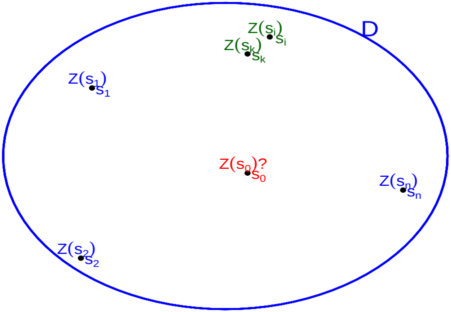
GEOESTADISTICA

Parte 1 : Variograma

Marzo 2019

Geoestadistica

$D \subset \mathbb{R}^2$
continuo



$Z = \{Z(s)\}_{s\in D}$ un proceso estocastico (campo aleatorio) i.e., una colleccion de variables aleatorias indexadas espacialmente $s \in D$ $Z(s_1), Z(s_2), \ldots, Z(s_n)$ observados,

Preguntas

- $Z(s_0)$ no medida, prediccion $\widehat{Z}(s_0)$ de $Z(s_0)$, de $\int_B Z(s)ds$?
- estimar la distribucion de probabilidad de $Z(s)$, de $\varphi(Z(s))$?
- estimacion de la relaciones de dependencia de $(Z(s_i), Z(s_k))$?

Estimacion de la media

Hypotesis

- $E(Z(s)) = \mu \qquad \forall s \in D$
- $\text{cov}(Z(s_i), Z(s_j)) = c_{ij}$

$Z(s_1), Z(s_2), \ldots, Z(s_n)$ observados
Estimacion de $E(Z(s)) = \mu$?

restricciones

- estimador lineal : $\widehat{\mu} = \sum_{i=1}^{n} \lambda_i Z(s_i)$
- insesgado : $\text{E}(\widehat{\mu}) = \mu$
- varianza $\text{Var}(\widehat{\mu} - \mu)$ minima

BLUE (Best Linear Unbiased Estimator)

Estimacion de la media

(λ_i) son soluciones de

$$\min \sum_{i,j=1}^{n} \lambda_i \lambda_j c_{ij} \qquad ; \qquad \sum_{i=1}^{n} \lambda_i = 1$$

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} & 1 \\ c_{21} & c_{22} & \dots & c_{2n} & 1 \\ \\ \dots & \dots & \dots & \dots & \dots \\ \\ c_{n1} & c_{n2} & \dots & c_{nn} & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \\ \dots \\ \\ \lambda_n \\ m \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \\ \dots \\ \\ 0 \\ 1 \end{pmatrix}$$

$$\widehat{\mu} = \sum_{i=1}^{n} \lambda_i Z(s_i) \qquad ; \qquad \mathsf{Var}(\widehat{\mu} - \mu) = -m$$

Necesita conocer a los c_{ij}

Proceso estacionario

Definiciones

- Z *es un proceso de segundo orden si para todos los* $s \in D, \mathsf{E}(Z(s)^2) < \infty$.
- *La media de* Z *es la funcion*

$$\begin{aligned} m : D &\rightarrow \mathbb{R} \\ s &\rightarrow m(s) = \mathsf{E}(Z(s)) \end{aligned}$$

- *La covarianza de* Z *es la funcion*

$$\begin{aligned} c : D \times D &\rightarrow \mathbb{R} \\ (s,t) &\rightarrow c(s,t) = \mathsf{Cov}(Z(s), Z(t)) \end{aligned}$$

Definicion

Z *es un proceso gaussiano en* D *si para cada parte finita de* $S \subset D$ *y cada succesion real* $a = (a_s, s \in S)$, $\sum_{s \in S} a_s Z(s)$ *es una variable gaussiana.*

