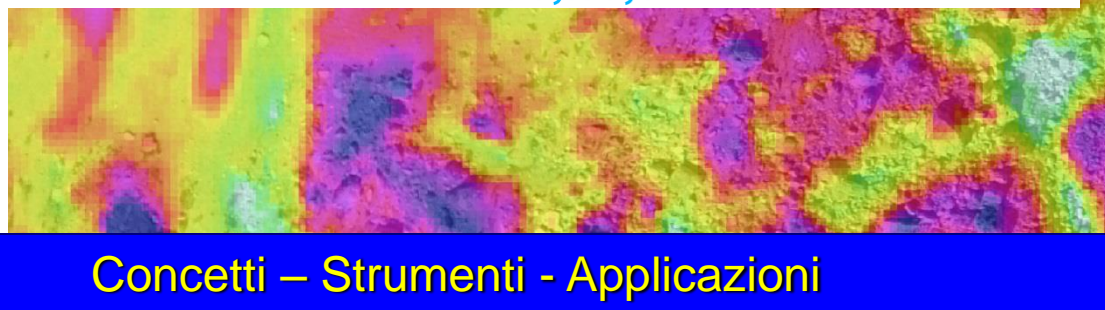
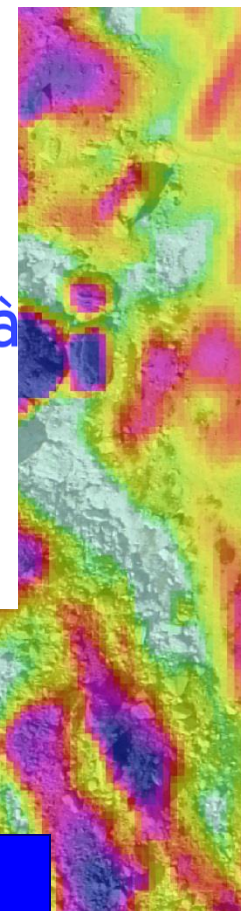


Mapping dei dati geo-ambientali: una breve introduzione alla geostatistica



Concetti – Strumenti - Applicazioni

La geostatistica è una branca della statistica spaziale (spaziotemporale)

La statistica spaziale comprende diverse tipologie di analisi, le più comuni sono:

- Point pattern analysis
- Studio della distribuzione geografica di entità vettoriali (ad esempio analisi distribuzione tridimensionale dei piani di faglia, dati direzionali, ecc..)
- Analisi di rete (ad esempio flussi di traffico in una rete viaria)
- Analisi dati spazialmente esaustivi di natura discreta (ad es. Fragstat, analisi ecologiche)
- Analisi dati spazialmente esaustivi di natura continua (ad es. Image texture)
- Geostatistica
- Regressione geografica

Molte di queste tecniche sono implementate o implementabili nei GIS. Le diverse tecniche a volte vengono utilizzate in maniera integrata. Sovente approcci mediante Machine Learning

In ogni caso, molti dei concetti geostatistici sono funzionali nell'ambito delle altre tipologie di analisi spaziale (e non solo...vedi serie temporali...)

La geostatistica non va intesa come statistica di dati di tipo geologico. Essa è una branca della statistica spaziale e si interessa di dati spaziali (spazio-temporali)

“Il tuo studio della geostatistica non scalzerà le tue conoscenze specifiche; piuttosto le renderà più fruibili”. *Tratto e adattato da D.E. Myers (il quale lo ha parafrasato da W.E. Deming)*

Inizialmente è stata sviluppata per esigenze pratiche, lontano dall'ambito puramente statistico

Anni 50/60

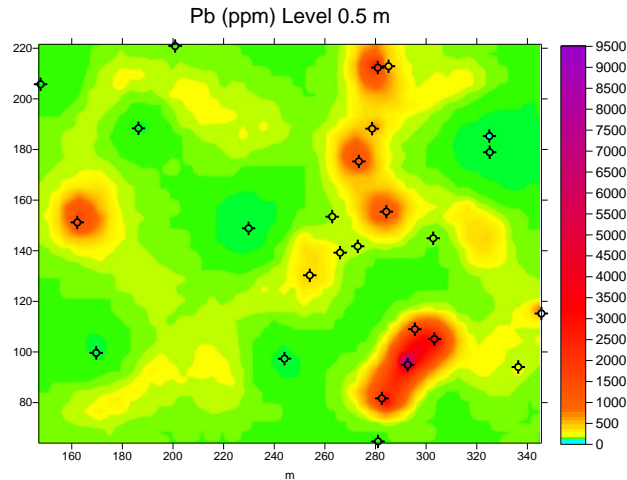
G. Matheron (Ecole Normale Supérieure des Mines de Paris)

Y. Ghandin (analisi oggettiva)

Applicazioni: giacimenti minerari, idrogeologia, meteorologia, ambiente, risorse naturali, scienza dei suoli, geologia petrolifera, ecc.

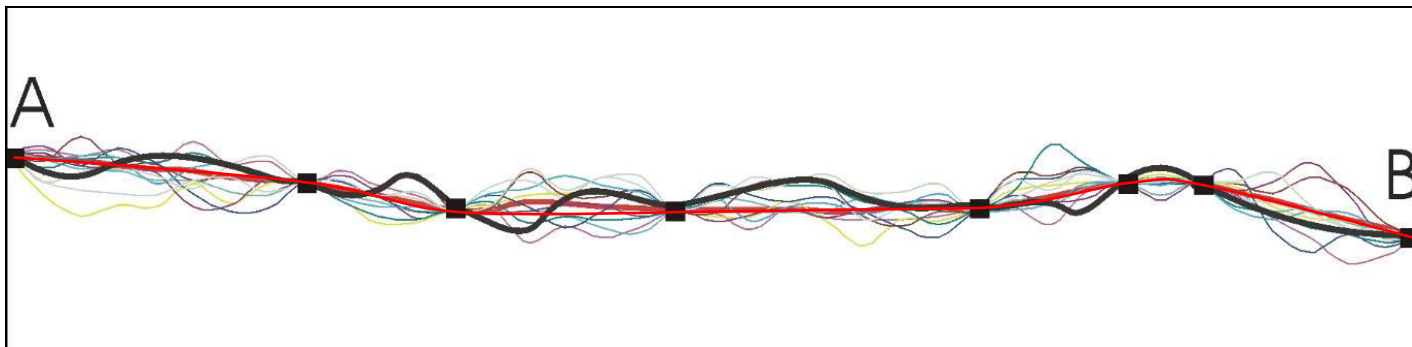
Il problema:

Ricostruire entro un dato dominio dimensionale la distribuzione spaziale (spazio-temporale) completa di una determinata proprietà fisico-chimica (o attributo)



Compito difficile:

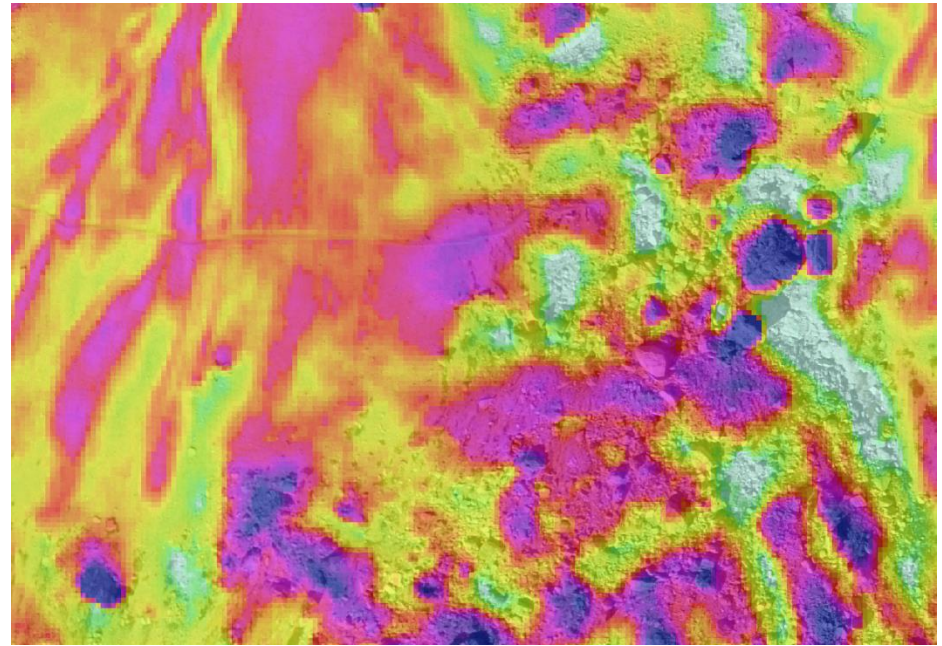
- 1) Eterogeneità caratterizzante la distribuzione spaziale
- 2) Informazione disponibile frammentaria ed eterogenea



Incertezza spaziale:

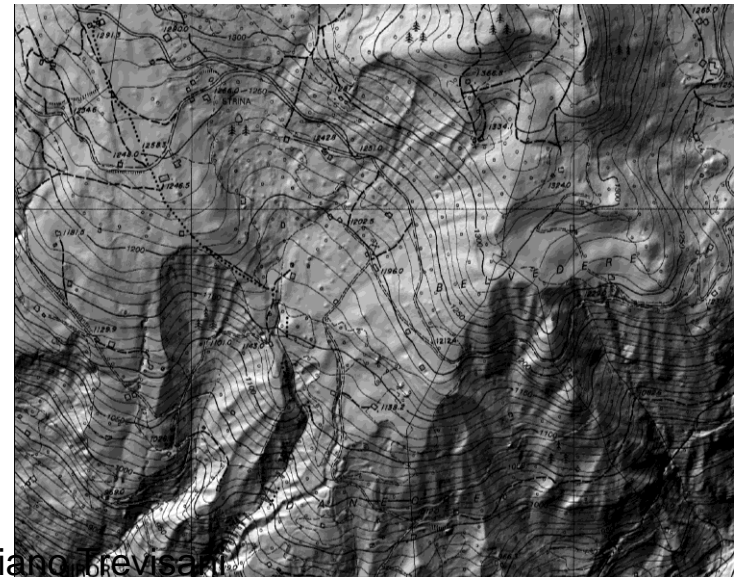
Più scenari sono compatibili con le informazioni disponibili

Molto spesso, almeno entro un dato dominio dimensionale, la distribuzione spaziale della proprietà studiata è caratterizzata da una qualche struttura



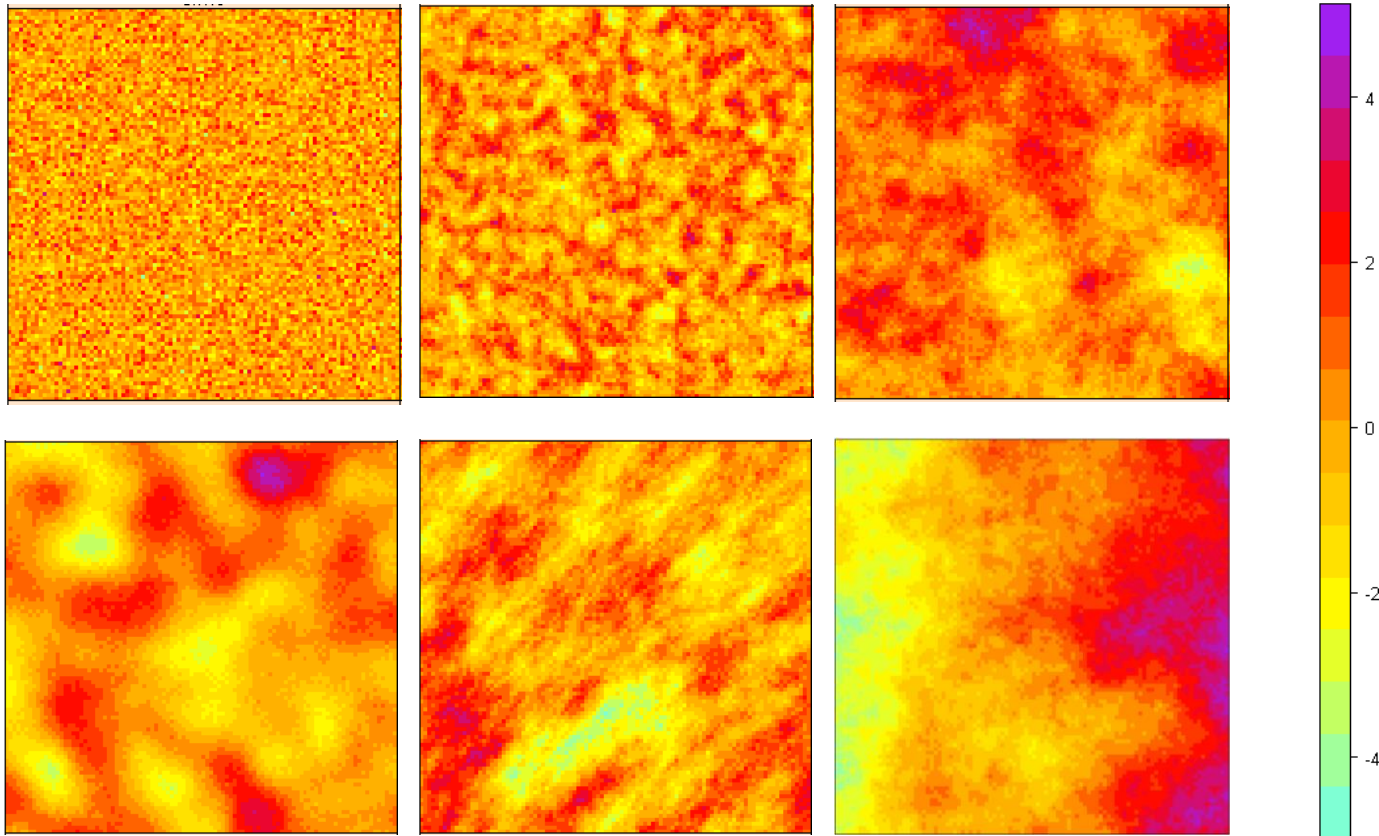
In particolare si osserva la presenza di continuità spaziale.

L'idea è quella di utilizzare tale struttura per "estendere" l'informazione spazialmente

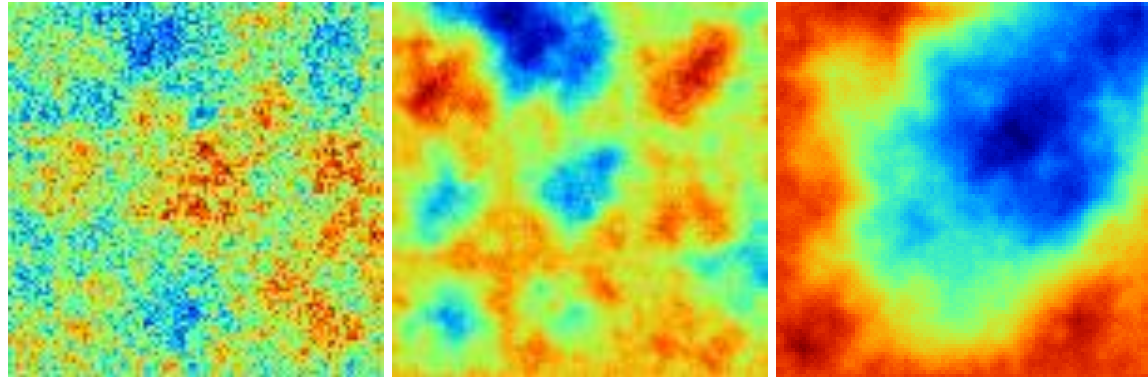




Continuità spaziale



Concettualizzando la distribuzione spaziale della proprietà studiata come risultato di un processo stocastico possiamo utilizzare ciò che vi è di strutturato nei dati per quantificare e gestire l'incertezza spaziale



Il parametro del modello stocastico a cui siamo maggiormente interessati è la continuità spaziale

La continuità spaziale è quantificabile (variogramma, covariogramma, ecc.)

