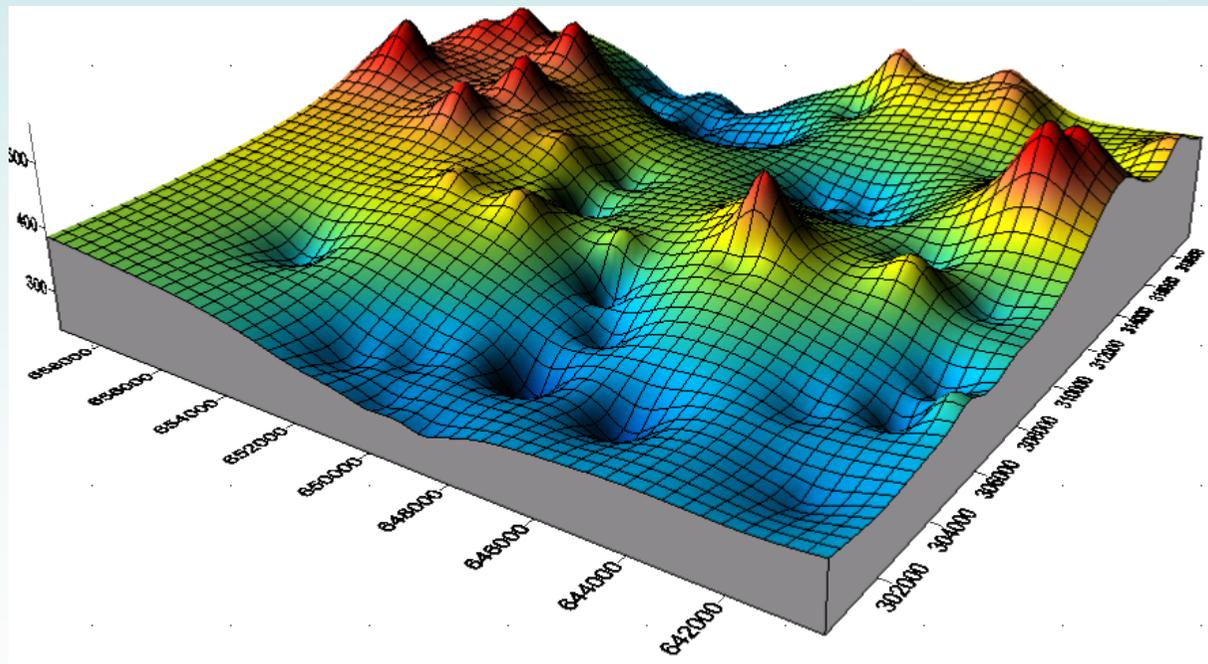


# Geostatistica e interpolazione spaziale

Valerio Noti



<http://it.linkedin.com/in/valerionoti>

[www.spatialanalysisonline.com](http://www.spatialanalysisonline.com)

# Metodi geostatistici

La superficie interpolata è concettualizzata come una delle superfici possibili che potrebbero essere osservate, tutte coerenti con i dati di input.

La geostatistica richiede l'assunzione che la variabile studiata sia casuale, che la realtà osservata sia il risultato di uno o più processi casuali.

Per ogni punto dello spazio non c'è solo un valore per una proprietà ma un intero insieme di valori. Il valore osservato diventa quindi uno tra gli infiniti valori possibili assunti da qualche distribuzione di probabilità per una qualche legge.