Proceso de segundo orden

Propiedad

Una funcion de covarianza c es una funcion semi definida positiva,

$$\forall n \geq 1, \forall a \in \mathbb{R}^n, \forall (s_1, s_2, \ldots, s_n) \sum_{i=1}^{n} \sum_{j=1}^{n} a_i a_j c(s_i, s_j) \geq 0$$

c *es definida positiva si la desigualdad es estricta para* $a \neq 0$*.*

Proceso estacionario

Definicion

Z es un proceso aleatorio de segundo orden estacionario en D si la media m de Z es constante y si la funcion de covarianza c de Z es invariante por translacion:

$$\begin{aligned} \forall s \in D \quad m(s) = \mathsf{E}(Z(s)) &= \mu \\ \forall (s,t) \in D \times D \quad c(s,t) = \mathsf{cov}(Z(s), Z(t)) &= C(t-s) \end{aligned}$$

Definicion

Z es un proceso aleatorio isotropo si es estationario y si su funcion de covarianza verifica

$$\forall h \in \mathbb{R}^2 \qquad C(h) = C(\|h\|)$$

Funciones de covarianza

Propiedad

$C(h)$ *la funcion de covarianza de un campo aleatorio estacionario*

- $\forall h \in \mathbb{R}^2 \quad C(h) = C(-h)$
- $C(0) = \mathit{Var}(Z(s))$
- $\forall h \in \mathbb{R}^2 \quad |C(h)| \leq C(0)$
- $C(h)$ *es definida positiva i.e.*
  $\forall s_1, \ldots, s_n, \quad \forall \lambda_1, \ldots, \lambda_n \quad \sum_{ij} \lambda_i \lambda_j C(s_i - s_j) \geq 0$

Propiedad

- la suma y el producto de dos funciones definidas positivas es una funcion definida positiva.
- las combinaciones lineales positivas (y sus limites) de funciones definidas positivas son definidas positivas.

Funciones de covarianza

Teorema (Bochner)

$C(h)$ es una funcion definida positiva de $\mathbb{R}^d$ si

$$C(h) = \int_{\mathbb{R}^d} \cos(\omega' h) G(d\omega)$$

G es una medida positiva.

Teorema (Shoenberg)

$C(h)$ es una foncion definida positiva isotropa en $\mathbb{R}^d$ si

$$C(h) = \int_{\mathbb{R}} e^{-h^2/t^2} G(dt)$$

G es una medida positiva.

Incrementos estacionarios

Definicion

Z es un proceso con incrementos estacionarios si los incrementos de Z son estacionarios de segundo orden, i.e.

$$E(Z(s+h)-Z(s))=0 \qquad Var(Z(s+h)-Z(s))=2\gamma(h)$$

La estacionaridad de segundo orden conlleva la estacionaridad de los incrementos.

Ejemplo: W_t movimiento Browniano

- $P(W_0=0)=1$
- $\mathrm{E}(W_t)=0, \quad \mathrm{Var}(W_t)=t, \quad \mathrm{E}(W_tW_s)=\min(t,s)$
- los incrementos son independientes
- $\Delta_t^h=W_t-W_{t-h}\sim\mathcal{N}(0,h)$
- $\mathrm{E}(\Delta_t^h)=0, \quad \mathrm{E}(\Delta_t^h\Delta_{t+s}^h)=0, \quad \mathrm{Var}(\Delta_t^h)=h$

Variograma

Definicion

El semivariograma de un proceso con incrementos estacionarios es la funcion

$$\gamma(h) = \frac{1}{2}\mathit{Var}(Z(s+h) - Z(s))$$

Propiedad

- $\gamma(h) \geq 0$, $\gamma(0) = 0$, $\gamma(-h) = \gamma(h)$
- $\lim_{\|h\|\to\infty} \frac{\gamma(h)}{\|h\|^2} = 0$
- *si el proceso es estationario de segundo orden*

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

- *si* $\lim_{|h|\to\infty} \gamma(h) = \ell < +\infty$ *entonces el proceso es estacionario de segundo orden y* $\ell = C(0)$

Variograma

Definicion

Una combinacion lineal autorizada (CLA) es una combinacion lineal que verifica:

$$Var\big(\sum_i \lambda_i Z(s_i)\big) < +\infty$$

Proposicion

Las combinaciones lineales autorizadas para un proceso con incrementos estacionarios son las que verifican

$$\sum_i \lambda_i = 0$$

Propiedad

La funcion variograma $\gamma(h)$ *es condicionalmente definida negativa i.e.*

$$\forall s_1, \ldots, s_n, \quad \forall \lambda_1, \ldots, \lambda_n \; con \; \textstyle\sum_{i=1}^n \lambda_i = 0$$
$$\textstyle\sum_{ij} \lambda_i \lambda_j \gamma(s_i - s_j) \leq 0$$