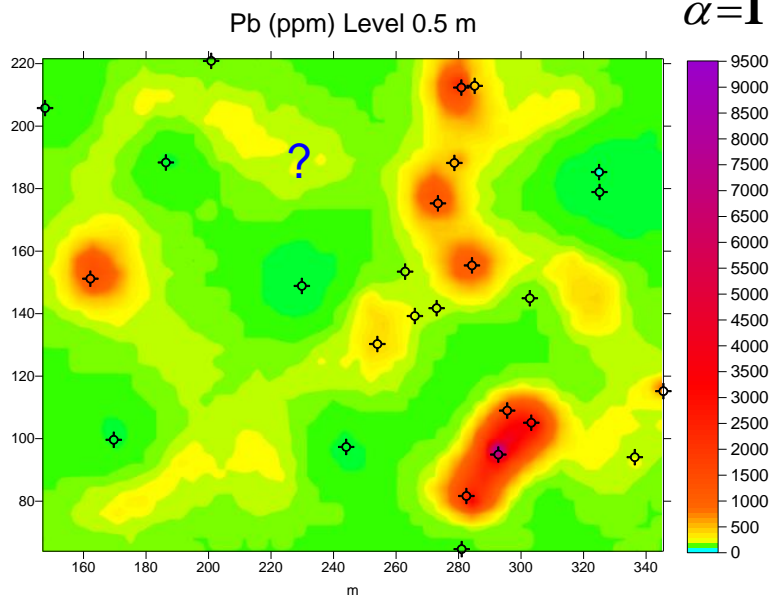
Lo studio geostatistico può fornirci importanti informazioni circa i processi

Le conoscenze esperte possono aiutare il processo di inferenza

In realtà gran parte degli interpolatori classici presuppongono una forma di continuità spaziale: fissandola a priori

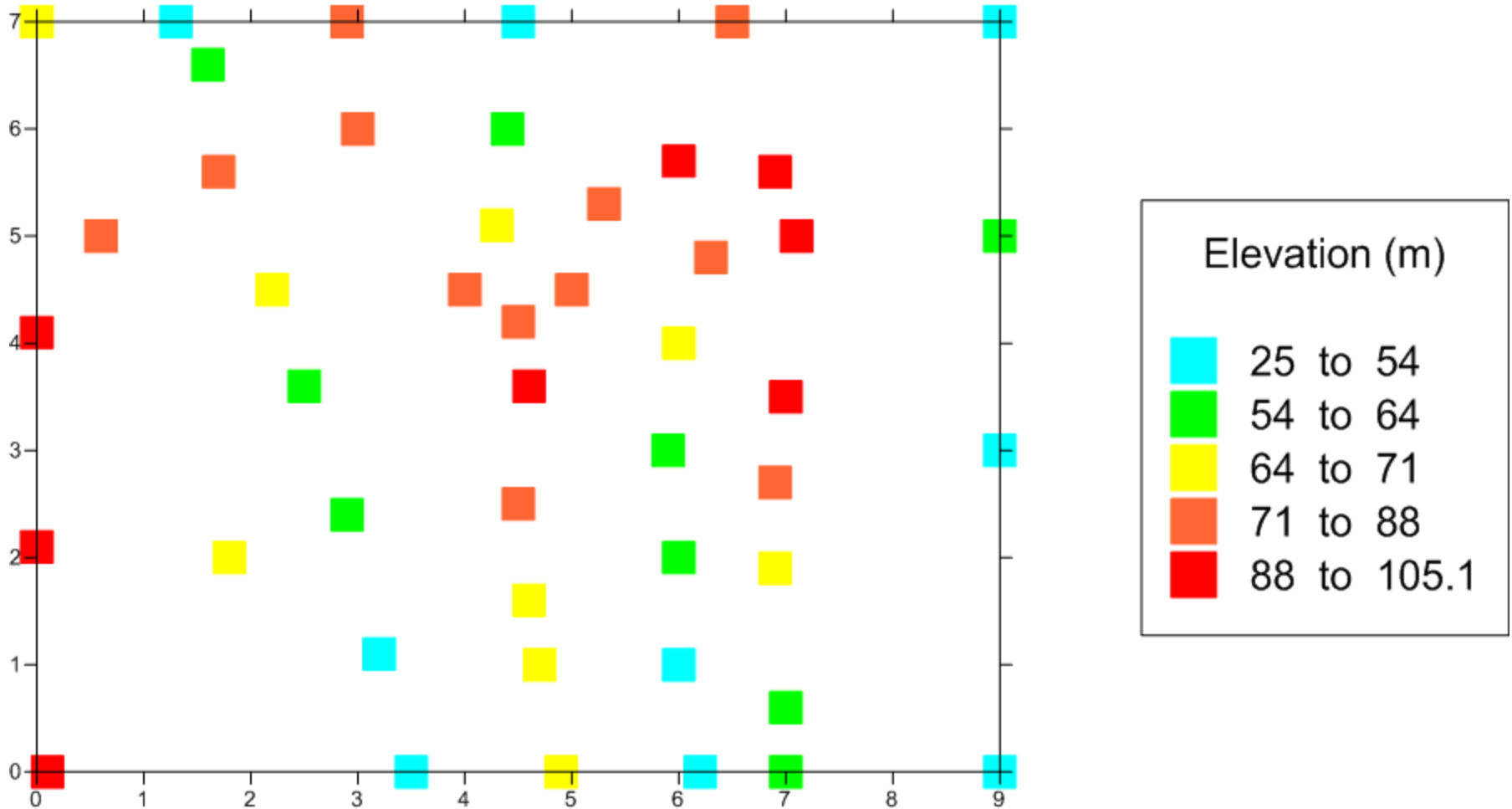
Alcuni tengono conto solo della distanza dei dati noti dal punto di stima (triangolazione e metodi dell'inverso della distanza) e altri tengono conto anche del livello di clustering (ad esempio il Natural Neighbor).

$$? = \sum_{\alpha=1}^n \lambda_{\alpha} z(\mathbf{u}_{\alpha})$$

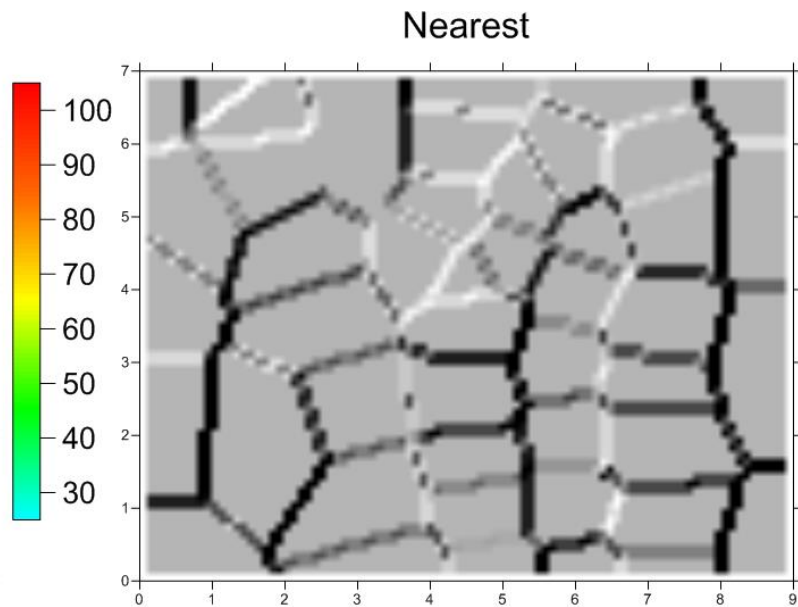
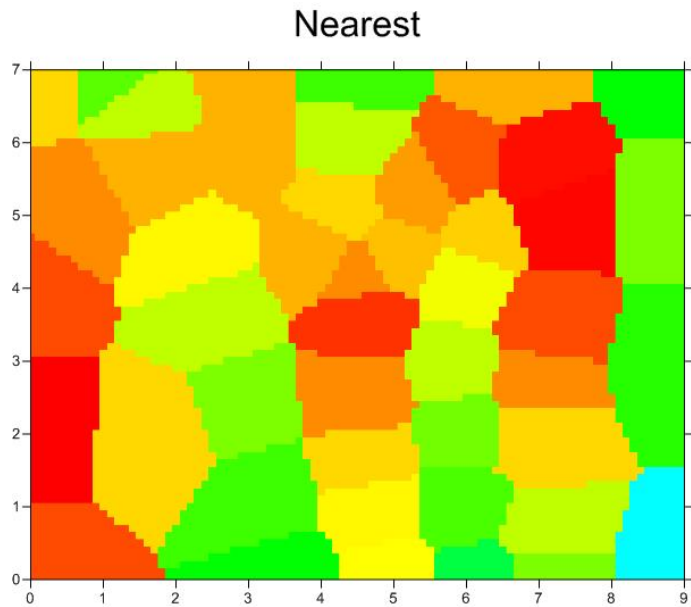
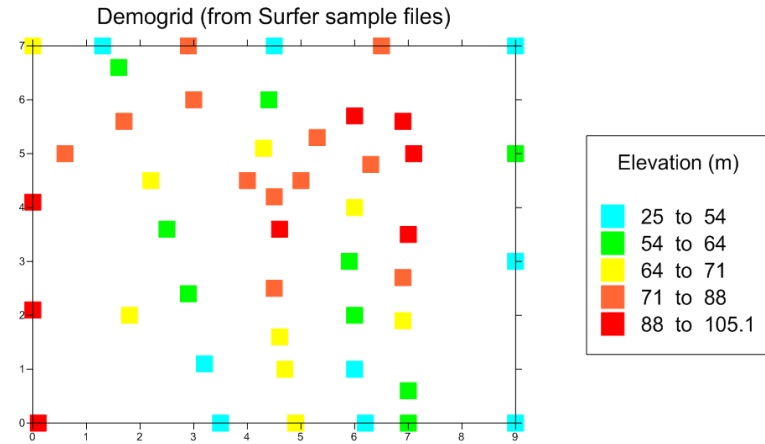


Con i metodi geostatistici si tiene conto della distanza dei dati noti dal punto di stima e del clustering secondo un criterio di distanza statistico derivato per inferenza dai dati.

Demogrid (from Surfer sample files)

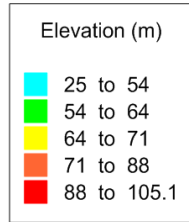
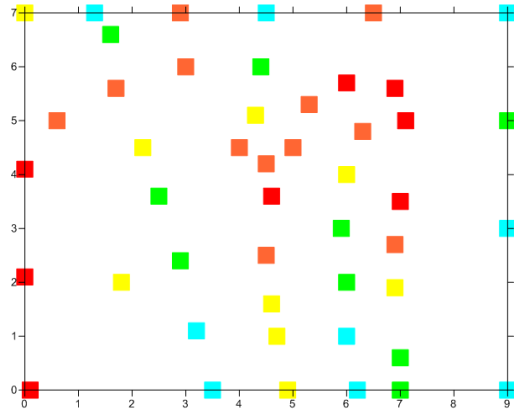


La problematica di base: interpolazioni

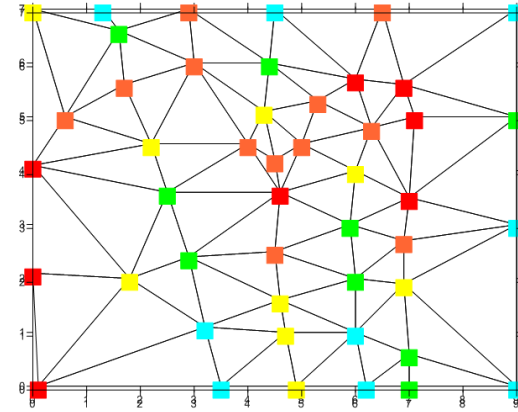


La problematica di base: interpolazioni

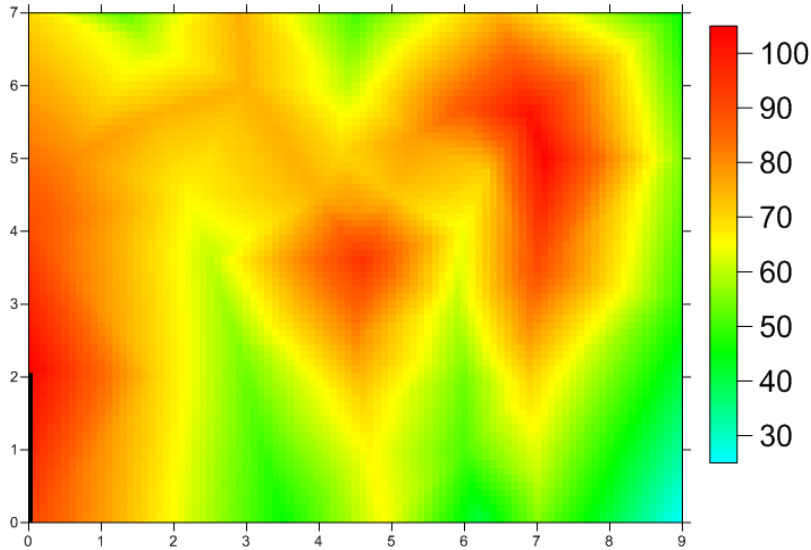
Demogrid (from Surfer sample files)



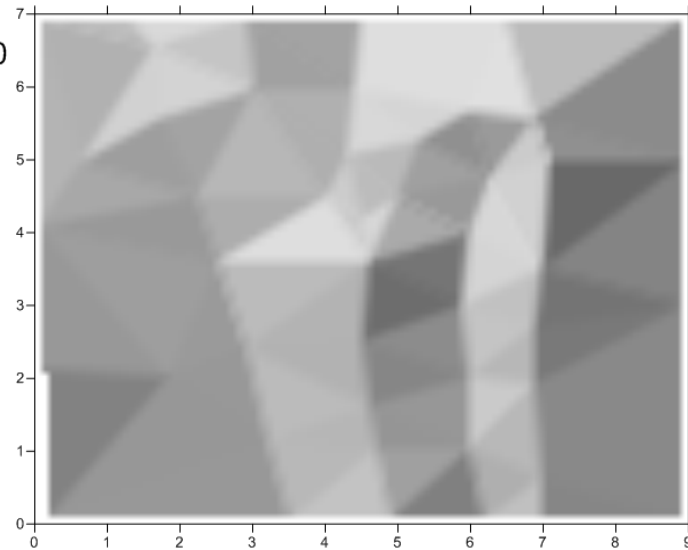
Demogrid (from Surfer sample files)



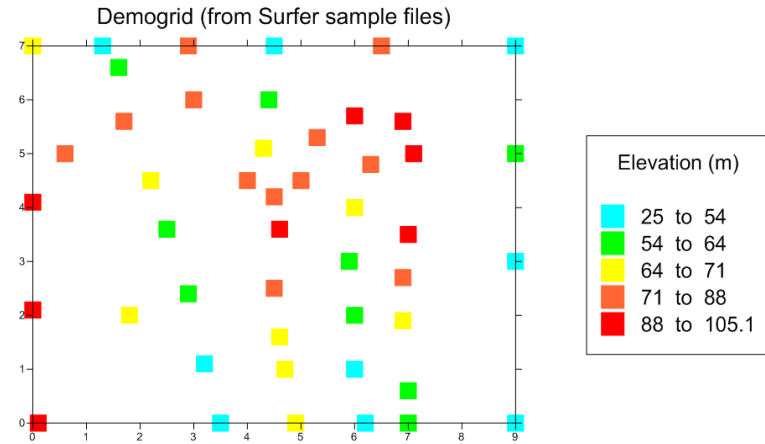
Triangolazione



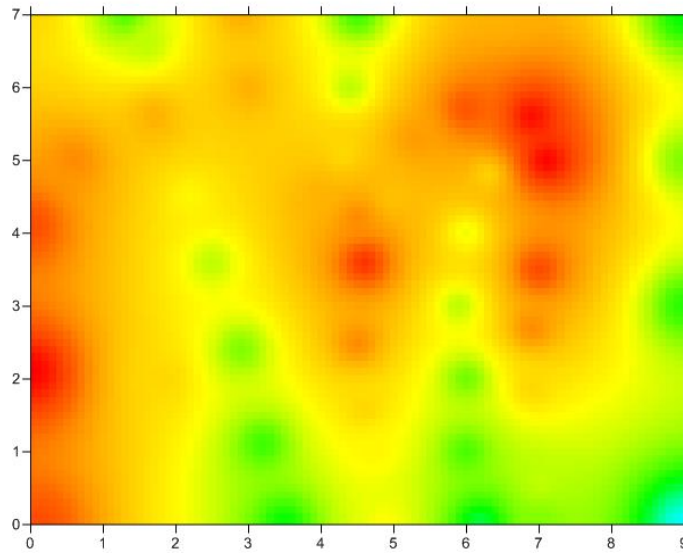
Triangolazione



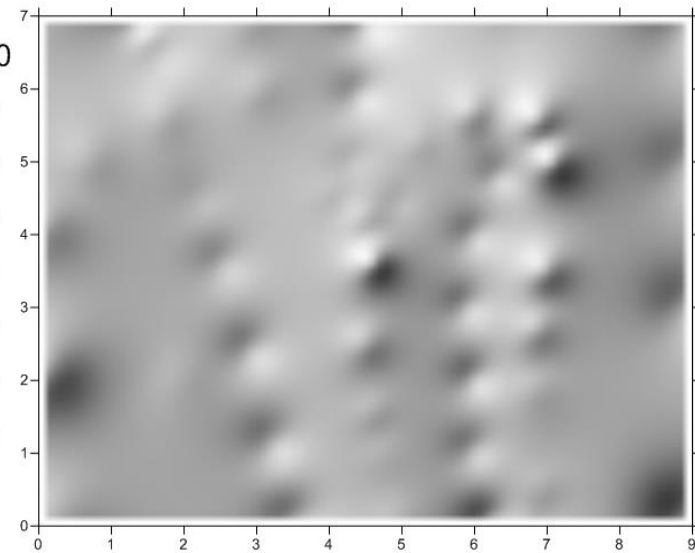
La problematica di base: interpolazioni



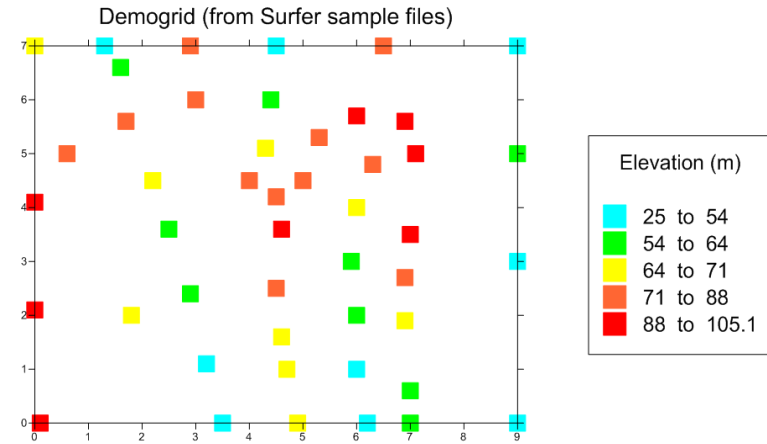
IDW (Esp=2)



