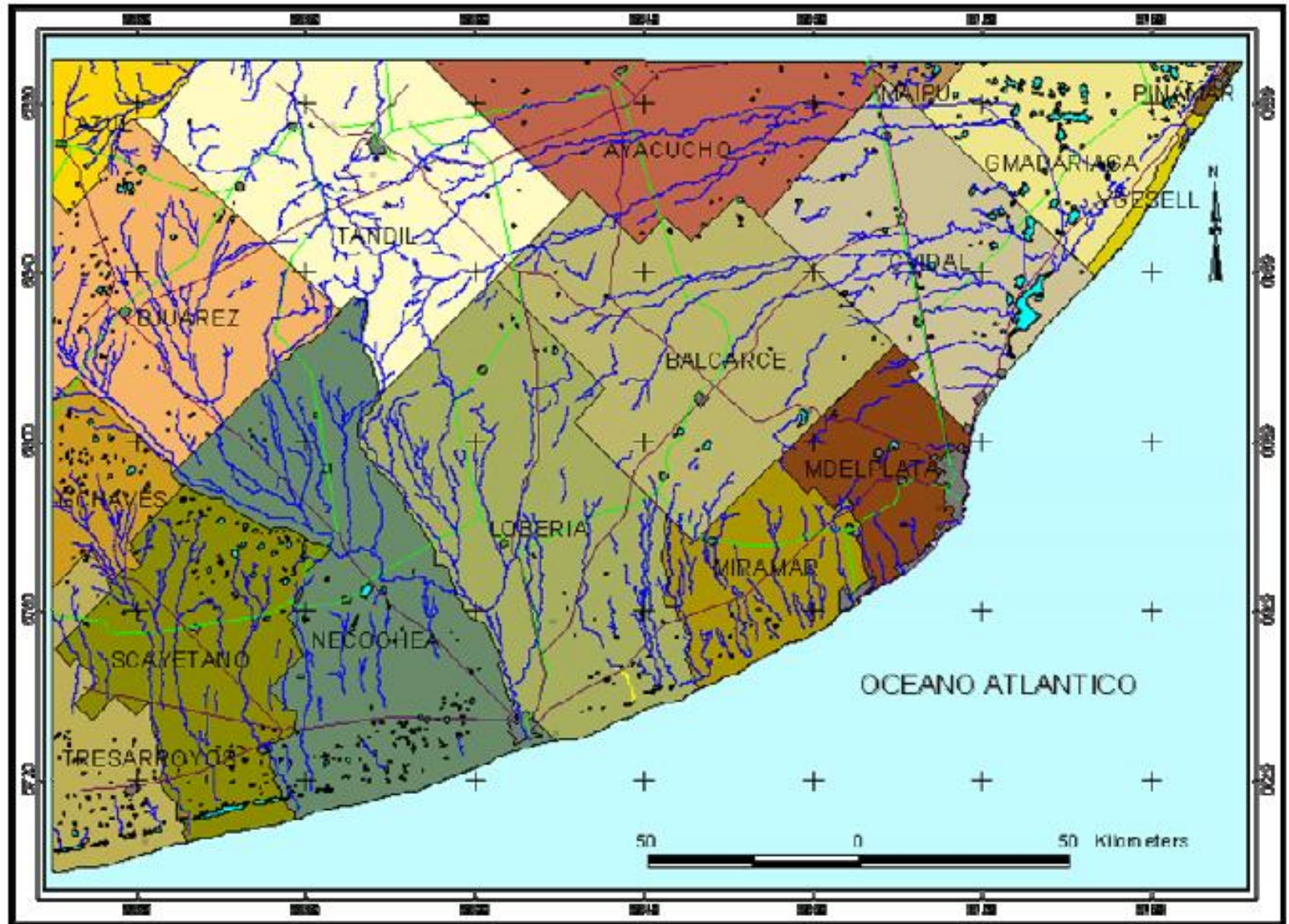
- **Aplicación** de krigeado ordinario (KO) a datos multivariados de aguas subterráneas.
- **Región:** SE de la Prov. de Bs. As. 5500 m – 5700 m (O - E), y 5700 m – 5850 m (S – N).
- **Datos:**