Variograma

Teorema

Si γ *es una funcion continua en* $\mathbb{R}^d$*, que verifica* $\gamma(0) = 0$*,*
entonces γ *es condicionalmente definida negativa es equivalente a*

$$\forall a > 0 \qquad e^{-a\gamma(h)} \quad \textit{es definida positiva}$$

Funciones variograma admisibles

- pepita $\gamma(h) = C$
- exponencial $\gamma(h) = \sigma^2(1 - \exp(-|h|/\rho)) \quad \sigma^2 > 0 \quad \rho > 0$
- esferico $\gamma(h) = \begin{cases} \sigma^2(\frac{3}{2}\frac{|h|}{\rho} - \frac{1}{2}\frac{|h|^3}{\rho^3}) & \text{si} \quad 0 \leq |h| \leq \rho \\ \sigma^2 & \text{si} \quad |h| \geq \rho \end{cases}$
- gaussiano $\gamma(h) = \sigma^2(1 - \exp(-|h|^2/\rho)) \quad \sigma^2 > 0 \quad \rho > 0$
- potencia $\gamma(h) = \sigma^2|h|^\alpha \quad \alpha < 2$
- ...

Variograma

Clase de Matern

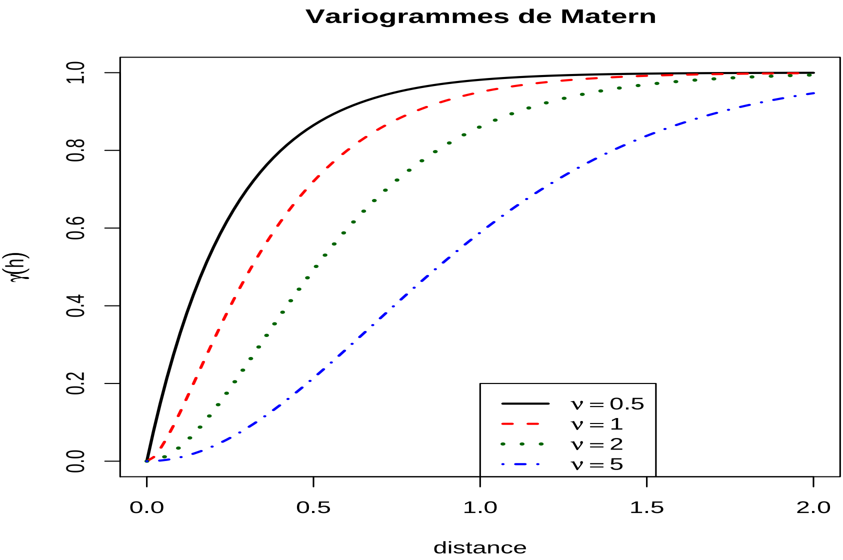
$$\gamma(h) = \sigma^2 \Big[1 - \frac{1}{2^{\nu-1}\Gamma(\nu)} \Big(\frac{2\nu^{1/2}h}{\rho} \Big) \mathcal{K}_\nu \Big(\frac{2\nu^{1/2}h}{\rho} \Big) \Big]$$

$\mathcal{K}_\nu$: funcion de Bessel modificada de tercera especie, de orden ν

ν : parametro que regula la regularidad en 0.