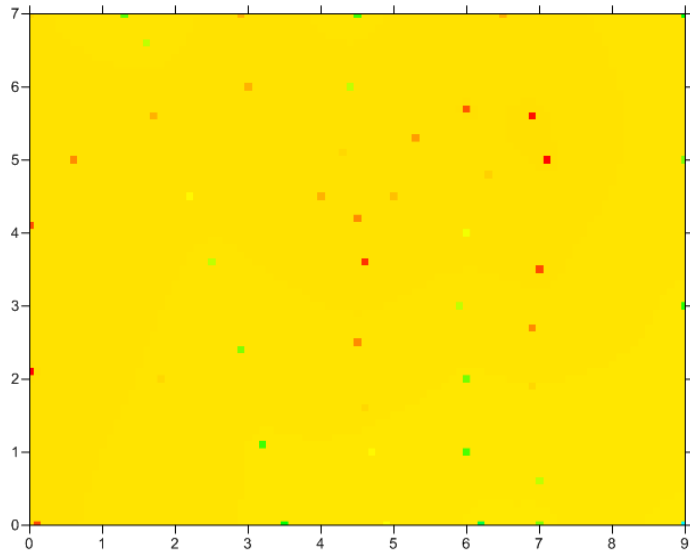
IDW (Esp=2)



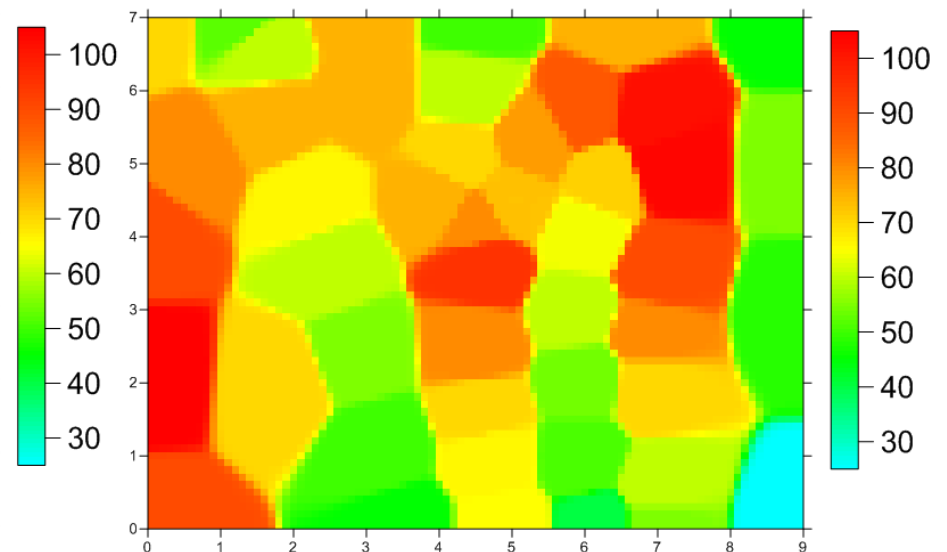
La problematica di base: interpolazioni



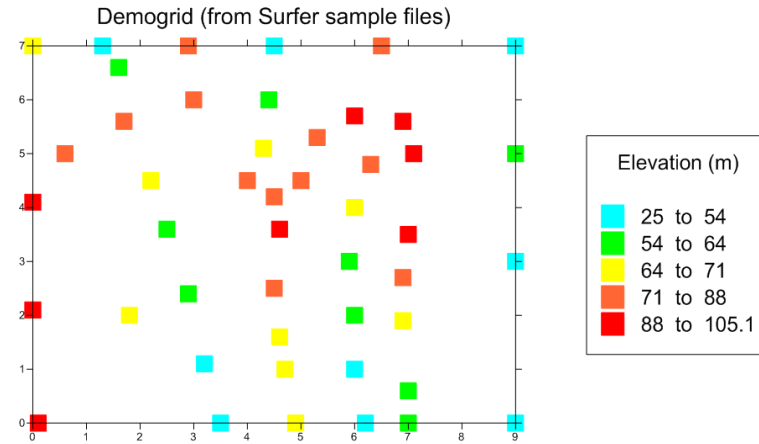
IDW (Esp=0.1)



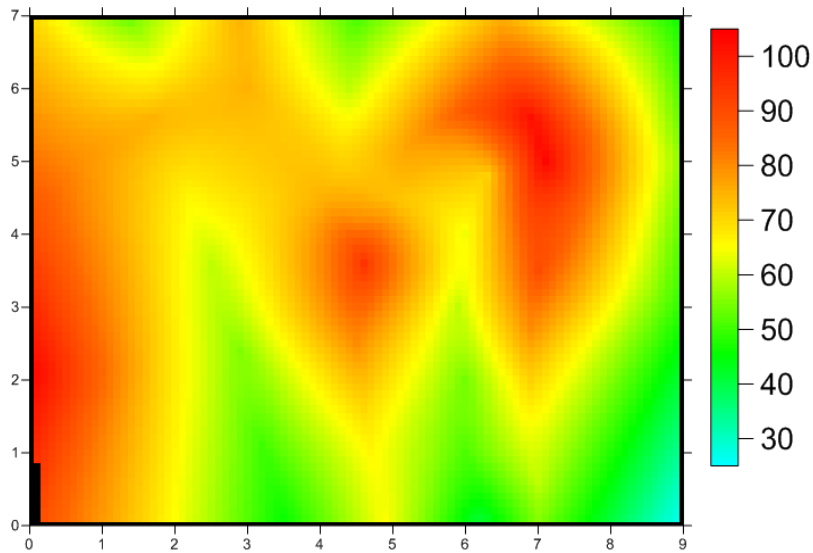
IDW (Esp=20)



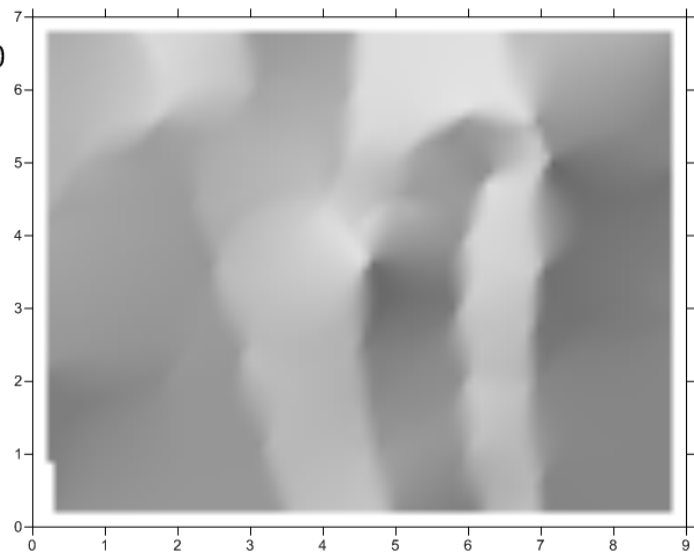
La problematica di base: interpolazioni



Natural

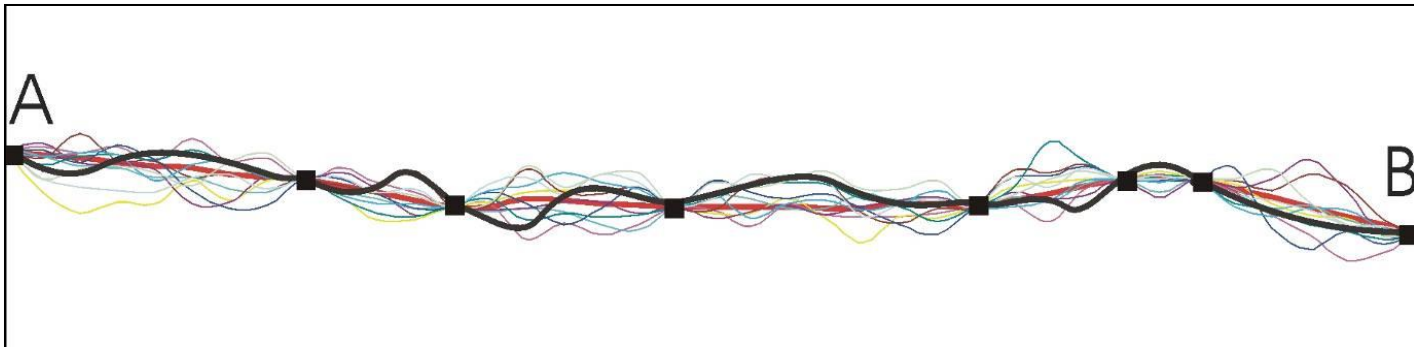


Natural



Vi è l'esigenza di uno strumento matematico che sia in grado di cogliere la struttura sfumata che generalmente caratterizza la distribuzione spaziale di una certa proprietà fisico-chimica.

Astrazione: la distribuzione spaziale di una data proprietà nello spazio è generata da un processo di tipo stocastico, matematicamente individuato da una o più funzioni aleatorie



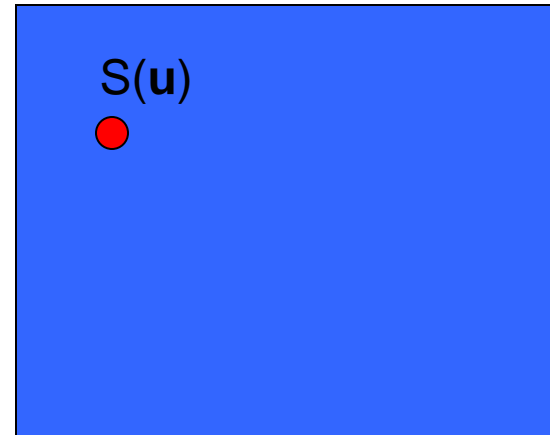
Generalmente, non abbiamo abbastanza informazioni o l'esigenza di definire la legge completa della funzione aleatoria. Ne possiamo definire solo alcuni parametri o momenti.

Una variabile discreta $S(\mathbf{u}), \mathbf{u} \in A$

in cui \mathbf{u} indica un determinato punto del dominio spaziale considerato A , può assumere K stati mutuamente esclusivi, ordinati o meno, e ad ognuno di questi stati è associata una probabilità:

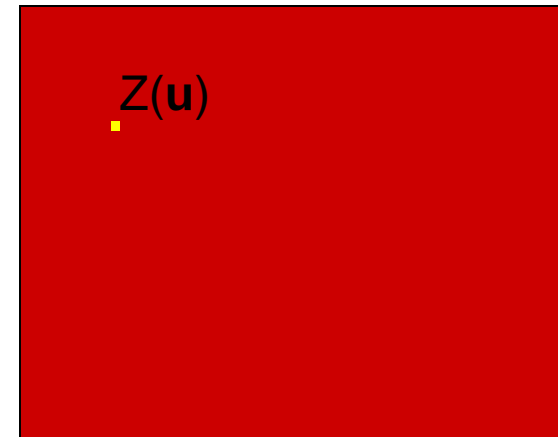
$$p(\mathbf{u}; s_k) = \text{Prob} \{S(\mathbf{u}) = s_k\}$$

$$p(\mathbf{u}; s_k) \in [0,1] \quad k=1, \dots, K$$



$$\sum_{k=1}^K p(\mathbf{u}; s_k) = 1$$

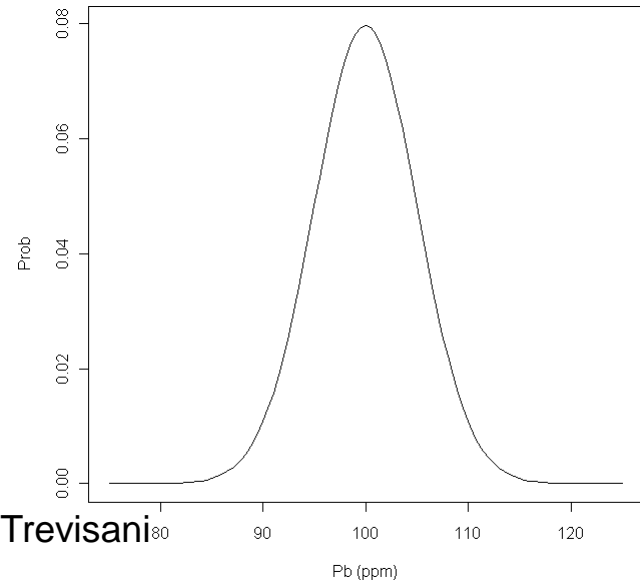
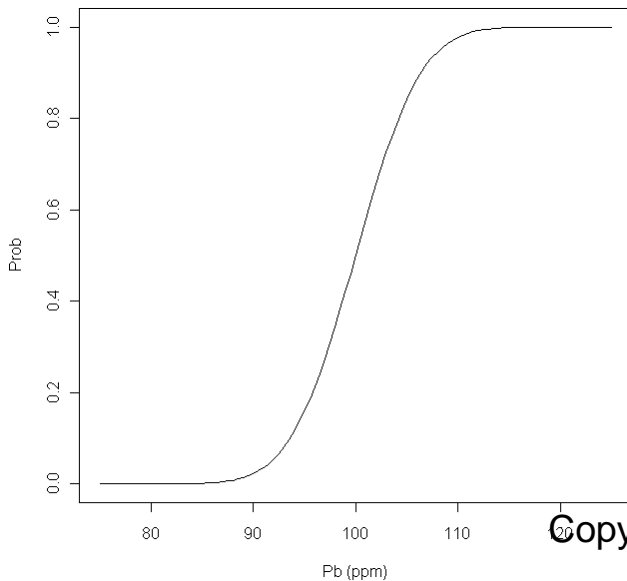
Una variabile aleatoria continua $Z(\mathbf{u}), \mathbf{u} \in \mathbf{A}$ è completamente definita dalla funzione di distribuzione cumulativa o di ripartizione (**c.d.f.** cumulative distribution function) $F(\mathbf{u}; z)$ o dalla funzione di densità di probabilità (**p.d.f.**: probability density function) $f(\mathbf{u}; z)$.



$$F(\mathbf{u}; z) = \text{Prob}\{Z(\mathbf{u}) \leq z\} \quad \forall z$$

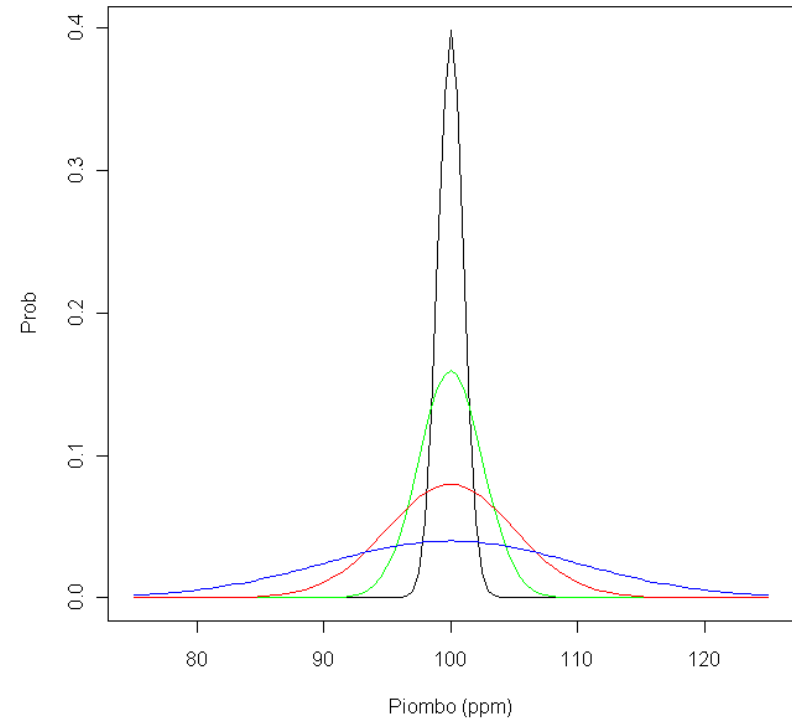
$$F(\mathbf{u}; z) \in [0,1] \quad \forall z$$

$$F(\mathbf{u}, z) = \int_{-\infty}^z f(\mathbf{u}, z) dz$$



Un esempio di variabile aleatoria continua può essere rappresentato da una gaussiana che definisce la funzione di densità di probabilità di una determinata proprietà (es. la concentrazione di piombo nel suolo). Quello della gaussiana è un caso alquanto fortunato....

La gaussiana è definita dalla media e dalla deviazione standard (nel disegno media 100 ppm e deviazioni standard 1, 2.5, 5, 10 ppm).