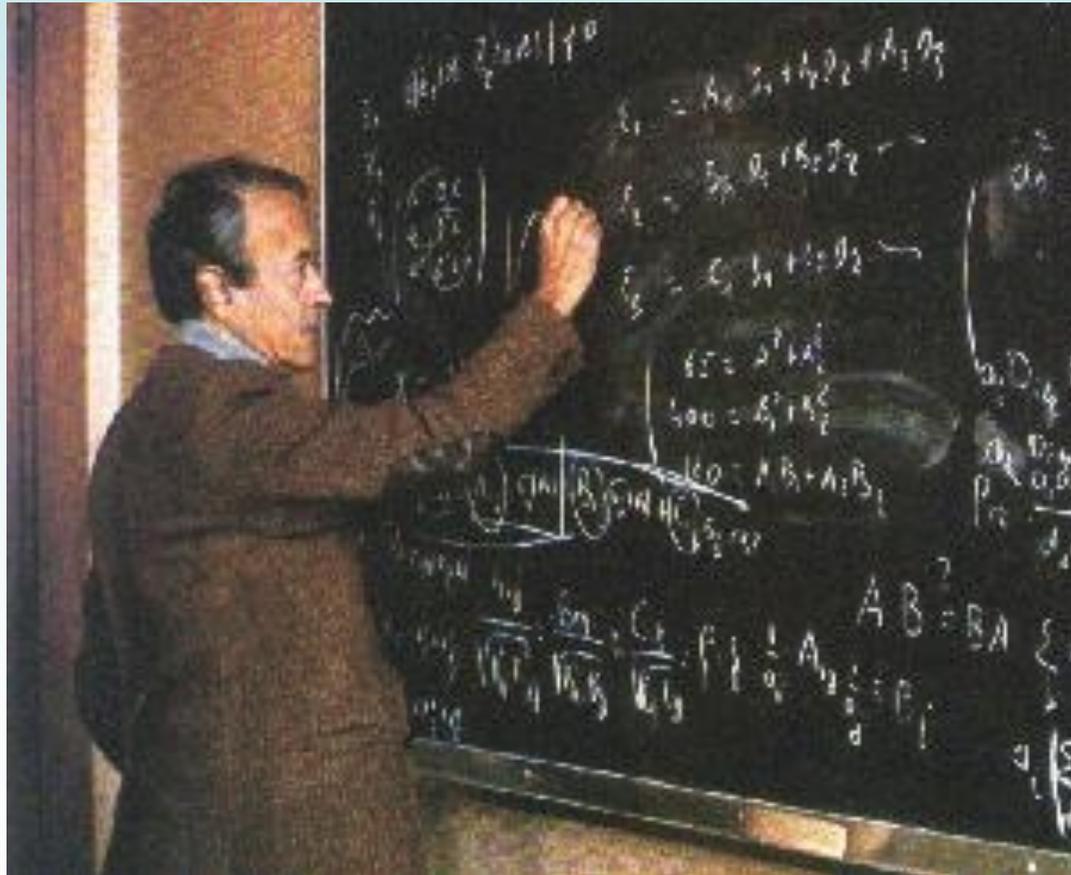
In ogni punto dello spazio  $X_0$  c'è una **variazione** di una proprietà  $Z(x_0)$ , concetto non presente nella teoria classica. Tale proprietà è trattata come una variabile casuale (aleatoria) generalmente continua, con una media, una varianza e una funzione di probabilità di densità cumulativa. **Questa variabile ha una distribuzione di probabilità da cui viene estratto il valore reale.**

Se noi conosciamo approssimativamente la distribuzione possiamo stimare i valori nei punti non campionati dai vicini campionati e assegnare gli errori alle predizioni.

# Metodi geostatistici

## Teoria delle Variabili Regionalizzate di Matheron (1965)

I valori delle VR tendono ad essere correlati (luoghi più vicini sono più simili ad altri maggiormente distanti) a certe scale. La teoria di Matheron quantifica questa correlazione.



# Le Variabili Regionalizzate

Una variabile regionalizzata  $Z(x)$  è una variabile con valore fortemente dipendente dalla posizione spaziale.

$$Z(x) = \alpha + R(x)$$

$\alpha$  = componente casuale

$R(x)$  = componente regionalizzata

La condizione è che  $R(x)$  sia preponderante rispetto alla componente casuale

**La geostatistica è la scienza che studia  
le Variabili Regionalizzate**

**La funzione utilizzata per il loro studio è il variogramma**

# Geostatistica: flusso di lavoro

**Campionamento  
(progettazione, acquisizione dati)**

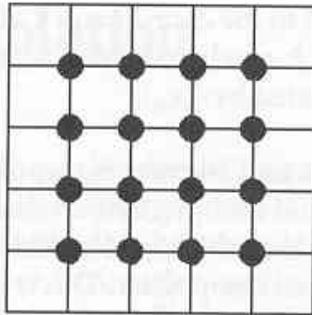
**Analisi statistica (explanatory data analysis):** Analizza le proprietà spaziali e statistiche dei dati esame distribuzione, identificazione ed eliminazione trend, identificazione anisotropie, ecc.