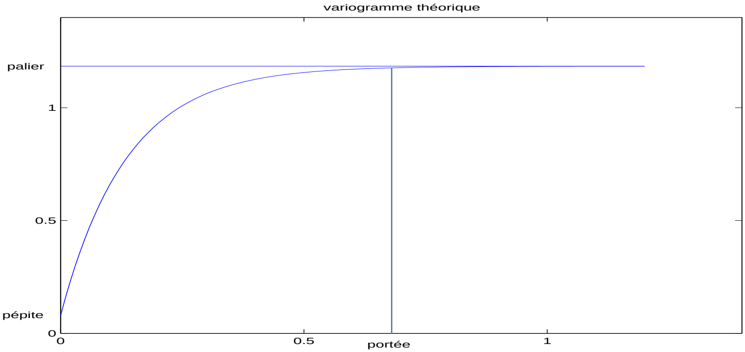
- $\nu = 1/2$: modelo exponencial
- $\nu \to +\infty$: modelo Gaussiano

Variograma de Matern



Efecto de los parametros

- silo (sill) $\sigma^2 = \lim_{h\to\infty} \gamma(h)$
- rango (range) a t.q. $\|h\| \geq a \to \sigma^2 - \gamma(h) \leq \varepsilon$
- pepita (nugget) $\tau^2 = \lim_{h\to 0} \gamma(h)$



Efecto de los parametros

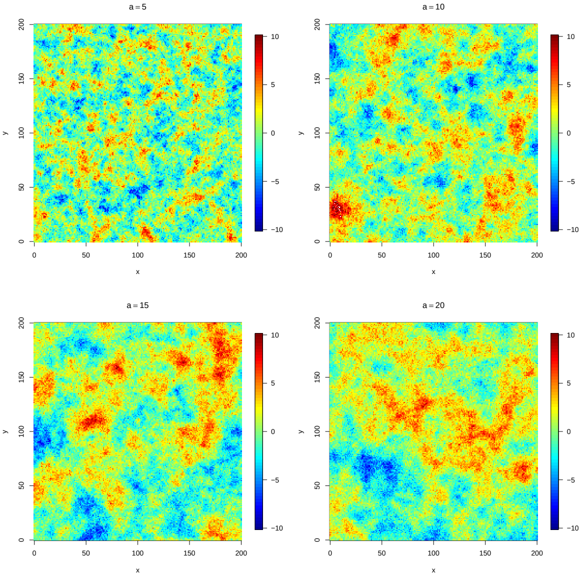
El silo
Si es finito, el proceso Z es estacionario de segundo orden y

$$\sigma^2 = \mathsf{Var}(Z(s))$$

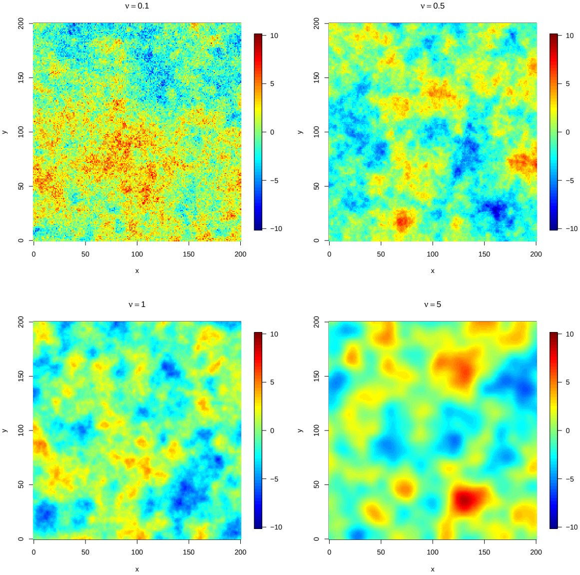
La escala (scale)
ρ es una medida de la fuerza de la correlacion, la distancia a partir de la cual ya no hay correlacion est el rango a. Esta relacionada con ρ.

La pepita
- si $\tau^2 \neq 0$ el proceso es muy irregular (ruido blanco, error de medida)
- si $\tau^2 = 0$, la regularidad del variograma en 0 (parametro ν en el variograma de Matern) indica la regularidad del proceso.