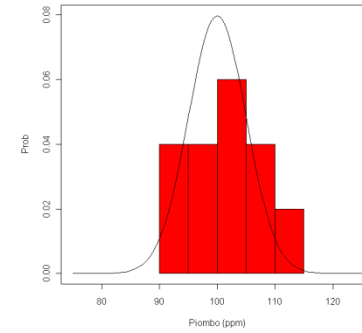
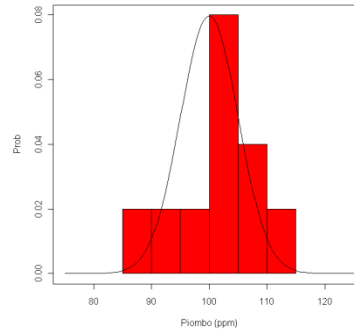
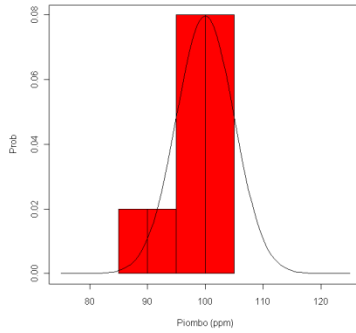


In accordo alla legge probabilistica si possono “pescare” casualmente dei valori (realizzazioni). Se prendo un grande numero di realizzazioni l'istogramma sperimentale approssima la gaussiana (e se calcolo la media e varianza queste approssimano i valori teorici).

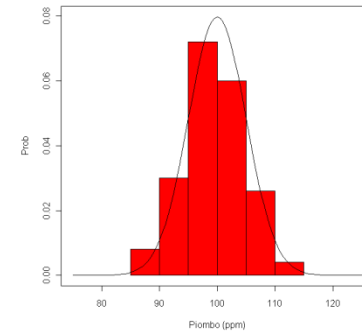
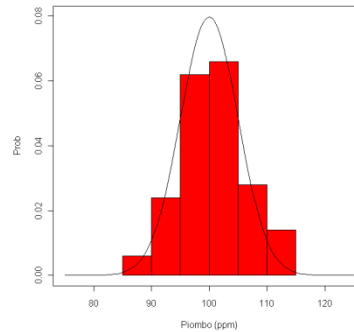
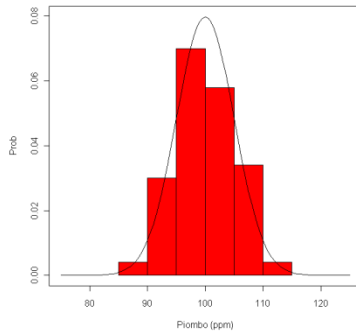
Variabili aleatorie continue: esempi

Riprendiamo la gaussiana con media 100 ppm e std 5 ppm proviamo a fare un po' di tentativi di campionamento....

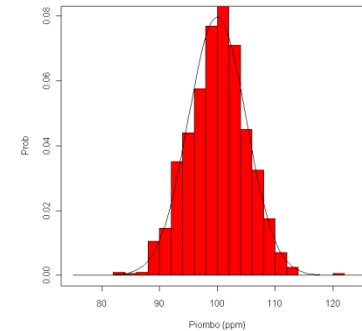
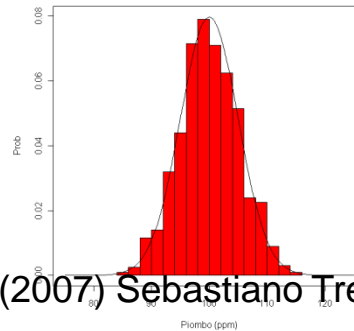
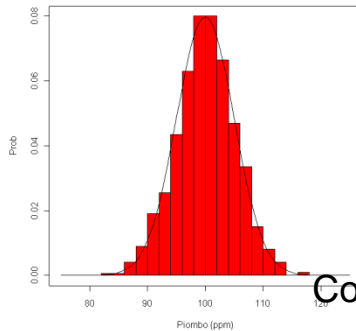
10 alla volta:



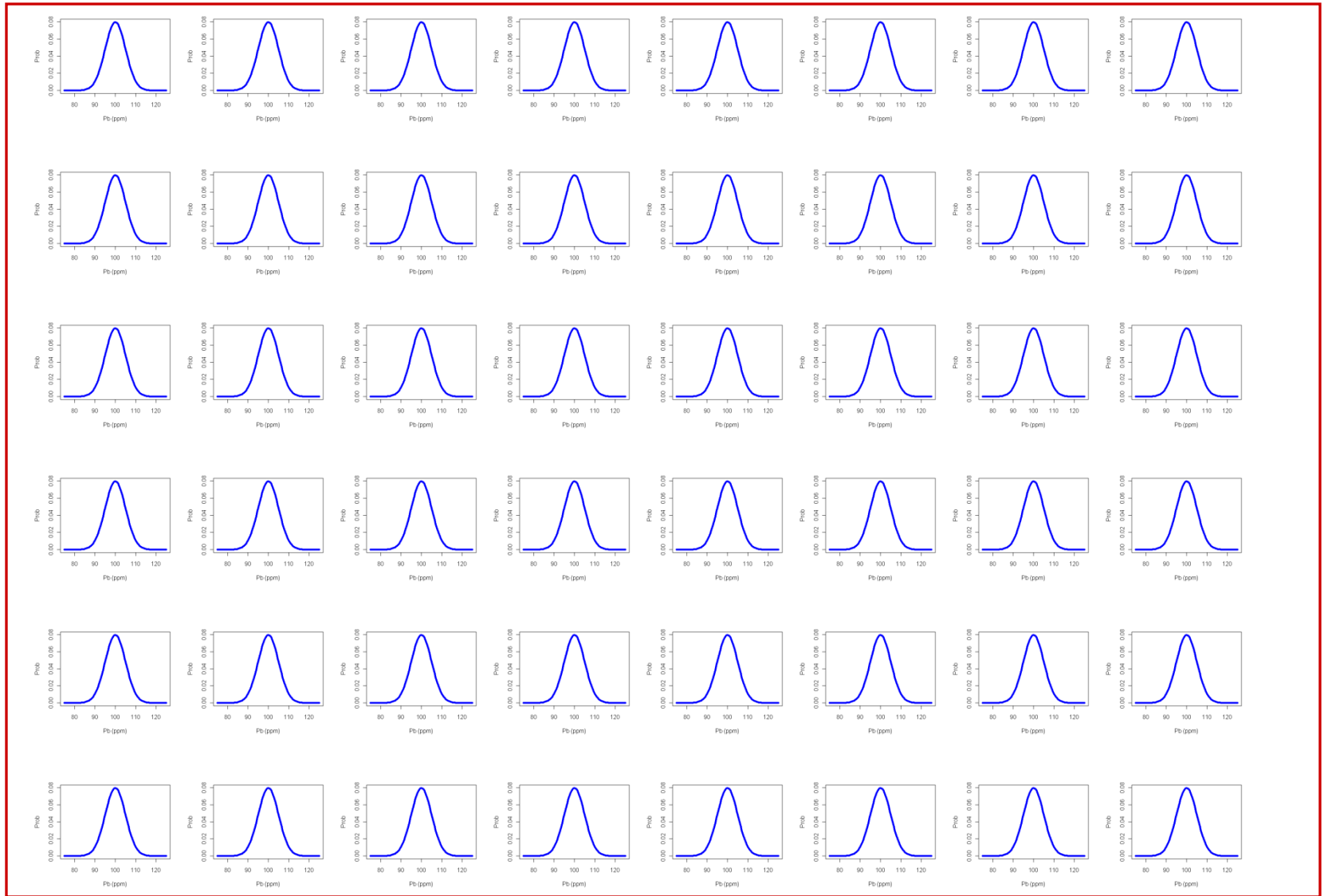
100 alla volta:



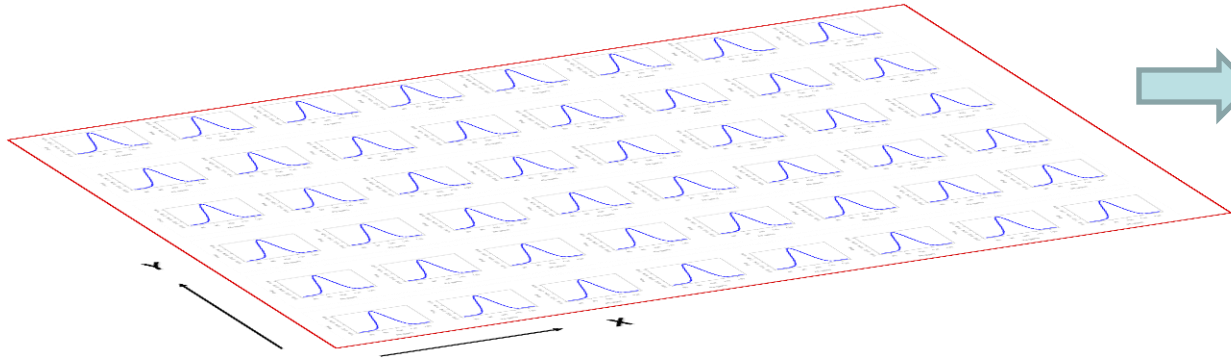
1000 alla volta:



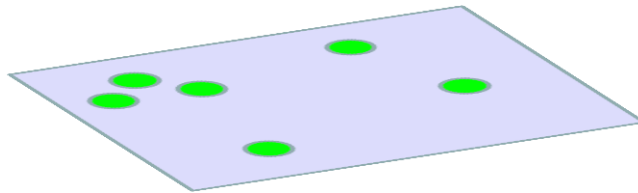
La funzione aleatoria definisce in ogni punto dello spazio una variabile aleatoria



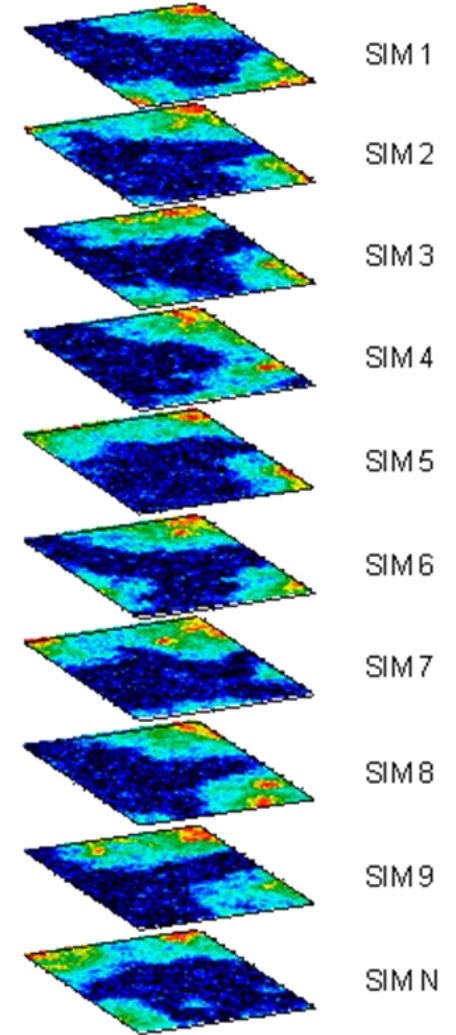
Modello



I dati



Scenari



La legge spaziale della funzione aleatoria

$$F(\mathbf{u}_1, \dots, \mathbf{u}_N; \mathbf{z}_1, \dots, \mathbf{z}_N) = \text{Prob} \{Z(\mathbf{u}_1) \leq z_1, \dots, Z(\mathbf{u}_N) \leq z_N\} \quad \forall \mathbf{u}_1, \dots, \mathbf{u}_n, z_1, \dots, z_n$$

Il momento del primo ordine considerato è il valore atteso (o speranza matematica). Il valore atteso di una variabile aleatoria $Z(\mathbf{u})$ è definito come

A noi interessano solo alcuni parametri della legge spaziale

$$m(\mathbf{u}) = E\{Z(\mathbf{u})\} \quad \text{con} \quad E\{Z(\mathbf{u})\} < \infty$$

$$m(\mathbf{u}) = \int_a^b z \cdot f(\mathbf{u}; z) dz$$

Momenti del secondo ordine:

Varianza

$$\sigma^2(\mathbf{u}) = \text{Var}\{Z(\mathbf{u})\} = E\{[Z(\mathbf{u}) - E\{Z(\mathbf{u})\}]^2\} = E\{Z(\mathbf{u})^2\} - E\{Z(\mathbf{u})\}^2$$

posto che $E\{Z(\mathbf{u})^2\} < \infty$

Per quel che riguarda la c.d.f. congiunta bivariata i momenti del secondo ordine considerati sono

$$\text{Covarianza} = C(\mathbf{u}, \mathbf{u}') = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u}')\} - E\{Z(\mathbf{u})\} \cdot E\{Z(\mathbf{u}')\}$$

$$\text{Correlogramma} = \rho(\mathbf{u}, \mathbf{u}') = \frac{C(\mathbf{u}, \mathbf{u}')}{\sqrt{C(\mathbf{u}, \mathbf{u}) \cdot C(\mathbf{u}', \mathbf{u}')}}}$$

$$\text{Variogramma} = 2 \gamma(\mathbf{u}, \mathbf{u}') = \text{Var}\{Z(\mathbf{u}) - Z(\mathbf{u}')\}$$



Ma se abbiamo una realizzazione sola come possiamo calcolare i "momenti"?



Stazionarietà

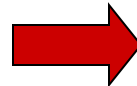


Stazionarietà in senso stretto

E' l'ipotesi più stringente in cui si ammette che l'intera legge spaziale sia indipendente dalla posizione considerata.

Stazionarietà di secondo ordine
Con questa ipotesi si ammette che i momenti del primo e del secondo ordine esistano e siano indipendenti dalla posizione considerata. Più precisamente si dice che una funzione aleatoria $Z(\mathbf{u})$ è stazionaria del secondo ordine se:

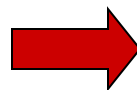
$$E\{Z(\mathbf{u})\} = m \quad \forall \mathbf{u} \in A$$



$$C(\mathbf{h}) = E\{Z(\mathbf{u}) \cdot Z(\mathbf{u} + \mathbf{h})\} - m^2 \quad \forall \mathbf{u} \in A$$

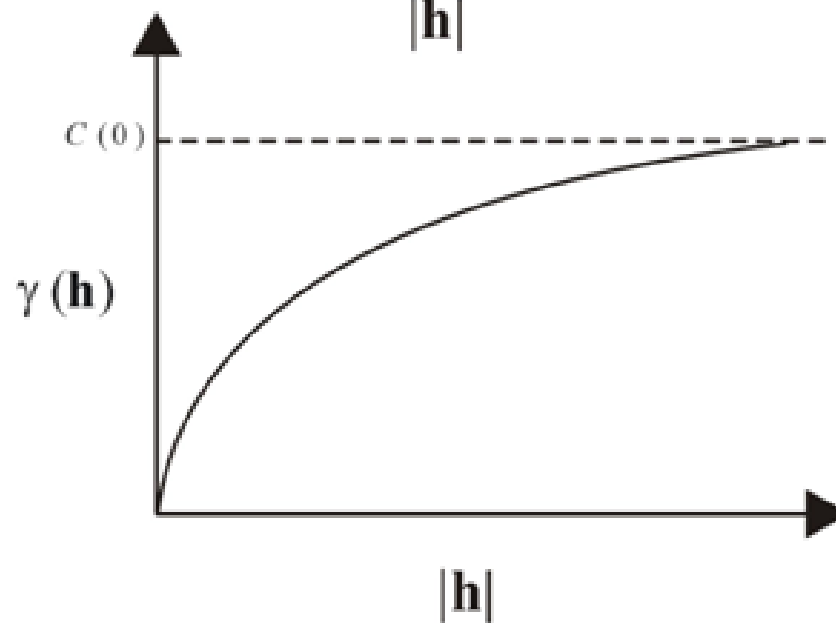
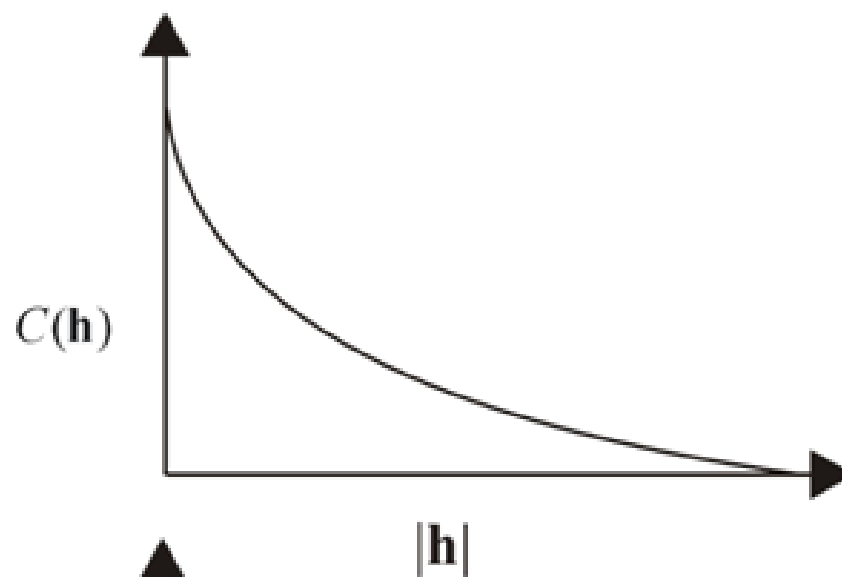
$$C(\mathbf{h}) = C(0) - \gamma(\mathbf{h})$$

Stazionarietà intrinseca



$$E\{Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})\} = 0$$

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



A

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h})$$

Funzione covarianza (covariogramma)

$$\hat{C}(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{u}_{\alpha}) \cdot z(\mathbf{u}_{\alpha} + \mathbf{h}) - \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{u}_{\alpha}) \cdot \frac{1}{N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} z(\mathbf{u}_{\alpha} + \mathbf{h})$$