**Stima del variogramma sperimentale e scelta del modello da utilizzare.**

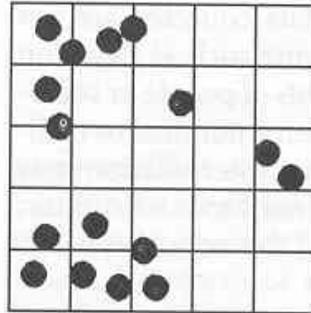
Interpolazione **Calcolo superficie statistica** di predizione e **Carta degli errori**. La predizione dei valori incogniti avviene attraverso il modello del variogramma, la configurazione spaziale dei dati e il valore dei punti misurati.

**Diagnostica:** comprendere “quanto bene” il modello predice i valori incogniti. Analisi di errore: cross validation, validation points, ecc.

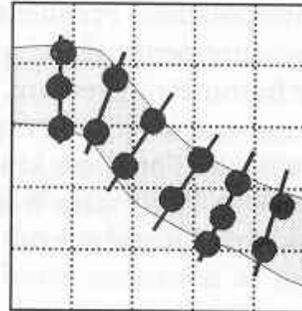
# Strategie di campionamento



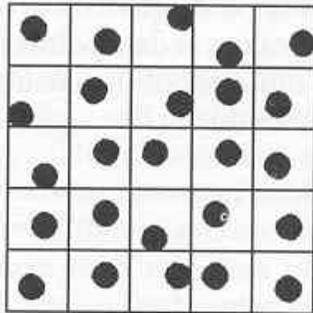
a) regular sampling



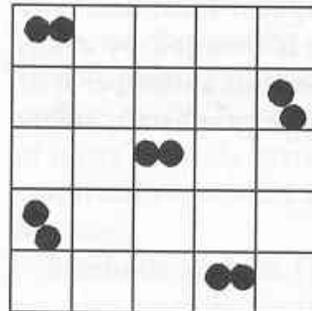
b) random sampling



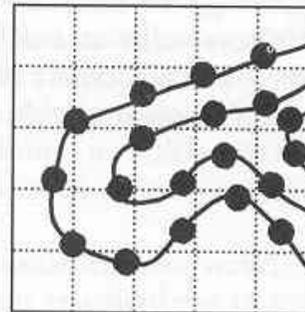
e) transect sampling



c) stratified random sampling



d) cluster sampling



f) contour sampling

Burrough, McDonnel, 1998. Principles of GIS

La strategia di campionamento è sicuramente più importante nell'approccio deterministico ma influenza fortemente anche i risultati delle elaborazioni stocastiche.

**Non esiste alcun modo di produrre una carta di qualità da punti misurati di bassa qualità**

# Campionamento

## Requisiti del dataset:

- Sufficientemente ampio (numero minimo di punti per analisi geostatistica: 30-50)
- Imparziale (es, nessuna preferenza verso misure nei luoghi più accessibili)
- Rappresentativo
- Indipendente
- Acquisito con una significativa precisione (la migliore possibile in base alla scala e alle caratteristiche degli strumenti)
- Uniformità delle misurazioni (es. stessa stagione, stessi strumenti, stesse condizioni, ecc.)
- Tenere conto delle anisotropie
- Evitare cluster di campioni