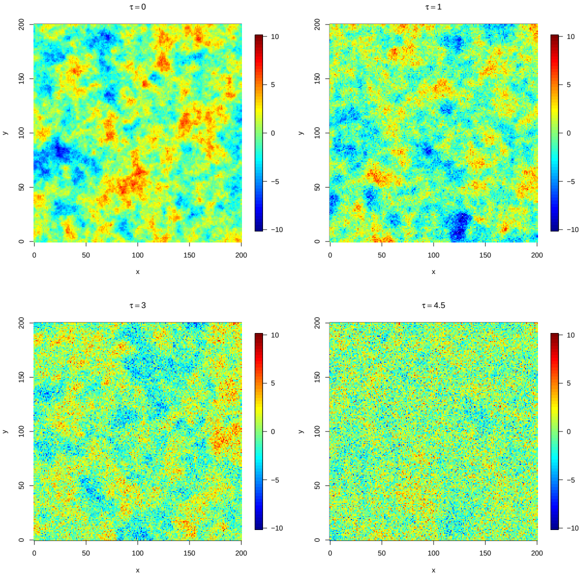
Efectos de los parametros : la escala



Efectos de los parametros : regularidad



Efectos de los parametros : pepita



Anisotropia

Variograma direccional (direccion $\overrightarrow{d}$)

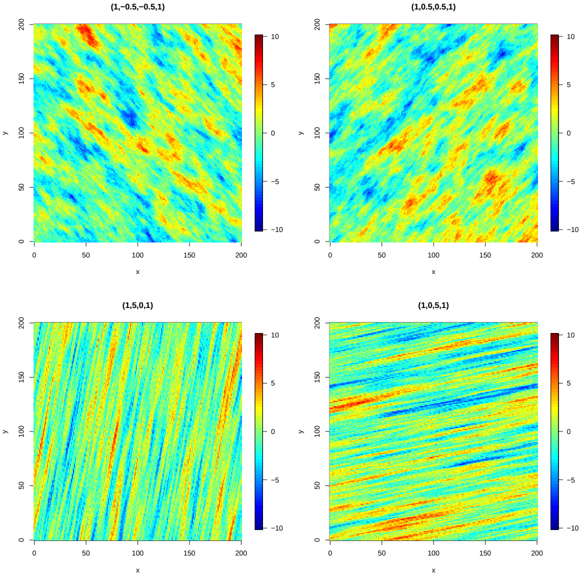
$$\gamma_{\overrightarrow{d}}(h) = \frac{1}{2}\mathsf{Var}(Z(s + h\overrightarrow{d}) - Z(s))$$

Si la funcion variograma depende de la direccion el proceso es anisotropo.

- **Anisotropia geometrica**
La escala varia segun una elipse.

- **Anisotropia zonal**
El silo depende de la direccion.

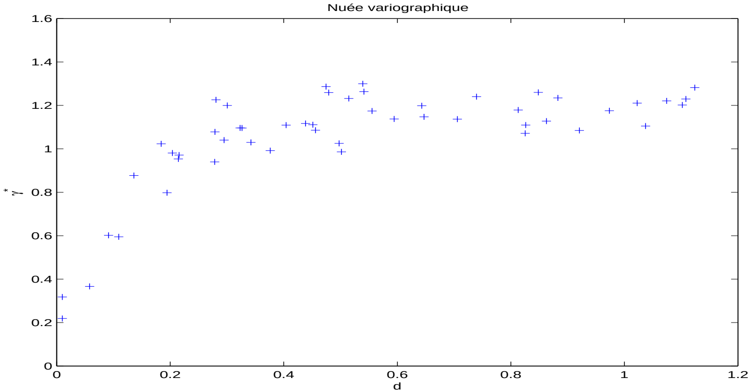
Efecto de los parametros



Estimacion del variograma

Nube variografica

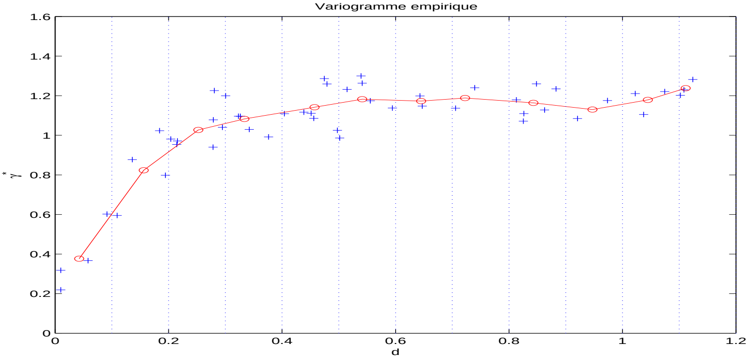
$$\gamma_{ij}^* = \frac{(z(s_i) - z(s_j))^2}{2} \qquad d_{ij} = \|s_i - s_j\|$$