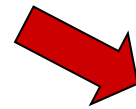
$$\text{con } \hat{C}(\mathbf{h}) = \hat{C}(-\mathbf{h})$$

Variogramma

$$2\gamma(\mathbf{h}) = \text{Var}\{Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})\} = E\{[Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})]^2\} - \cancel{E\{Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})\}^2}$$



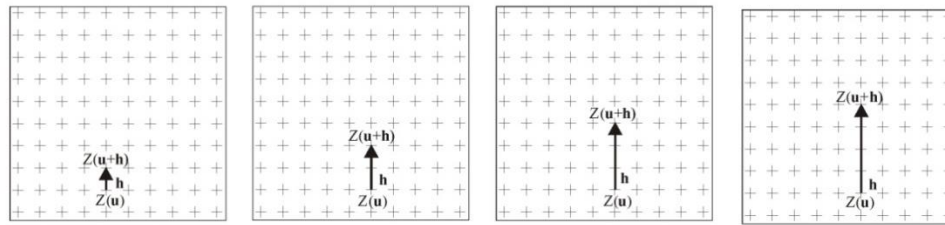
$$2\gamma(\mathbf{h}) = E\{[Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})]^2\}$$



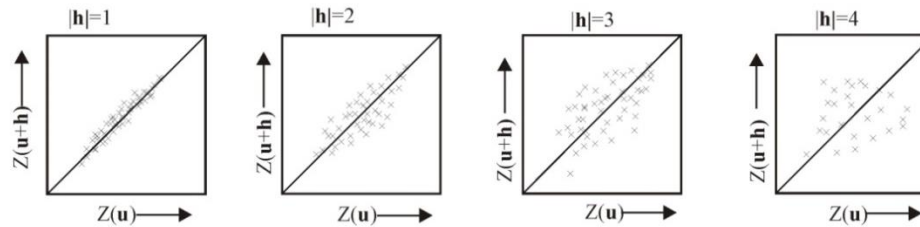
$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z(\mathbf{u}_{\alpha}) - z(\mathbf{u}_{\alpha} + \mathbf{h})]^2$$

$$\text{con } \hat{\gamma}(\mathbf{h}) = \hat{\gamma}(-\mathbf{h})$$

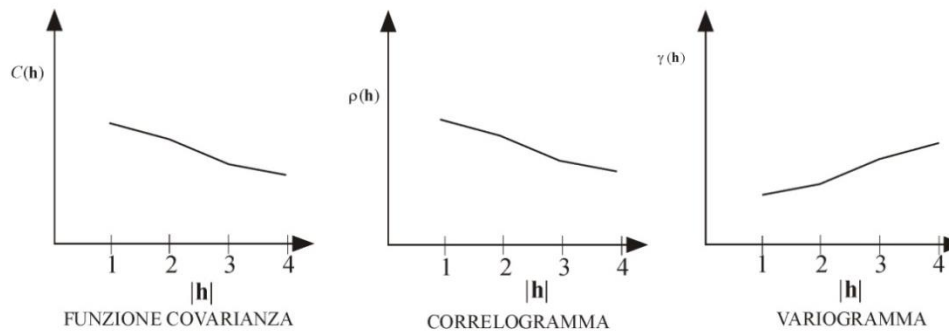
Indici continuità sperimentali



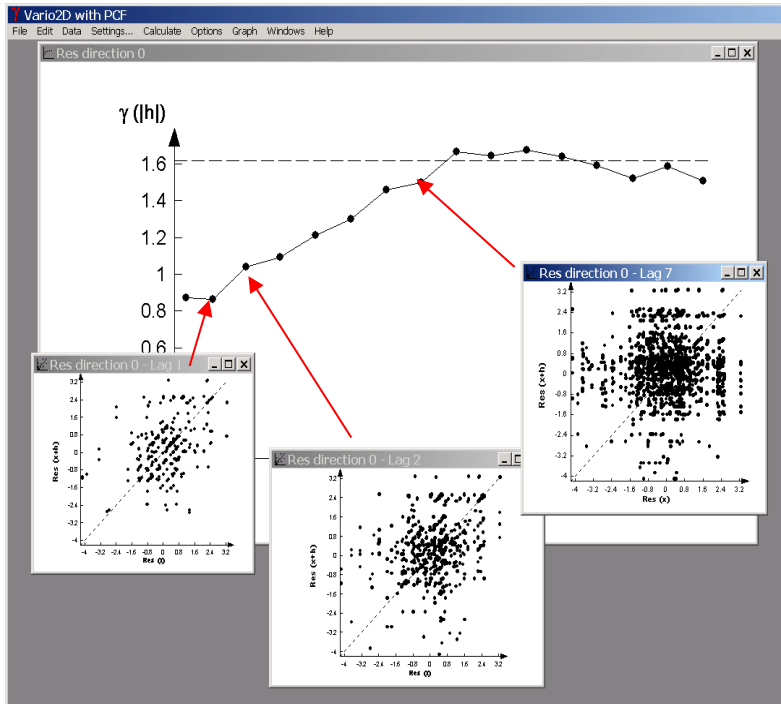
h- SCATTERPLOTS



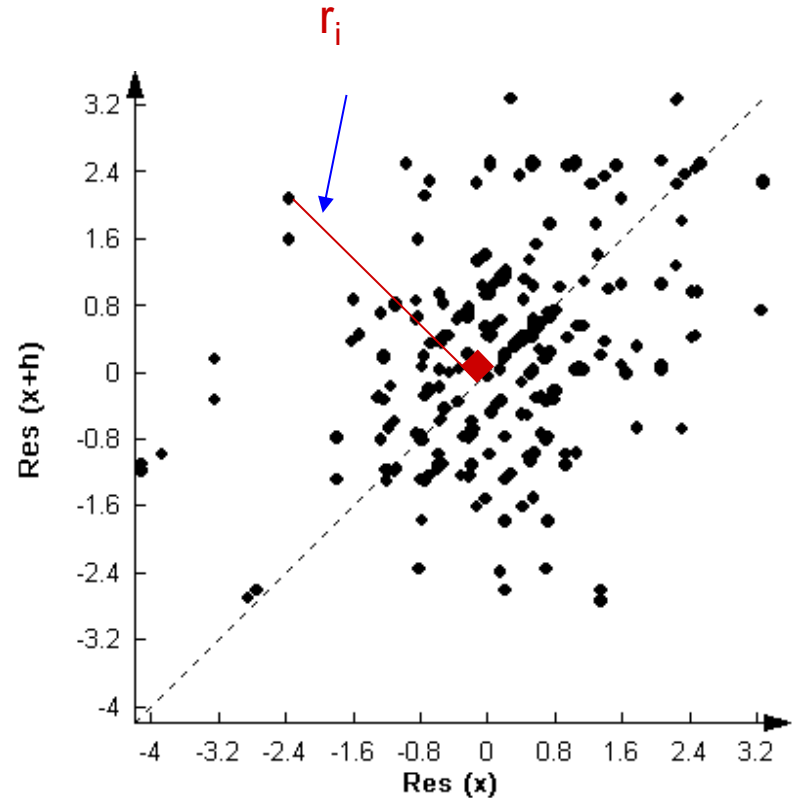
FUNZIONI DI CONTINUITA'



Il variogramma nel dettaglio



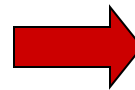
Momento di inerzia: $I = \sum m_i r_i^2$



Se $m_i = 1/n$

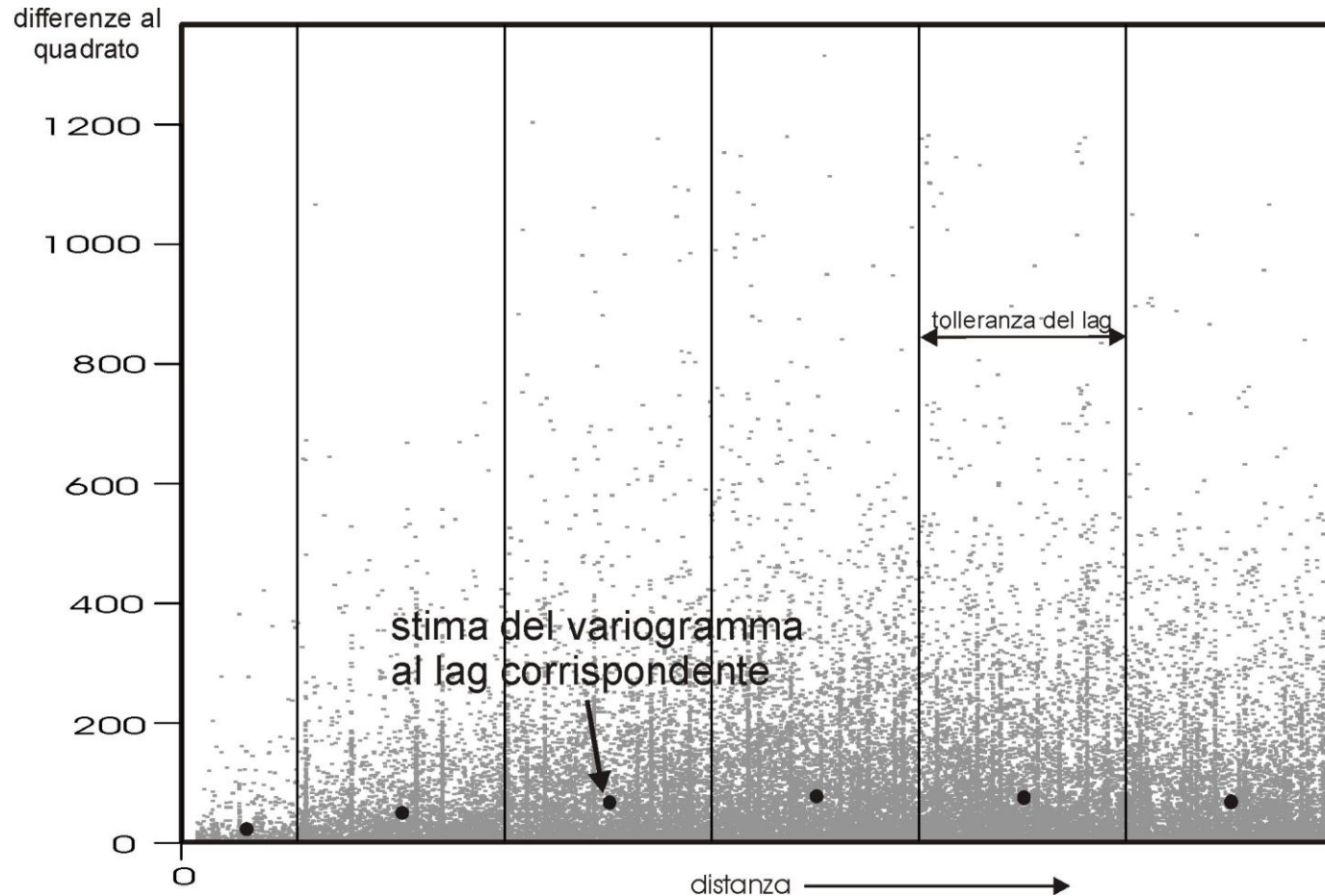
e $r_i^2 = (\text{Res}(x_i) - \text{Res}(x_i+h))^2 / 2$

$I = 1/2n \cdot \sum (\text{Res}(x_i) - \text{Res}(x_i+h))^2$



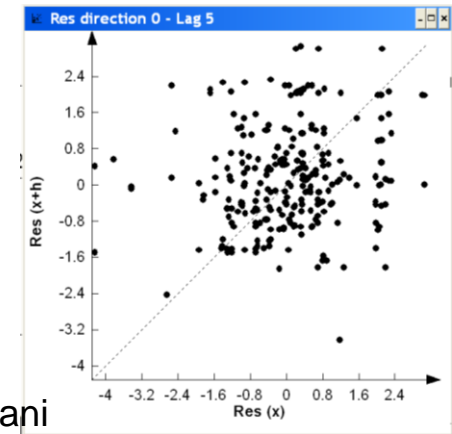
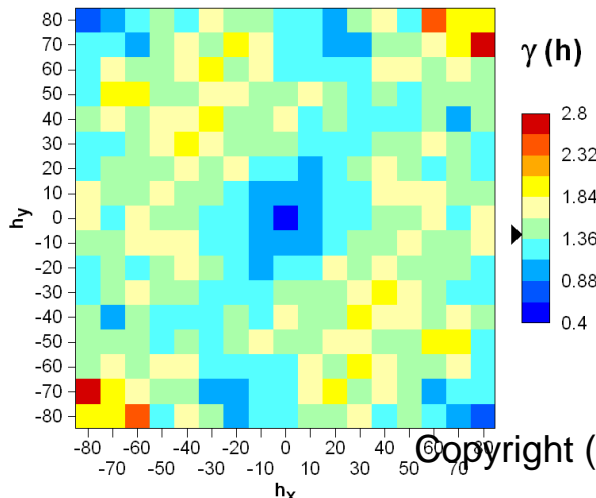
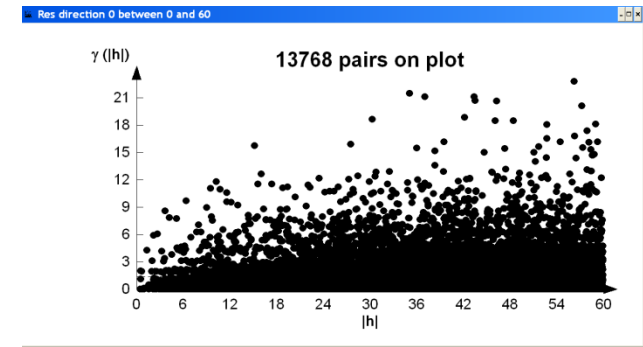
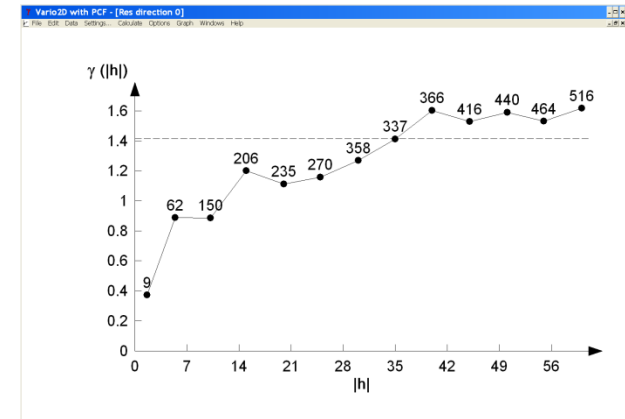
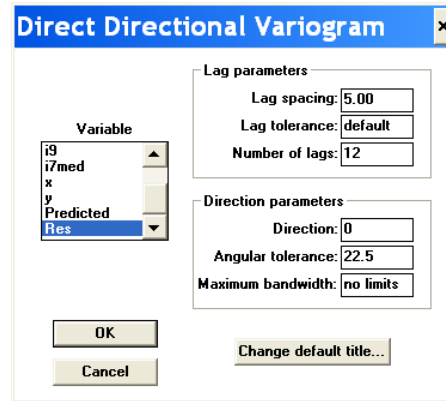
$$\{ (\| \mathbf{u}_j - \mathbf{u}_i \|, (z(\mathbf{u}_j) - z(\mathbf{u}_i))^2) \mid i \neq j, i, j = 1, \dots, n \}$$

Variogram cloud



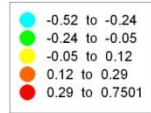
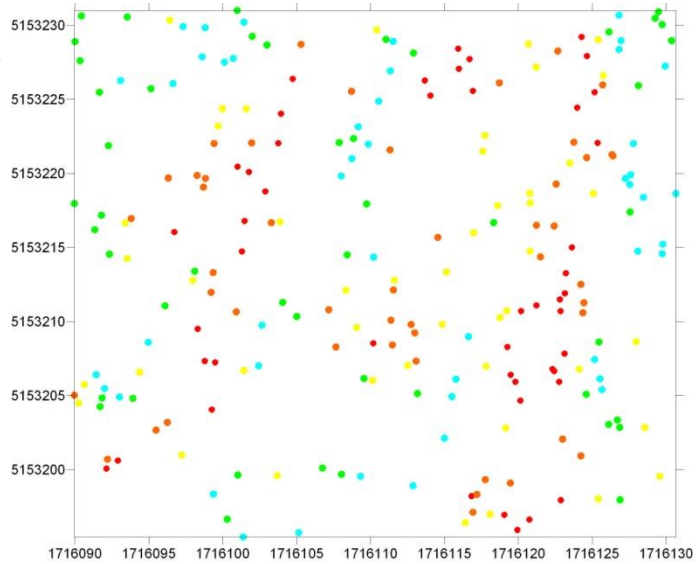
Gli strumenti per l'analisi di continuità

- H-scatterplot
- Variogram cloud
- Variogrammi direzionali
- Variogram map
- Altri stimatori di continuità:
(madogramma, variogramma standardizzato, ecc...)