Obs	X	Y	pH	CE	HCO3	Cl	SO4	Ca	Mg	Na	K	RAS
1	5502.97	5764.71	7.5	760.0	317.2	28.4	72.0	42.0	24.0	73.6	7.8	2.23
2	5505.55	5718.80	7.4	3603.3	608.0	669.8	446.4	46.7	44.0	728.3	18.2	18.4
...	...	...	...	...	...	...	...	...	...	...	...	...
44	5691.80	5769.80	7.3	2000.0	652.7	262.7	115.2	24.0	12.0	414.0	11.7	17.1



# Caso de aplicación

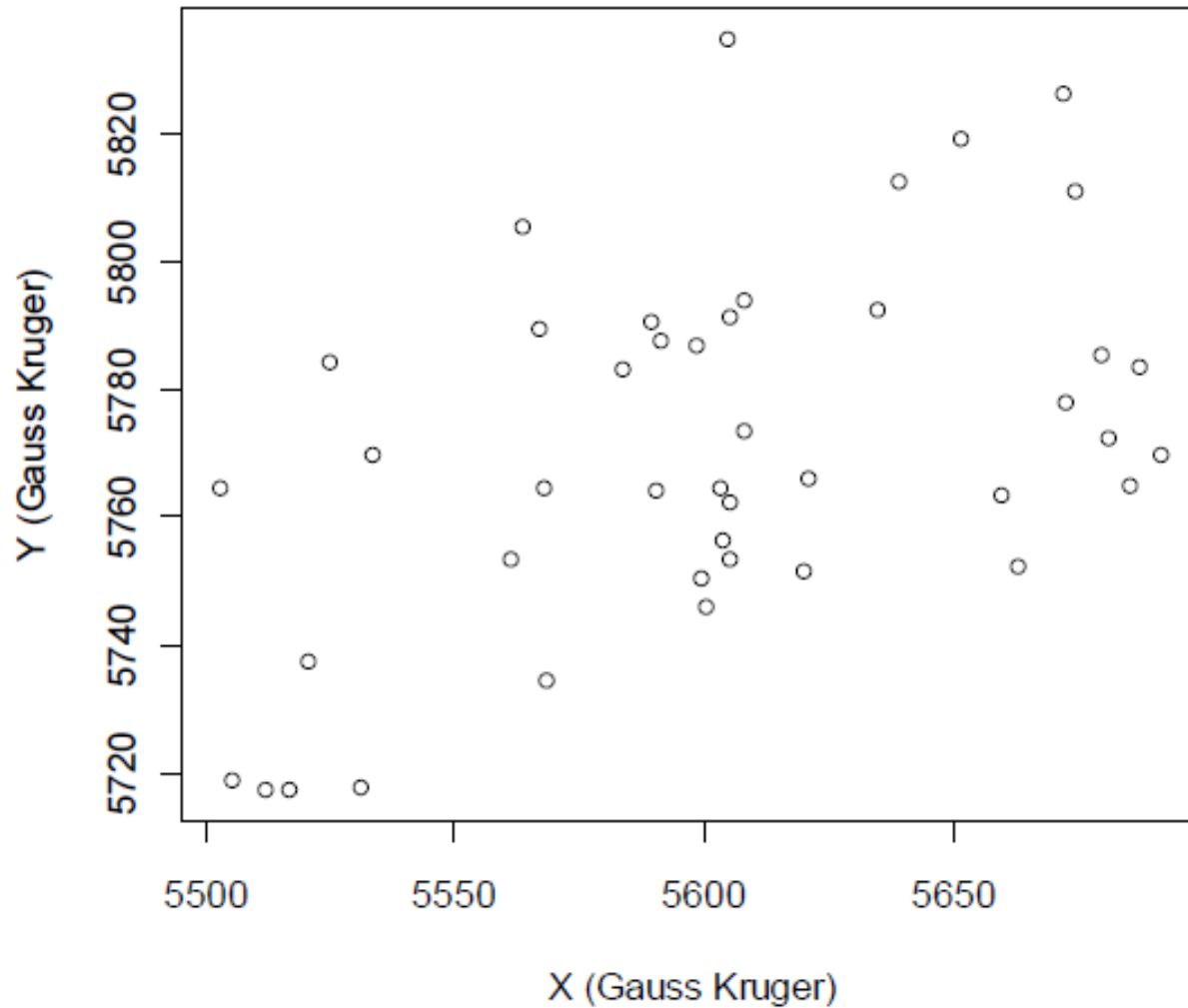