Estimacion del variograma

Variograma experimental

$$\widetilde{\gamma}(d_k) = \frac{1}{2n_c} \sum_{i,j \in C(k)} (Z(s_i) - Z(s_j))^2 \quad i,j \in C(k) \Leftrightarrow (k-1)\delta \leq \|s_i - s_j\|$$

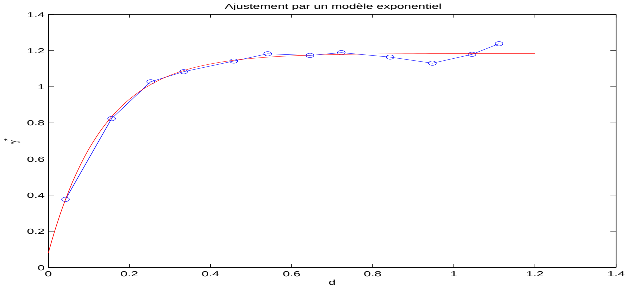


Estimacion del variograma

Ajuste

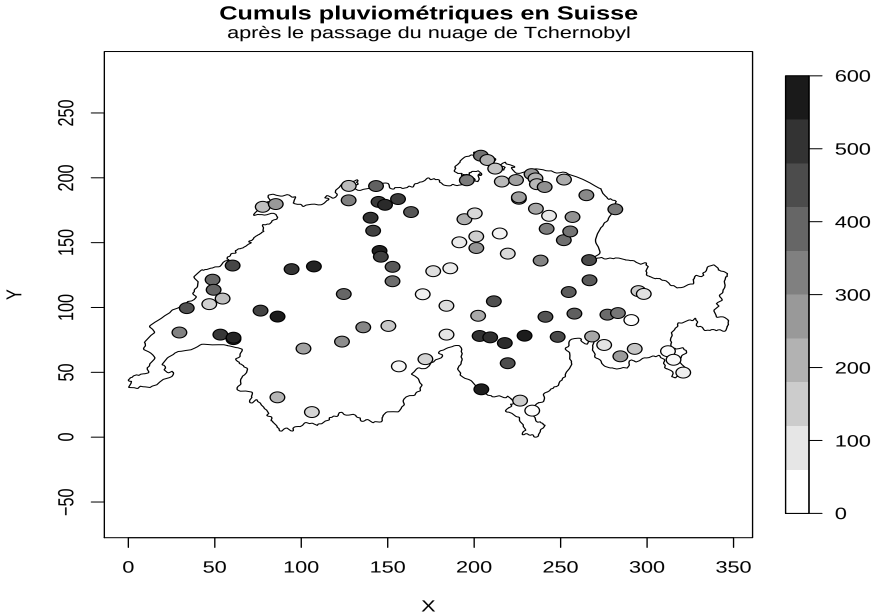
$\gamma_{\rho,\sigma^2,\nu,\tau^2}$ variograma admisible, $(\rho,\nu,\sigma^2,\tau^2)$ solucion del problema

$$\min_{\rho,\sigma^2,\nu,\tau^2} \sum_{k=1}^{K} (\gamma_{\rho,\sigma^2,\nu,\tau^2}(d_k) - \widetilde{\gamma}(d_k))^2$$