**Non esiste alcun modo di produrre una carta di qualità da punti misurati di bassa qualità**

# Passo di campionamento

**Il passo di campionamento deve essere relazionato alla scala di variazione della proprietà in quella regione**

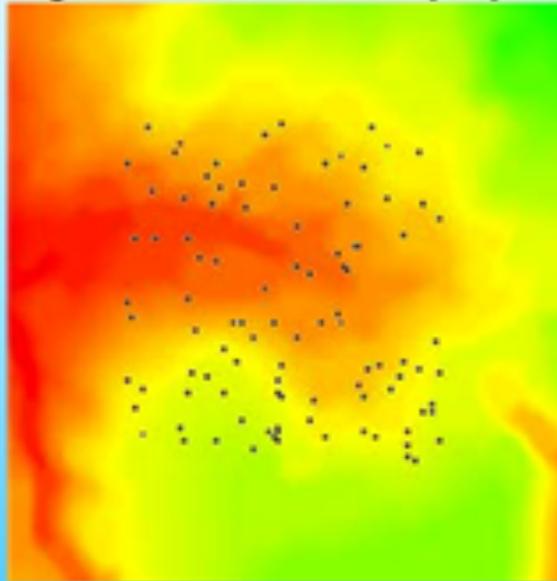
## **Due situazioni:**

Terra incognita (non sappiamo niente della scala di variabilità, rischio di sotto o sovra campionamento, alto rischio di perdita economica)

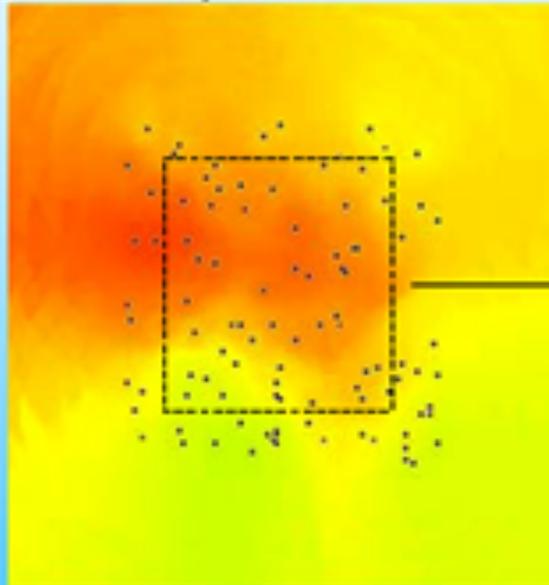
Abbiamo una precedente indagine (situazione ottimale, fondamentale l'analisi geostatistica sui vecchi dati prima di procedere)

.....attenzione ai margini dell'area di indagine

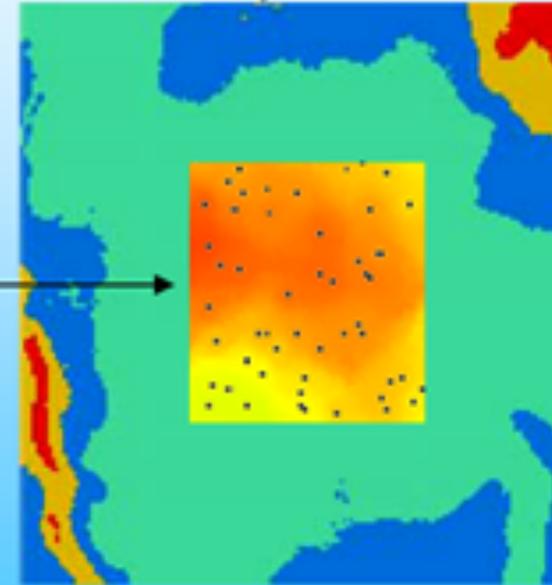
Original surface with sample points



Interpolated surface



Error map and extract



# Geostatistica: flusso di lavoro

Campionamento  
(progettazione, acquisizione dati)

**Analisi statistica (explanatory data analysis):** Analizza le proprietà spaziali e statistiche dei dati esame distribuzione, identificazione ed eliminazione trend, identificazione anisotropie, ecc.