## Región



# Caso de aplicación

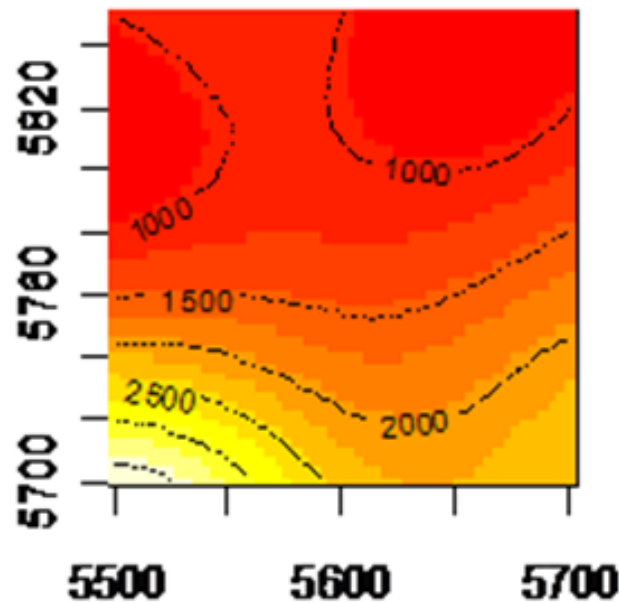
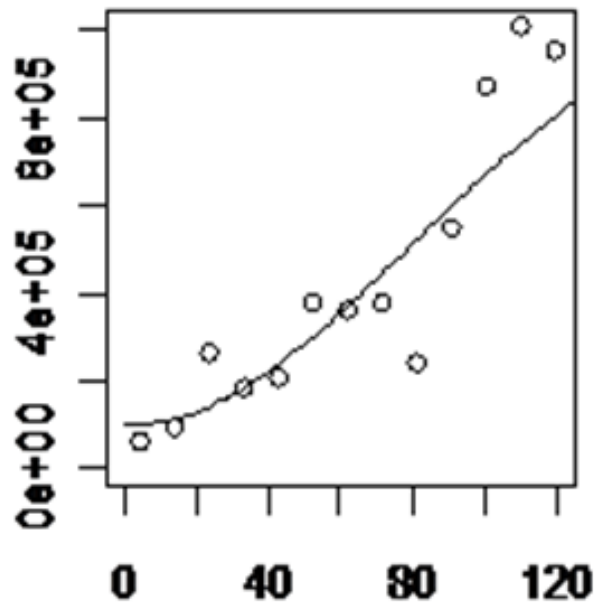
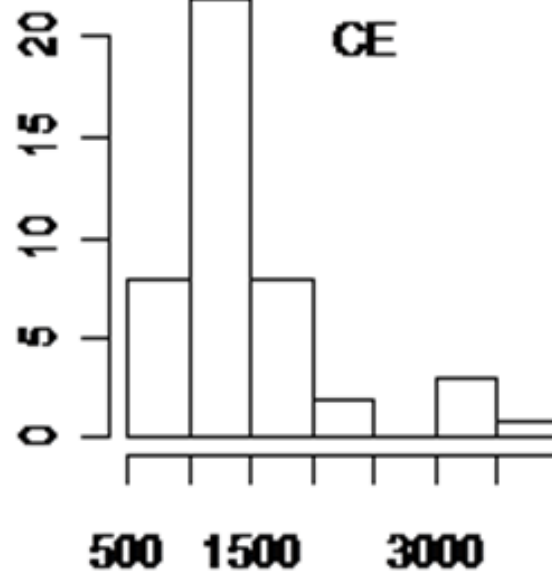
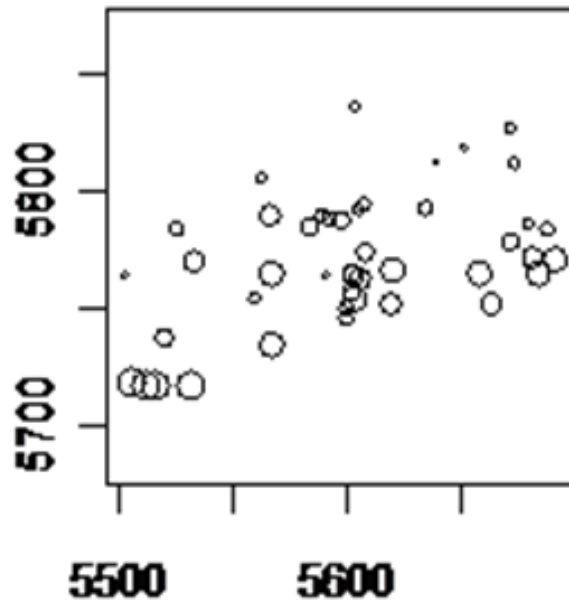
Datos localizados