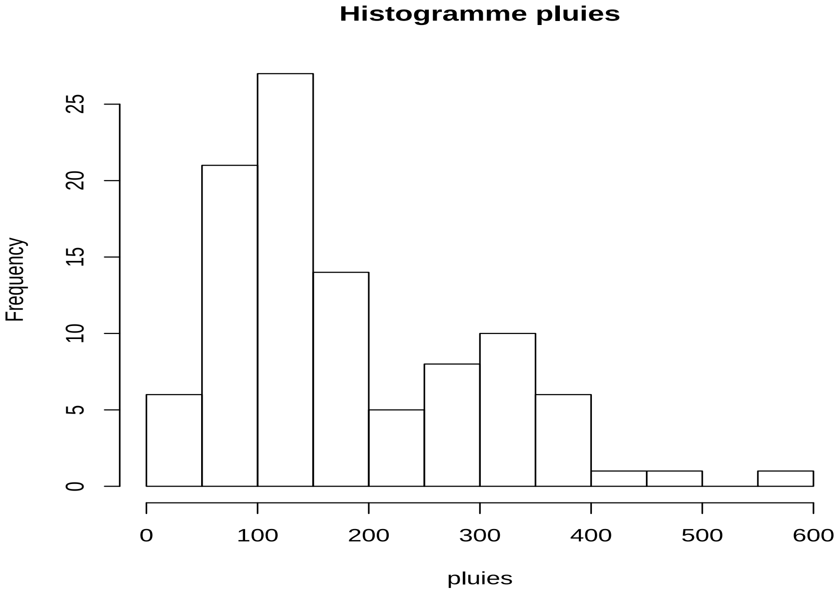


Otros metodos :
maxima verosimilitud,
minimos cuadrados generalizados ...