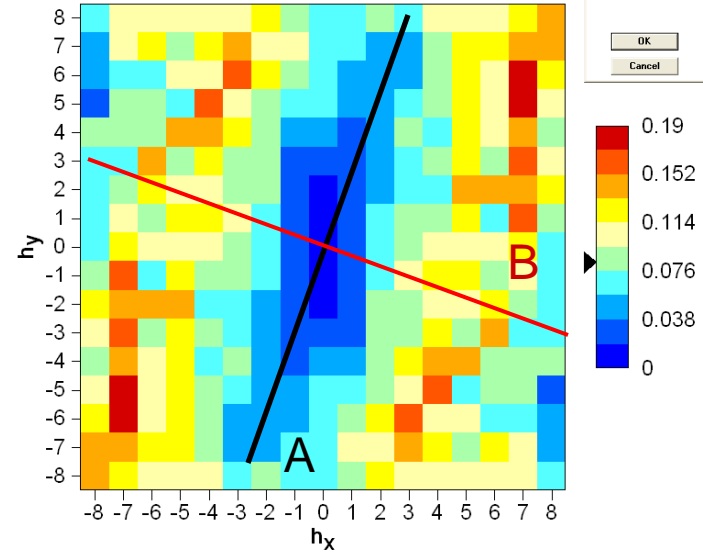
Gli strumenti per l'analisi di continuità

Distribuzione nello spazio di punti quotati (m)



N 234
 Min -0.52
 Max 0.75
 Mean 0.03
 Median 0.01
 First quartile -0.2
 Third quartile 0.22
 Std 0.28

Mappa del variogramma (m²)



Direct Variogram Surface

Variable:

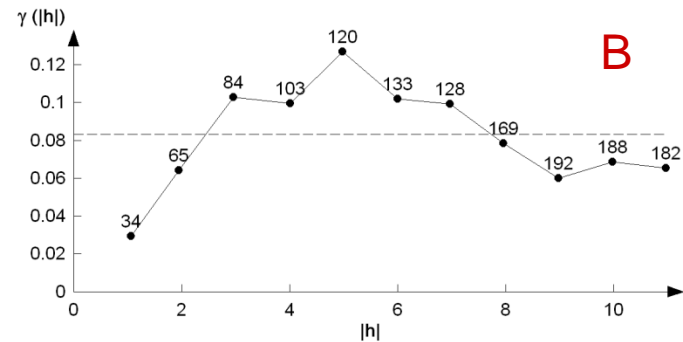
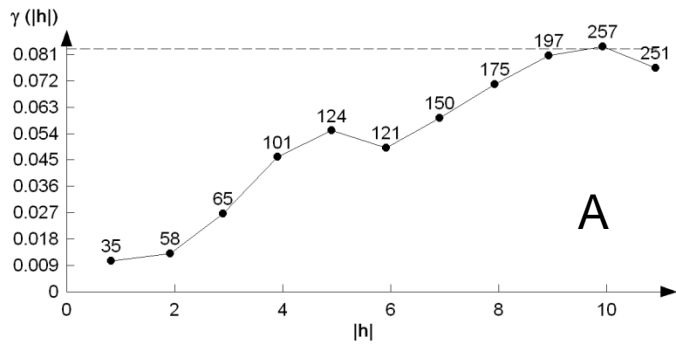
Parameters in the X-direction:
 Log spacing:
 Number of lags:

Parameters in the Y-direction:
 Log spacing:
 Number of lags:

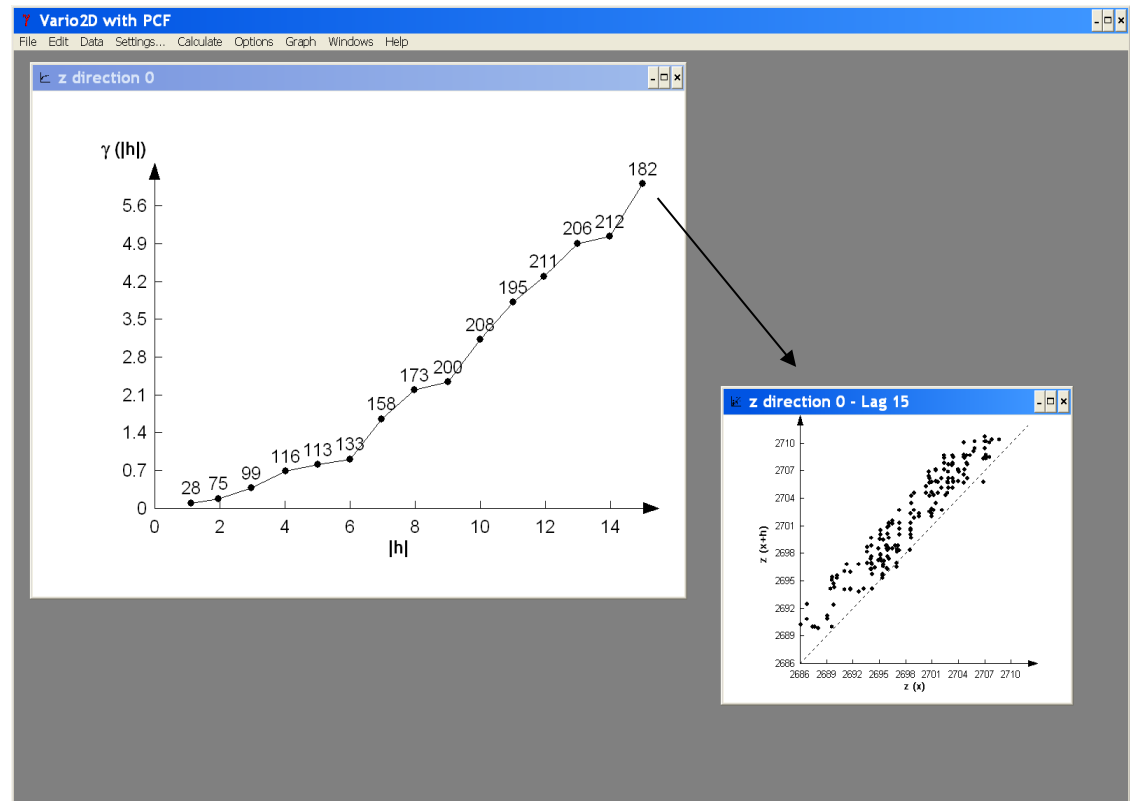
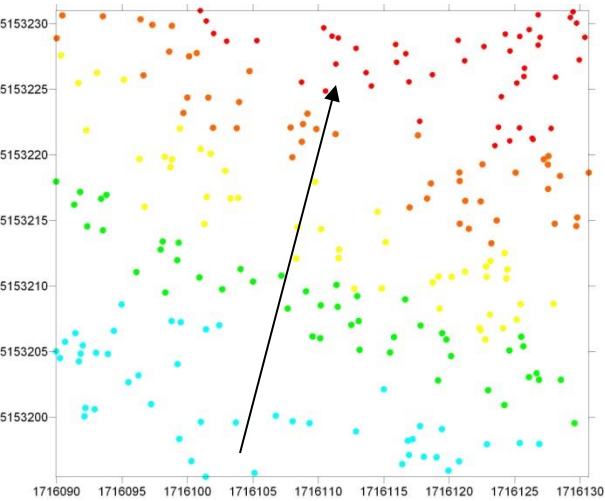
OK Cancel

Change default title...

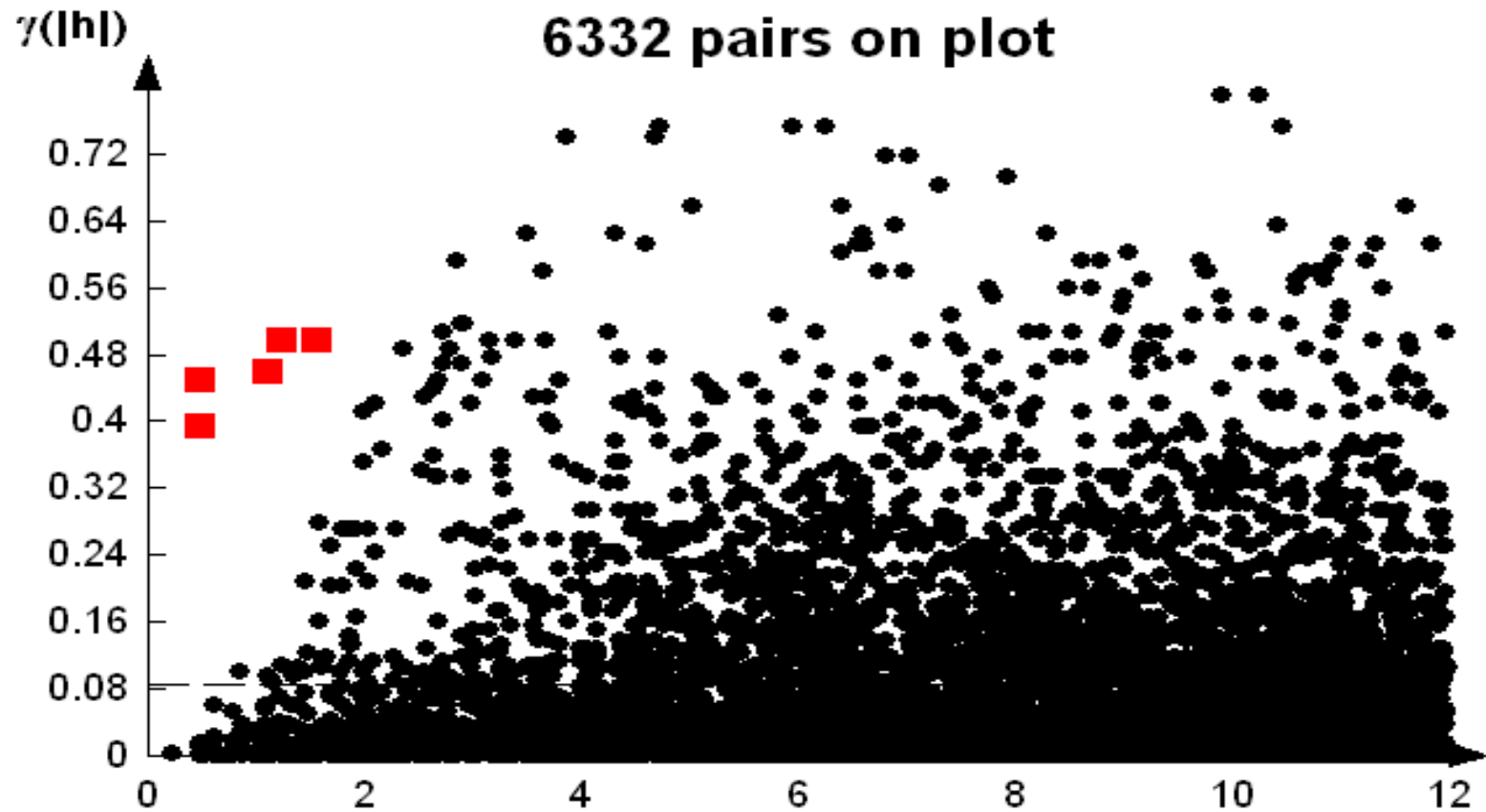
Variogrammi direzionali (m²)



Variogramma direzionale con presenza di trend

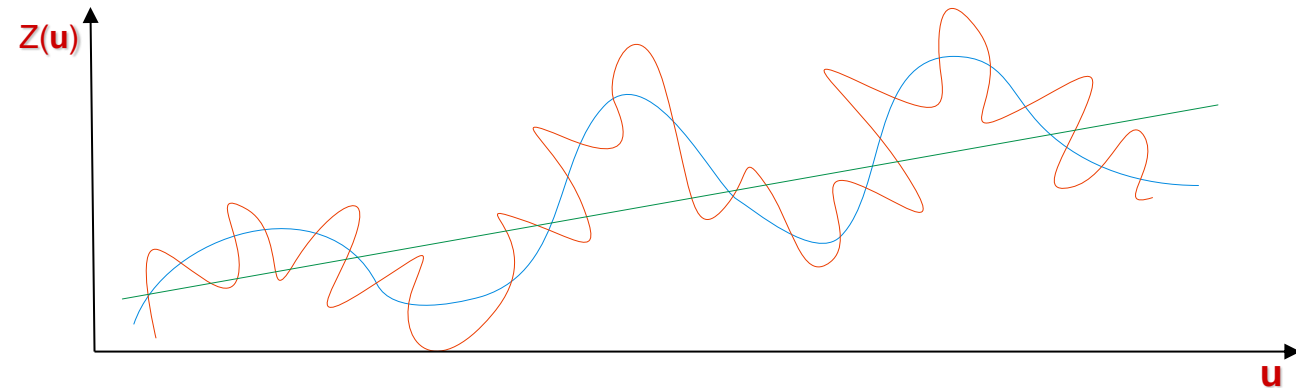


Individuazione di dati anomali locali mediante variogram cloud



$$Z(\mathbf{u}) = R(\mathbf{u}) + m(\mathbf{u})$$

$R(\mathbf{u})$ è una variabile aleatoria, chiamata residuo, che è stazionaria del secondo ordine (con funzione covarianza esistente e media nulla, a meno di ipotesi intrinseca), mentre $m(\mathbf{u})$ è un termine deterministico che rappresenta il modello di deriva (o trend)

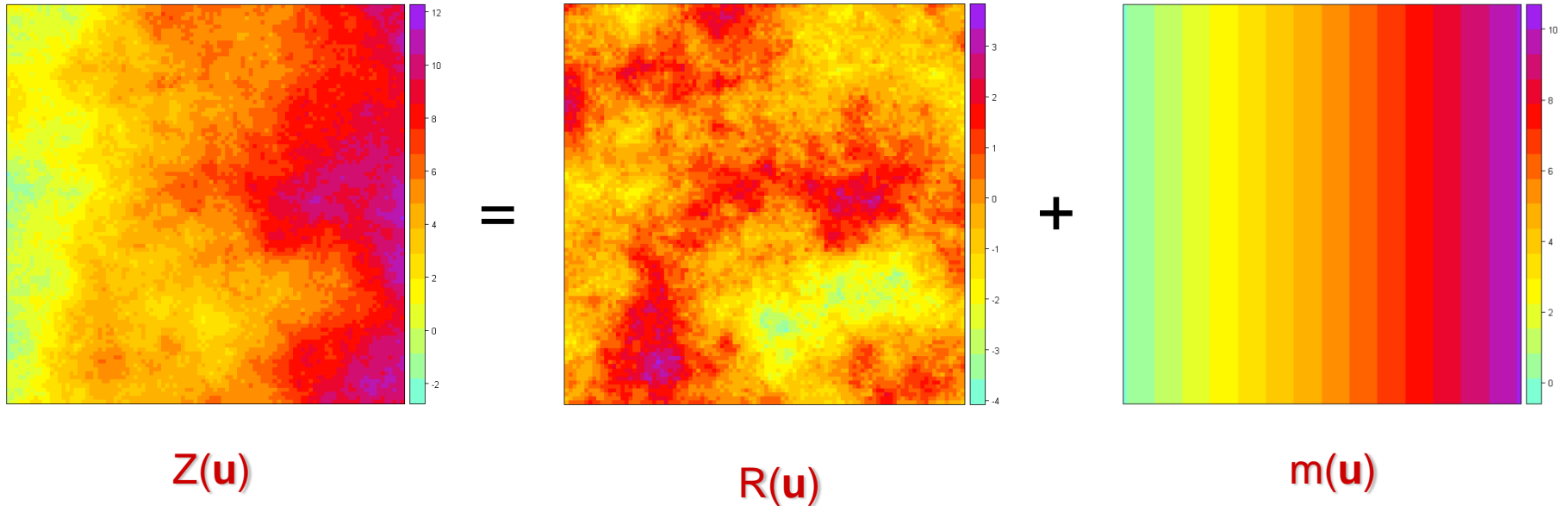


La scomposizione della proprietà in studio in una componente deterministica ed una aleatoria è un'operazione soggettiva ed è funzione della scala alla quale si effettua lo studio e della risoluzione del proprio dataset. Per tale motivo si cerca di utilizzare modelli di deriva semplici e possibilmente giustificati da processi fisico-chimici, caratterizzanti i fenomeni studiati.