puntos de observación



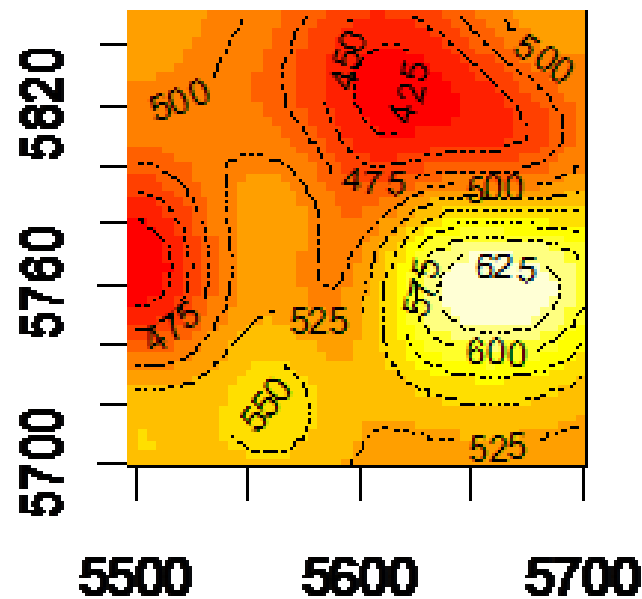
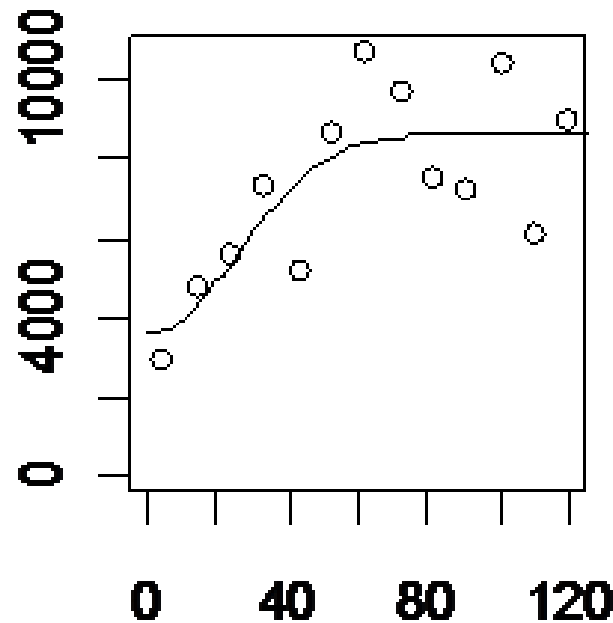
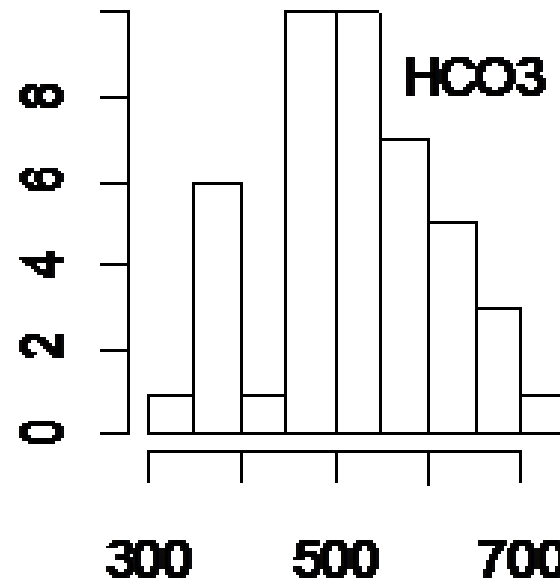
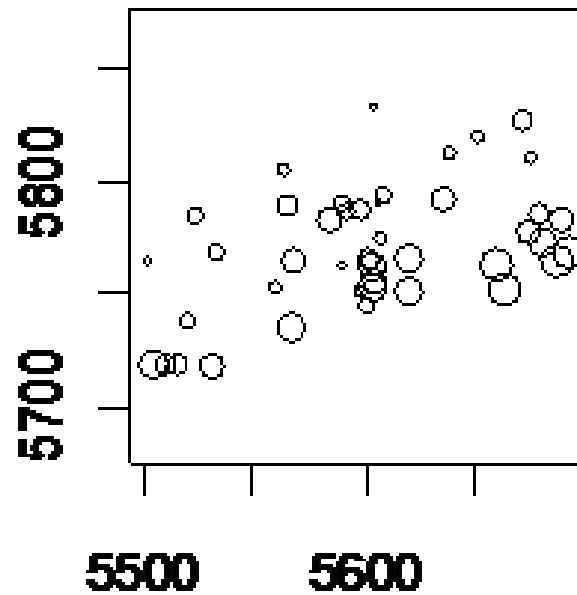
# Caso de aplicación