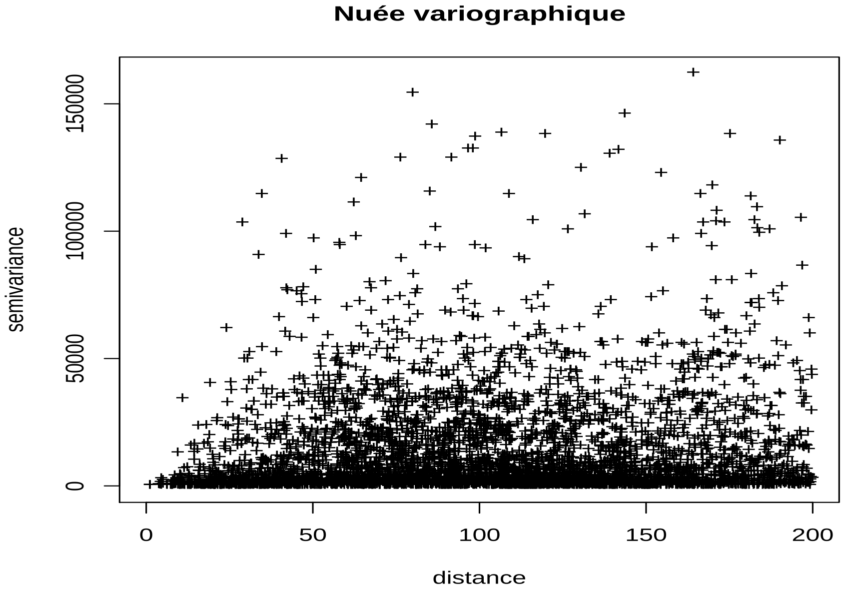
Lluvias suizas



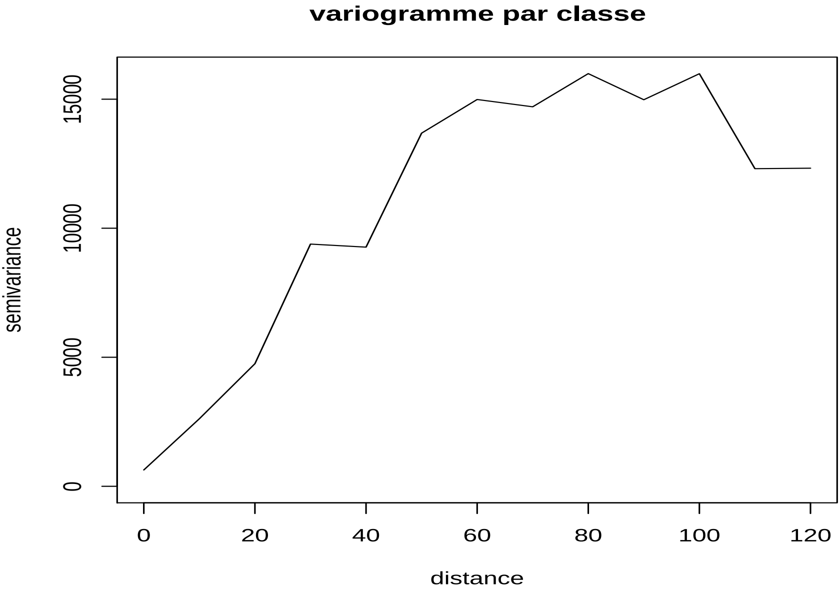
Lluvias suizas



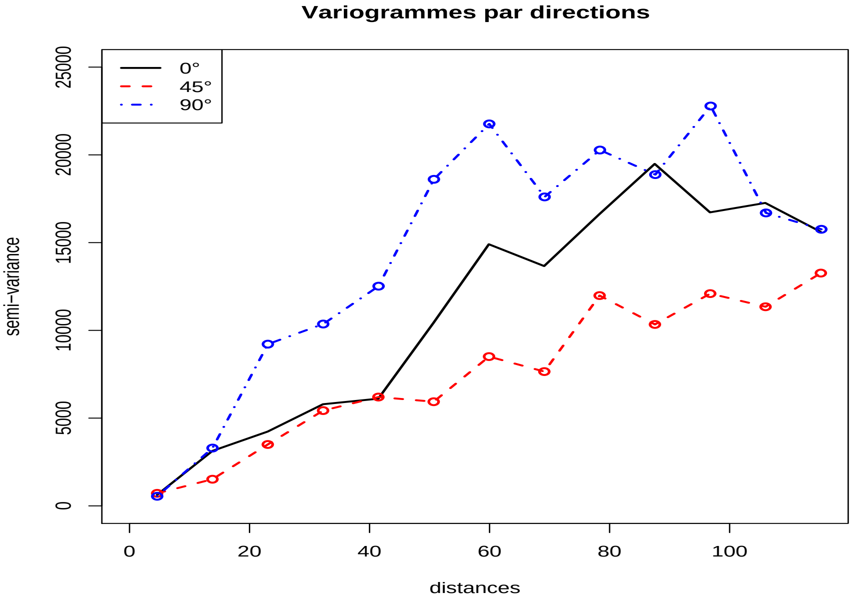
Lluvias suizas



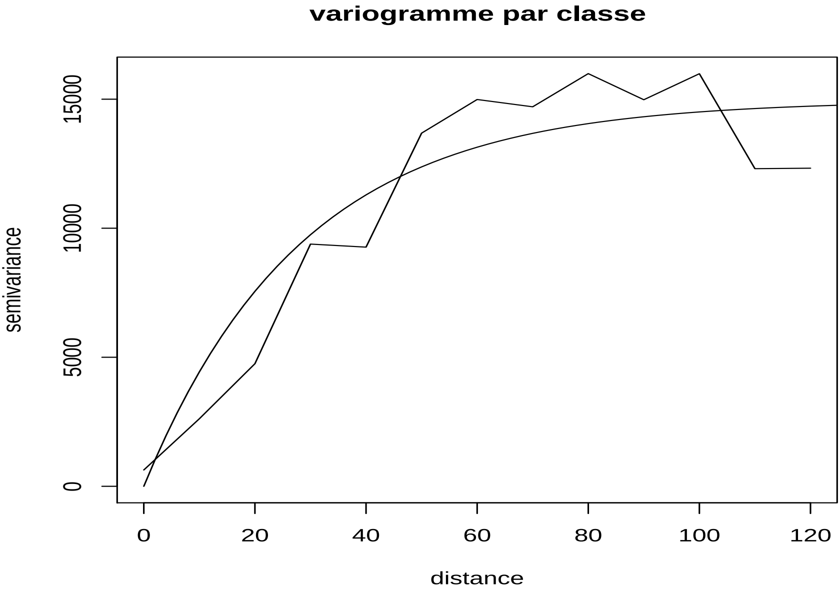
Lluvias suizas



Lluvias suizas



Lluvias suizas



Lluvias suizas