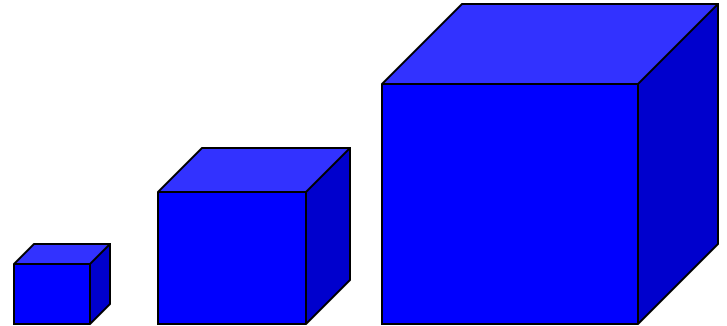
Dal punto di vista teorico si cade in una sorta di principio di indeterminazione:

- 1) Per calcolare il trend dovremmo conoscere il variogramma dei residui
- 2) Per calcolare il variogramma dei residui ci serve il trend

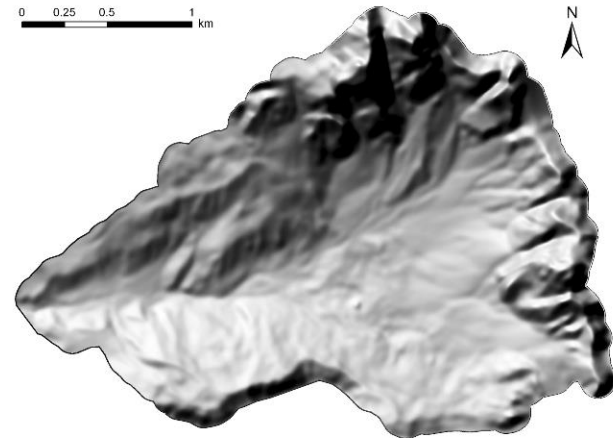
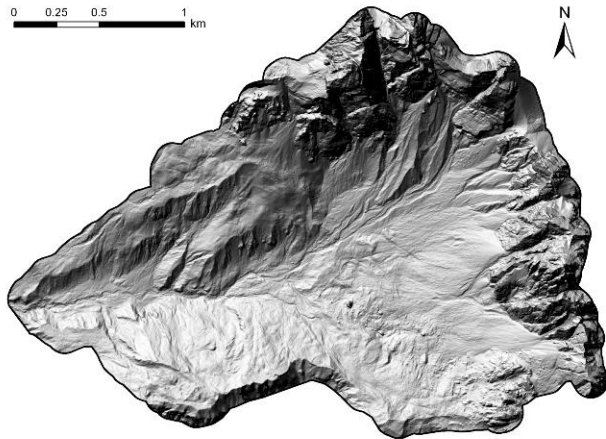
$$Z(\mathbf{u}) = R(\mathbf{u}) + m(\mathbf{u})$$



$$\sigma^2(\bullet, A) = \sigma^2(\bullet, B) + \sigma^2(B, A)$$

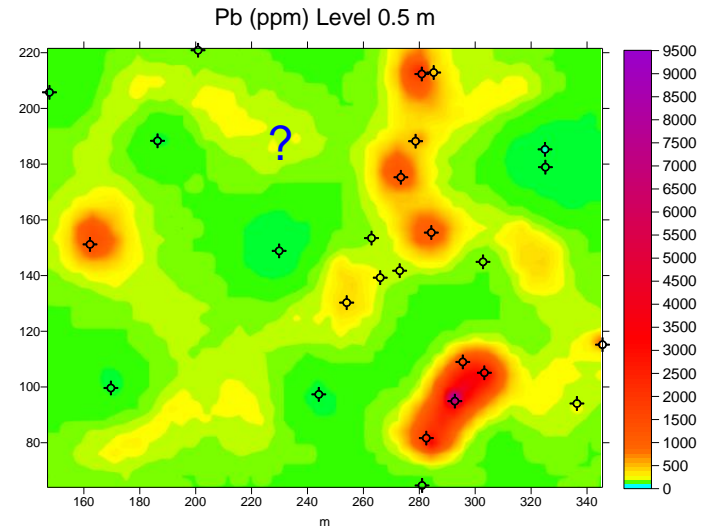


Stessa dimensione del pixel ma diverso supporto!!!



Con il termine Kriging si indica una famiglia di stimatori di regressione lineare non distorti e a minima varianza (BLUE Best Linear Unbiased Estimator)

L'algoritmo del kriging prende il nome da D.G: Krige (Sud Africa, 1950)



Stima

$$z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha}(\mathbf{u})z(\mathbf{u}_{\alpha})$$

Stimatore

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha}Z(\mathbf{u}_{\alpha})$$

I pesi λ_{α} vengono scelti in modo tale che lo stimatore sia ottimale e non distorto. L'ottimalità viene raggiunta minimizzando la varianza dell'errore (Isaaks e Srivastava, 1989, pp. 279-290):

$$\text{Var}\{Z^*(\mathbf{u})-Z(\mathbf{u})\}=\text{minima}$$

La non distorsione si ottiene imponendo la speranza matematica nulla:

$$E\{Z^*(\mathbf{u})-Z(\mathbf{u})\}=0$$

Kriging ordinario

In certi casi si può assumere che la media m , entro un sottodominio spaziale $W(\mathbf{u})$ del dominio di studio A , sia costante ma ignota. In questo caso, non conoscendola, è necessario filtrarla ponendo una condizione sui pesi:

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})$$

$$\begin{aligned} E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} &= E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha}) - Z(\mathbf{u})\right\} = \\ &= E\left\{\sum_{\alpha=1}^n \lambda_{\alpha} R(\mathbf{u}_{\alpha}) - R(\mathbf{u}) + \sum_{\alpha=1}^n \lambda_{\alpha} m - m\right\} = \\ &= \sum_{\alpha=1}^n \lambda_{\alpha} m - m \end{aligned}$$

$$E\{Z^*(\mathbf{u}) - Z(\mathbf{u})\} = 0$$

sotto la condizione che:

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

o ponendo $\lambda_0 = -1$

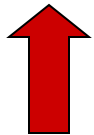
$$\sum_{\alpha=0}^n \lambda_{\alpha} = 0$$

Minimizzazione della
varianza dell'errore



$$Var\left\{\sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right\} = Var\left\{\sum_{\alpha=0}^n \lambda_{\alpha} R(\mathbf{u}_{\alpha})\right\} = \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} C_R(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta})$$

$$Var\left\{\sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right\} = Var\left\{\sum_{\alpha=0}^n \lambda_{\alpha} R(\mathbf{u}_{\alpha})\right\} = - \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta})$$



Non distorsione

Sistema di kriging ordinario (equazioni normali della regressione lineare)

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

Varianza di stima

$$\sigma^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^n \lambda_{\alpha} C(\mathbf{u}_{\alpha} - \mathbf{u}) - \mu(\mathbf{u})$$

$$2\gamma = E\{[Z(\mathbf{u}) - Z(\mathbf{u} + \mathbf{h})]^2\} = E\{[R(\mathbf{u}) - R(\mathbf{u} + \mathbf{h})]^2\}$$

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

Kriging in termini di matrici

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



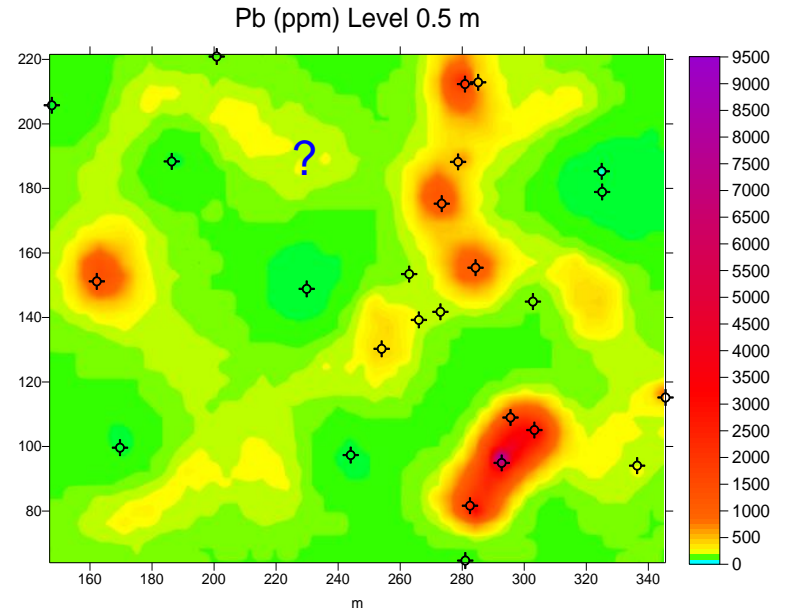
$$\begin{bmatrix} Cr_{11} & \dots & Cr_{1n} & 1 \\ \vdots & \ddots & \vdots & \vdots \\ Cr_{n1} & \dots & Cr_{nn} & 1 \\ 1 & \dots & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix} = \begin{bmatrix} Cr_{10} \\ \vdots \\ Cr_{n0} \\ 1 \end{bmatrix}$$

$$\mathbf{C} \cdot \mathbf{w} = \mathbf{D}$$

$$\mathbf{w} = \mathbf{C}^{-1} \cdot \mathbf{D}$$

Distanza

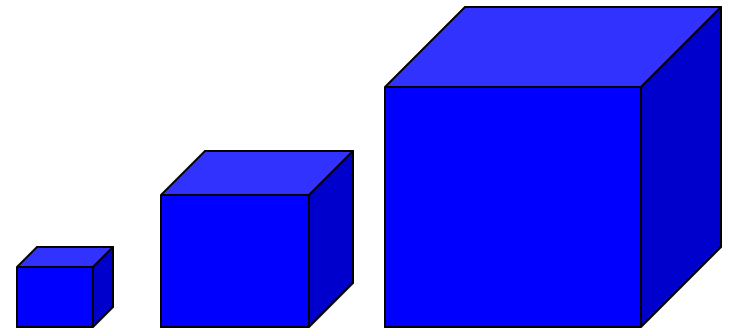
Clustering



- Criterio di pesatura (distanza e clustering)
- Previsore esatto (onora i dati noti)
- Stimatore non convesso (stime oltre il range dei dati, problemi con dati composizionali, probabilità e fuzzy membership)
- Varianza di stima (dipende solo dal variogramma e dalla geometria di campionamento)
- Smoothing (filtro passa basso, funzione della geometria dei dati e del variogramma)
- Possibilità di filtrare determinate componenti di variabilità

Block kriging: stime su supporti non puntuali:

- Filtraggio variabilità intra-blocco
- Riduzione della varianza di stima
- Implicito filtraggio del nugget



Modelli di variogramma

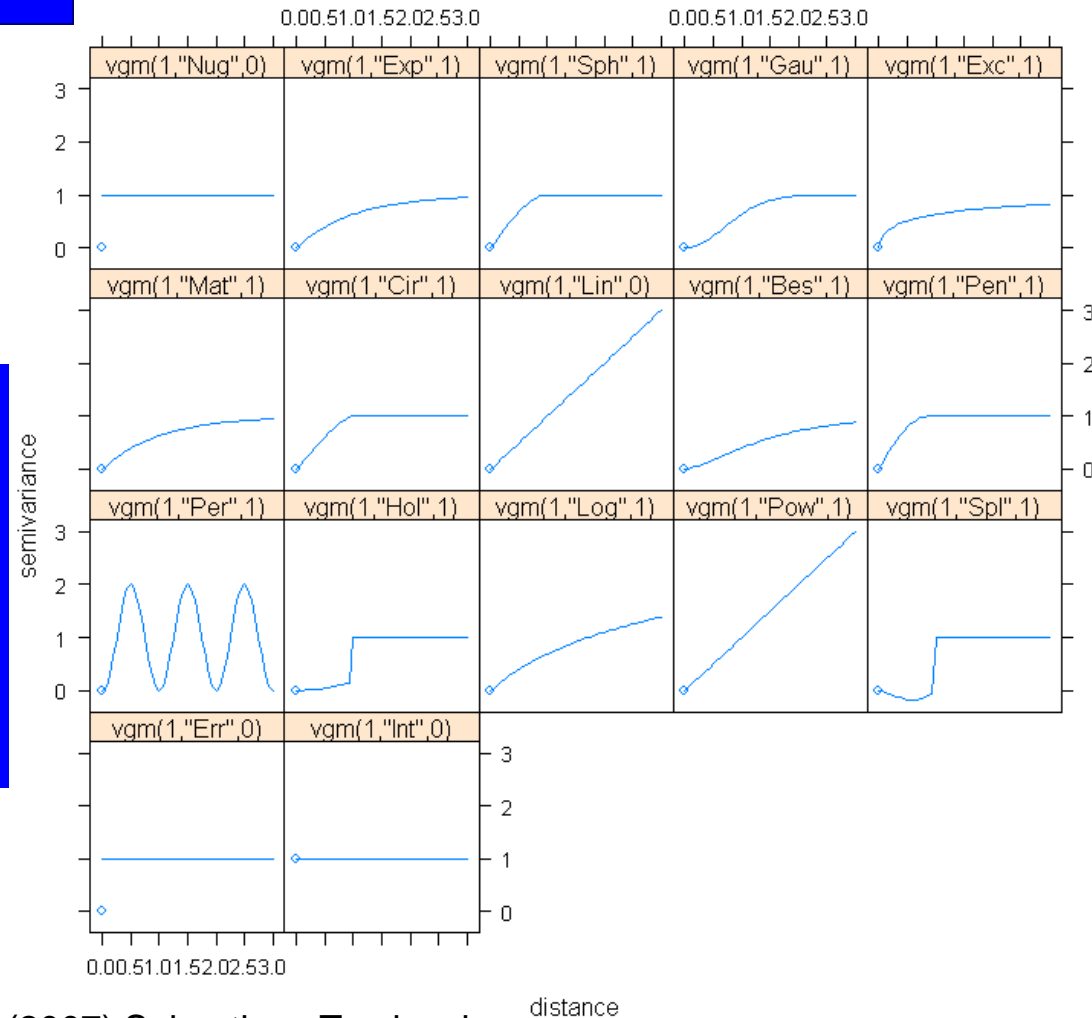
Il covariogramma e il negativo del variogramma devono essere delle funzioni condizionatamente definite positive.

$$\iff \text{Var}\left\{\sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha})\right\} = \text{Var}\left\{\sum_{\alpha=0}^n \lambda_{\alpha} R(\mathbf{u}_{\alpha})\right\} = -\sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} \gamma(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta})$$

Qualsiasi combinazione lineare risulta essere un modello autorizzato

- Nugget
- Range
- Sill
- Coefficienti di anisotropia
- Esponenti del modello Power

Informazioni esperte



Inferenza del modello variogramma

Ruolo fondamentale del variogramma:

-Fittaggio manuale

-Minimi quadrati pesati

-Massima verosimiglianza

fit	fit by	weight
0	-	- (no fit)
1	gstat	N_j
2	gstat	$N_j/\{\gamma(h_j)\}^2$
3	gnuplot	N_j
4	gnuplot	$N_j/\{\gamma(h_j)\}^2$
5	gstat	REML
6	gstat	no weights (OLS)
7	gstat	N_j/h_j^2

model	syntax	$\gamma(h)$	h range
Nugget	1 Nug(a)	0 1	$h = 0$ $h > 0$
Spherical	1 Sph(a)	$\frac{3h}{2a} - \frac{1}{2}(\frac{h}{a})^3$ 1	$0 \leq h \leq a$ $h > a$
Exponential	1 Exp(a)	$1 - \exp(-\frac{h}{a})$	$h \geq 0$
Linear	1 Lin(0)	h	$h \geq 0$
Linear-with-sill ¹	1 Lin(a)	$\frac{h}{a}$ 1	$0 \leq h \leq a$ $h > a$
Circular	1 Cir(a)	$\frac{2h}{\pi a} \sqrt{1 - (\frac{h}{a})^2} + \frac{2}{\pi} \arcsin \frac{h}{a}$ 1	$0 \leq h \leq a$ $h > a$
Pentaspherical	1 Pen(a)	$\frac{15h}{8a} - \frac{5}{4}(\frac{h}{a})^3 + \frac{3}{8}(\frac{h}{a})^5$ 1	$0 \leq h \leq a$ $h > a$
Gaussian	1 Gau(a)	$\gamma(h) = 1 - \exp(-(\frac{h}{a})^2)$	$h \geq 0$
Bessel ²	1 Bes(a)	$1 - \frac{h}{a} K_1(\frac{h}{a})$	$h \geq 0$
Logarithmic	1 Log(a)	0 $\log(h + a)$	$h = 0$ $h > 0$
Power	1 Pow(a)	h^a	$h \geq 0, 0 < a \leq 2$
Periodic	1 Per(a)	$1 - \cos(\frac{2\pi h}{a})$	$h \geq 0$

“Giocando” con il kriging ordinario

E{Z}-Kriging - Ordinary Kriging explained (version: 0.2, May 2002)

semivariogram

Ordinary Point Kriging weights

data configuration

presets: manual

number of sampling points: 6

id	easting	northing	value
0	149	149	21.8
1	99	62	3
2	141	49	3
3	162	58	14
4	199	236	50
5	108	194	2
6	236	129	59

semivariogram definition

semivariogram model: Spherical

c0: 0

c1: 100

a: 118

nugget: 0 range: 118 sill: 100

Ordinary Kriging type

Point Kriging Block Kriging

block size

x: 10

y: 10

results

Prediction: 21.8 Variance prediction error: 96.3

Matrices Help

Copyright 2001-2002 by Dennis Walvoort. Laboratory for Soil Science & Geology, Wageningen University, The Netherlands

Kriging universale: modelli polinomiali del trend

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C_R(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \sum_{k=0}^K \mu_k(\mathbf{u}) f_k(\mathbf{u}_{\alpha}) = C_R(\mathbf{u}_{\alpha} - \mathbf{u}) \\ \sum_{\beta=1}^n \lambda_{\beta} f_k(\mathbf{u}_{\beta}) = f_k(\mathbf{u}) \quad k = 0, \dots, K \end{cases} \quad \alpha = 1, \dots, n(\mathbf{u})$$

Possibilità di stima del trend o dei coefficienti di regressione

Informazione secondaria e uso integrato di diverse tipologie di informazione:

Esempi: intensità sismica e superficie quadratica, falda e direzione di deflusso, piovosità e dtm, profondità falda e dtm, salinizzazione suolo e distanza dalla costa, salinità acque marine e distanza dalle foci, porosità e tipologia di suolo, porosità ed impedenza acustica, sst e misurazioni in campo, ecc.