## Modelo gausseano



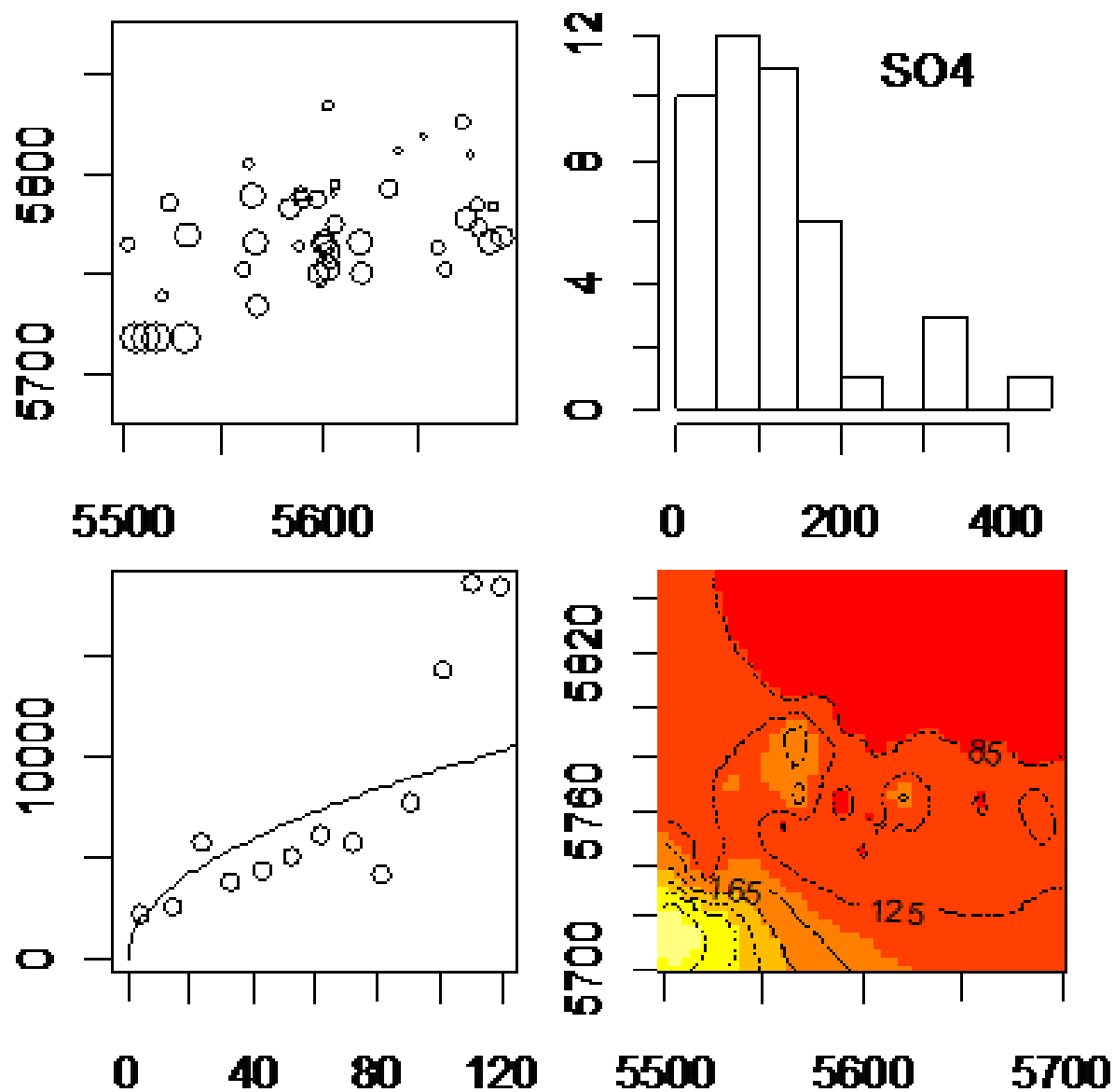
# Caso de aplicación

## Modelo gausseano



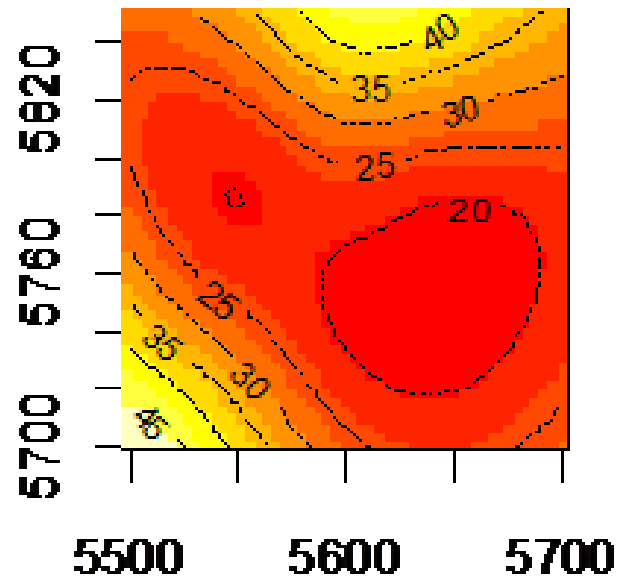
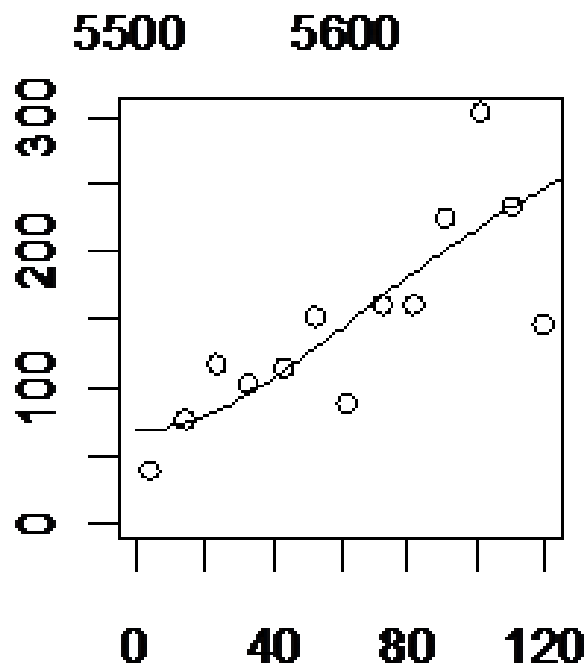
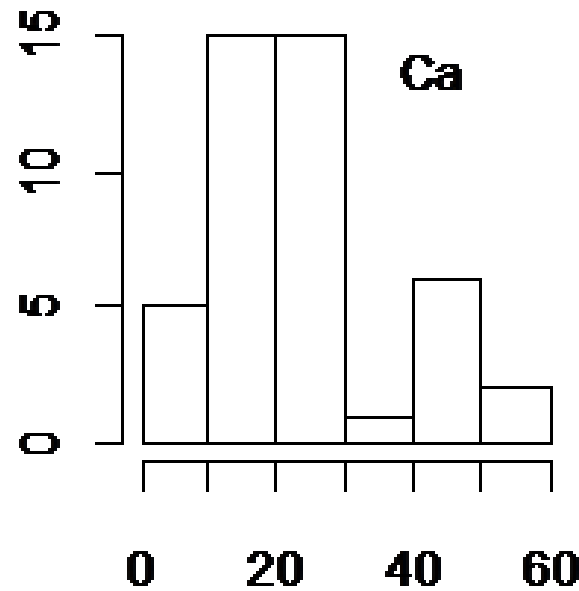
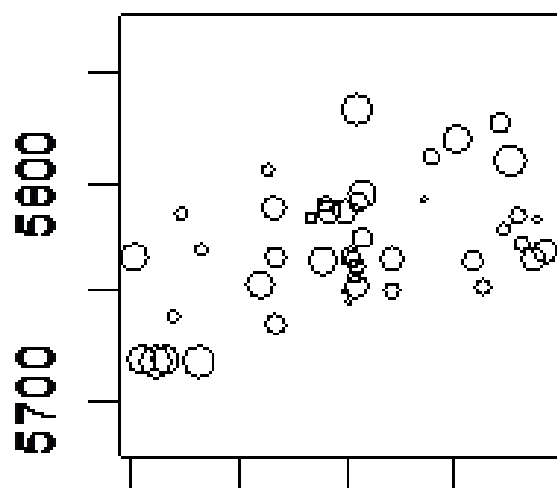
# Caso de aplicación

## Modelo exponencial



# Caso de aplicación

## Modelo circular



# ***GEO ESTADÍSTICA***

***¡Hasta la próxima!***

***Daniel***