Kriging con deriva esterna (e regression kriging): correlazione lineare tra trend ed una variabile secondaria nota in tutti i punti del dominio di interesse

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C_R(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu_0(\mathbf{u}) + \mu_1(\mathbf{u}) y(\mathbf{u}_{\alpha}) = C_R(\mathbf{u}_{\alpha} - \mathbf{u}) \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \\ \sum_{\beta=1}^n \lambda_{\beta} y(\mathbf{u}_{\beta}) = y(\mathbf{u}) \end{cases} \quad \alpha = 1, \dots, n$$

Cokriging: correlazione tra la componente aleatoria e una o più variabili ausiliarie (analisi e modellazione dei cross-variogrammi)

$$\begin{cases} \sum_{j=1}^{N_v} \sum_{\beta_j=1}^{n_j} \lambda_{\beta_j} C_{ij}(\mathbf{u}_{\alpha_i} - \mathbf{u}_{\beta_j}) + \mu_i(\mathbf{u}) = C_{i1}(\mathbf{u}_{\alpha} - \mathbf{u}) \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} = 1 \\ \sum_{\beta_i=1}^{n_i} \lambda_{\beta_i} = 0 \end{cases} \quad \begin{matrix} \alpha_i = 1, \dots, n_i \quad i = 1, \dots, N_v \\ i = 2, \dots, N_v \end{matrix}$$

Kriging universale o Kriging con trend

Si assume che $m(\mathbf{u})$ vari dolcemente entro un sottodominio $W(\mathbf{u})$, sia di forma polinomiale e funzione della posizione \mathbf{u} secondo dei coefficienti costanti entro il medesimo sottodominio:

$$m(\mathbf{u}') = \sum_{k=0}^K \mathbf{a}_k(\mathbf{u}') f_k(\mathbf{u}') \quad \text{con} \quad \mathbf{a}_k(\mathbf{u}') \approx \mathbf{a}_k \quad \forall \mathbf{u}' \in W(\mathbf{u})$$

Ad esempio, considerando un caso unidimensionale, una deriva quadratica (di ordine 2) potrebbe essere espressa come:

$$m(x) = a_0 + a_1 x + a_2 x^2$$

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha}) + \sum_{k=0}^K a_k(\mathbf{u}) \left[f_k(\mathbf{u}) - \sum_{\alpha=1}^n \lambda_{\alpha} f_k(\mathbf{u}_{\alpha}) \right]$$

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha}) \quad \text{con} \quad \sum_{\alpha=1}^n \lambda_{\alpha} f_k(\mathbf{u}_{\alpha}) = f_k(\mathbf{u}) \quad k = 0, \dots, K$$

$$VAR\left\{ \sum_{\alpha=0}^n \lambda_{\alpha} Z(\mathbf{u}_{\alpha}) \right\} = VAR\left\{ \sum_{\alpha=0}^n \lambda_{\alpha} R(\mathbf{u}_{\alpha}) \right\} = \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} C_R(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) \quad \rightarrow$$

Minimizzazione della varianza con nuove condizioni sui pesi

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C_R(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \sum_{k=0}^K \mu_k(\mathbf{u}) f_k(\mathbf{u}_{\alpha}) = C_R(\mathbf{u}_{\alpha} - \mathbf{u}) \\ \sum_{\beta=1}^n \lambda_{\beta} f_k(\mathbf{u}_{\beta}) = f_k(\mathbf{u}) \quad k = 0, \dots, K \end{cases} \quad \alpha = 1, \dots, n(\mathbf{u})$$

$$\sigma^2(\mathbf{u}) = C_R(0) - \sum_{\alpha=1}^n \lambda_{\alpha} C_R(\mathbf{u}_{\alpha} - \mathbf{u}) - \sum_{k=0}^K \mu_k(\mathbf{u}) f_k(\mathbf{u})$$

$$z_1^*(\mathbf{u}) = \sum_{i=1}^{N_v} \sum_{\alpha_i=1}^{n_i} \lambda_{\alpha_i}(\mathbf{u}) z_i(\mathbf{u}_{\alpha_i})$$

$$\text{Var}\left\{ \sum_{i=1}^{N_v} \sum_{\alpha_i=0}^{n_i} \lambda_{\alpha_i} z_i(\mathbf{u}_{\alpha_i}) \right\} = \sum_{i=1}^{N_v} \sum_{\alpha=0}^n \sum_{\beta=0}^n \lambda_{\alpha} \lambda_{\beta} C_{iR}(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta})$$

con

$$\sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} = 1$$

$$\sum_{\beta_i=1}^{n_i} \lambda_{\beta_i} = 0$$

$i = 2, \dots, N_v$

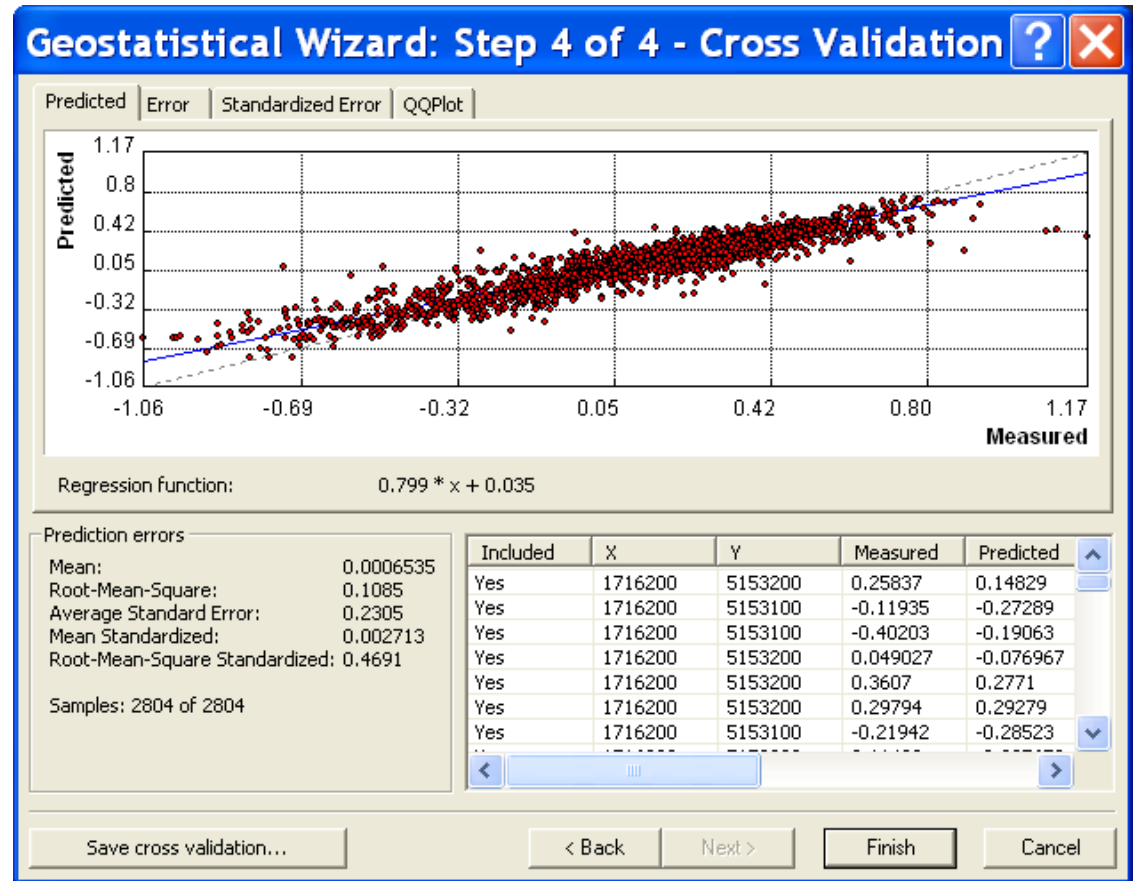
$$\left\{ \begin{array}{l} \sum_{j=1}^{N_v} \sum_{\beta_j=1}^{n_j} \lambda_{\beta_j} C_{ij}(\mathbf{u}_{\alpha_i} - \mathbf{u}_{\beta_j}) + \mu_i(\mathbf{u}) = C_{i1}(\mathbf{u}_{\alpha} - \mathbf{u}) \\ \sum_{\beta_1=1}^{n_1} \lambda_{\beta_1} = 1 \\ \sum_{\beta_i=1}^{n_i} \lambda_{\beta_i} = 0 \end{array} \right. \quad \begin{array}{l} \alpha_i = 1, \dots, n_i \quad i = 1, \dots, N_v \\ \\ i = 2, \dots, N_v \end{array}$$

Oltre al calcolo dei variogrammi diretti vi è la necessità di calcolare i cross-variogrammi. Difficoltà di inferenza di un modello di coregionalizzazione autorizzato.

$$\hat{\gamma}_{ab}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{\alpha=1}^{N(\mathbf{h})} [z_a(\mathbf{u}_{\alpha}) - z_a(\mathbf{u}_{\alpha} + \mathbf{h})] \cdot [z_b(\mathbf{u}_{\alpha}) - z_b(\mathbf{u}_{\alpha} + \mathbf{h})]$$

con $\hat{\gamma}_{ab}(\mathbf{h}) = \hat{\gamma}_{ab}(-\mathbf{h})$

- Cross-validation
- Jackknifing
- Analisi sensitività
- Training e validation datasets

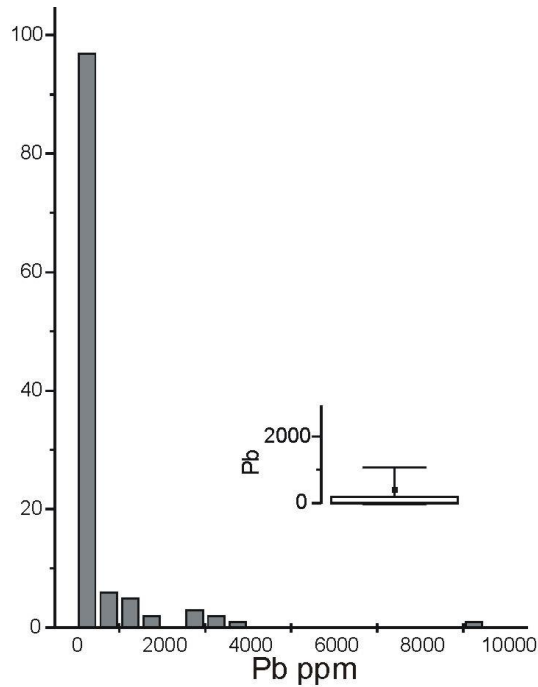




- Controllo dei dati: presenza errori, sistema di coordinate, considerazioni sul supporto di misura, variabilità temporale, variabili secondarie, ...
- Statistiche descrittive: istogrammi, boxplots, indici statistici, ...
- Analisi esplorativa spaziale: analisi della geometria di campionamento, mappe a classi di valori, analisi a finestre mobili, studio della continuità spaziale, ...

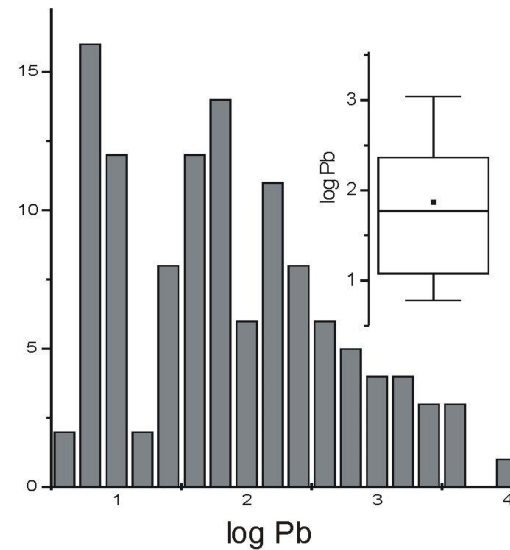
- Vi è continuità spaziale?
 - I dati sono sufficienti?
- Il dominio spaziale va scomposto?
 - C'è trend?
- I dati vanno trasformati?
- Vi è congruenza con la fenomenologia indagata?

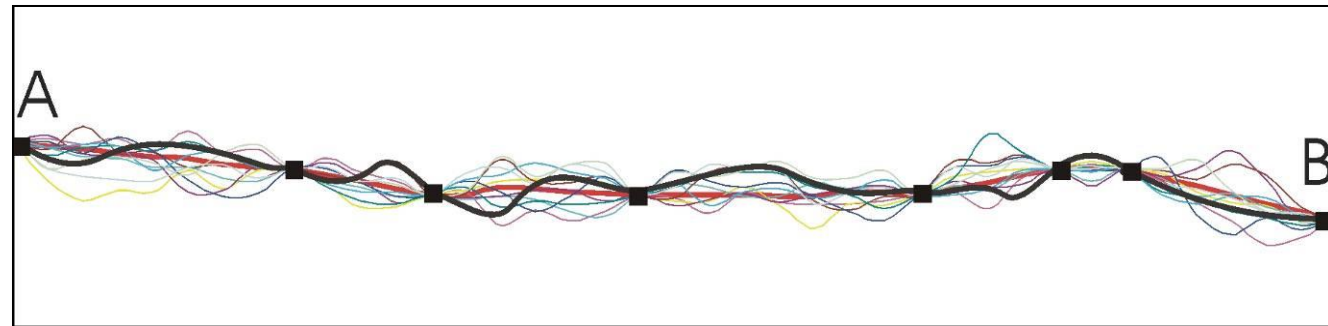
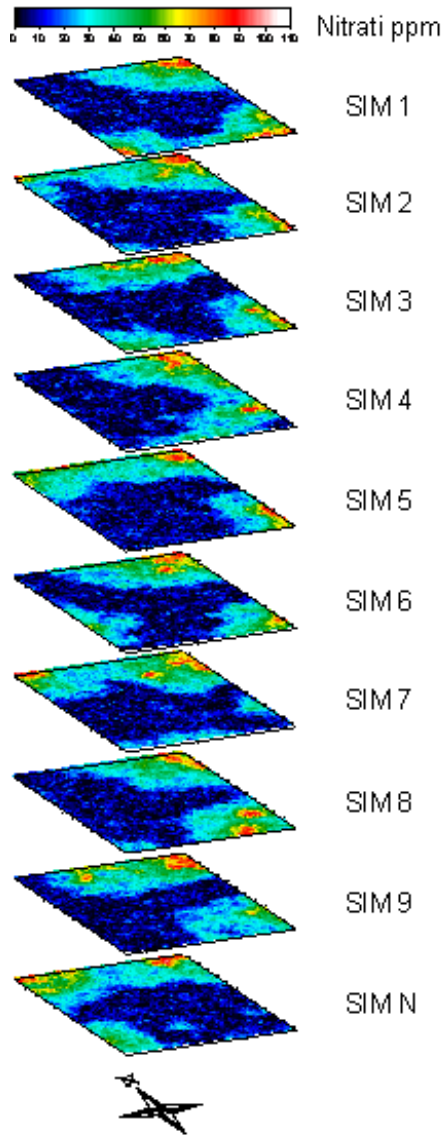
La trasformazione dei dati



Pb
Number of values 117
Minimum 4.99
Maximum 9499.99
Range 9494.99
Mean 430.13
Median 59.00
First quartile 11.99
Third quartile 229.99
Standard deviation 1126.45
Coefficient of variation 2.61

Log Pb
Number of values 117
Minimum 0.69
Maximum 3.97
Range 3.27
Mean 1.86
Median 1.77
First quartile 1.07
Third quartile 2.36
Standard deviation 0.81
Coefficient of variation 0.43





Simulazioni condizionate ovvero scenari compatibili con le informazioni disponibili:

- Onorano i dati noti
- Riproducono la struttura di variabilità dei dati

- Calcolo di P.D.F. condizionate congiunte
- Procedure di upscaling
- Funzioni di trasferimento (modelli di flusso sotterraneo, modelli idrologici, ecc.) e analisi degli scenari

- Burrough P. A. and McDonnel Rachael A., Principles of Geographical Information Systems: Oxford University Press, Oxford, 2000, 333 pp.
- Chilès J. P. and Delfiner P. , Geostatistics (Modeling Spatial Uncertainty): John Wiley & Sons INC., New York, 1999, 695 pp.
- **Clark I. and Harper W. V , Practical Geostatistics 2000: Ecosse North America Llc, Columbus Ohio, 2002, 400 pp.**
- Cressie N. E., Statistic for Spatial Data, revised edition: John Wiley & Sons Inc., New York, 1993, 900 pp.
- Davis J.C., Statistics and Data Analysis in Geology, John Wiley & Sons, New York, 2002, 638 pp.
- Deutsch C V. and Journel A. G., GSLIB Geostatistical Software Library and User's Guide: Oxford University Press, New York, 1998, 369 pp.
- Deutsch C V., Geostatistical Reservoir Modeling: Oxford University Press, New York, 2002, 376 pp.
- Diggle P. J., Tawn J. A., Moyed R. A., 1998, Model based geostatistics (with discussion): Applied Statistics, no. 43, pp. 299-350.

•Goovaerts P., Geostatistics for Natural Resources Evaluation: Oxford University Press, New York, 1997, 483 pp.

•Huijbregts C. J. and Journel A. G., Mining Geostatistics: Academic Press, New York, 1978, 600 pp.

•Isaaks E. H., Srivastava R. M., An Introduction to Applied Geostatistics: Oxford University Press, New York, 1989, 561 pp.

•Journel A. G., Fundamentals of Geostatistics in Five Lessons, short course in geology, volume 8: American Geophysical Union, Washington D.C., 1989, 40 pp.

•Kitanidis P. K., Introduction to Geostatistics: Cambridge University Press, New York, 1997, 249 pp.

•Pannatier Y., Variowin (Software for Spatial Data Analysis in 2D): Springer, New York, 1996, 91 pp.