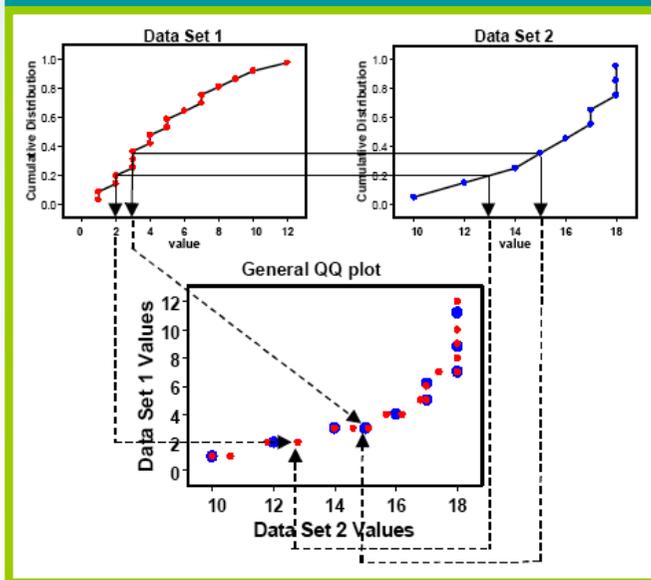
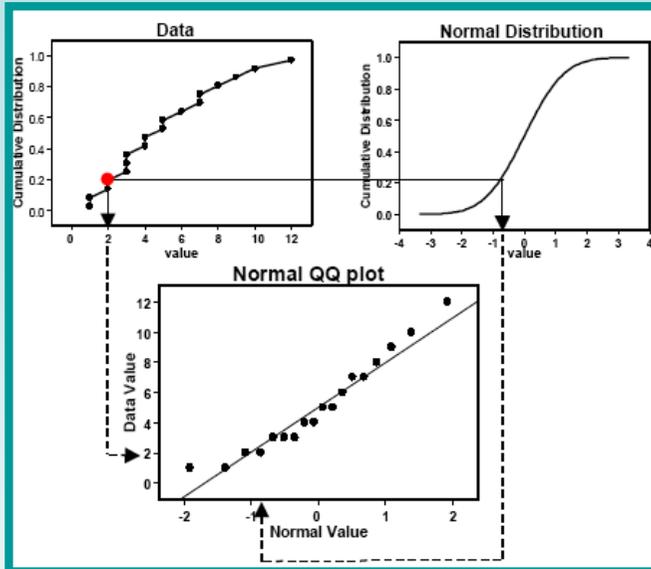
**Stima del variogramma sperimentale e scelta del modello da utilizzare.**

Interpolazione **Calcolo superficie statistica** di predizione e **Carta degli errori**. La predizione dei valori incogniti avviene attraverso il modello del variogramma, la configurazione spaziale dei dati e il valore dei punti misurati.

**Diagnostica:** comprendere “quanto bene” il modello predice i valori incogniti. Analisi di errore: cross validation, validation points, ecc.

# Explanatory Data Analysis

Analizza le proprietà spaziali e statistiche dei dati: esame distribuzione, ricerca outliers, ricerca trend e anisotropie, esame autocorrelazione spaziale, ecc.



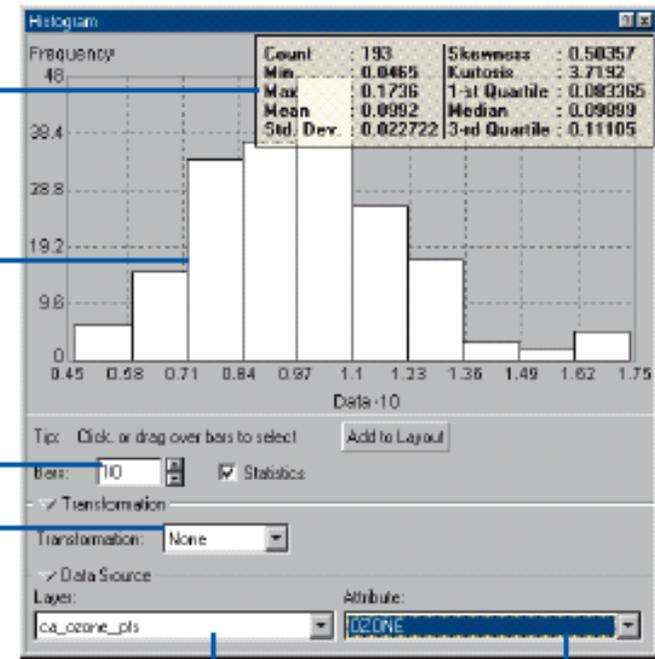
## Istogrammi

Distribution statistics

Data distribution

Number of histogram bars

Transformation selection



Selected layer

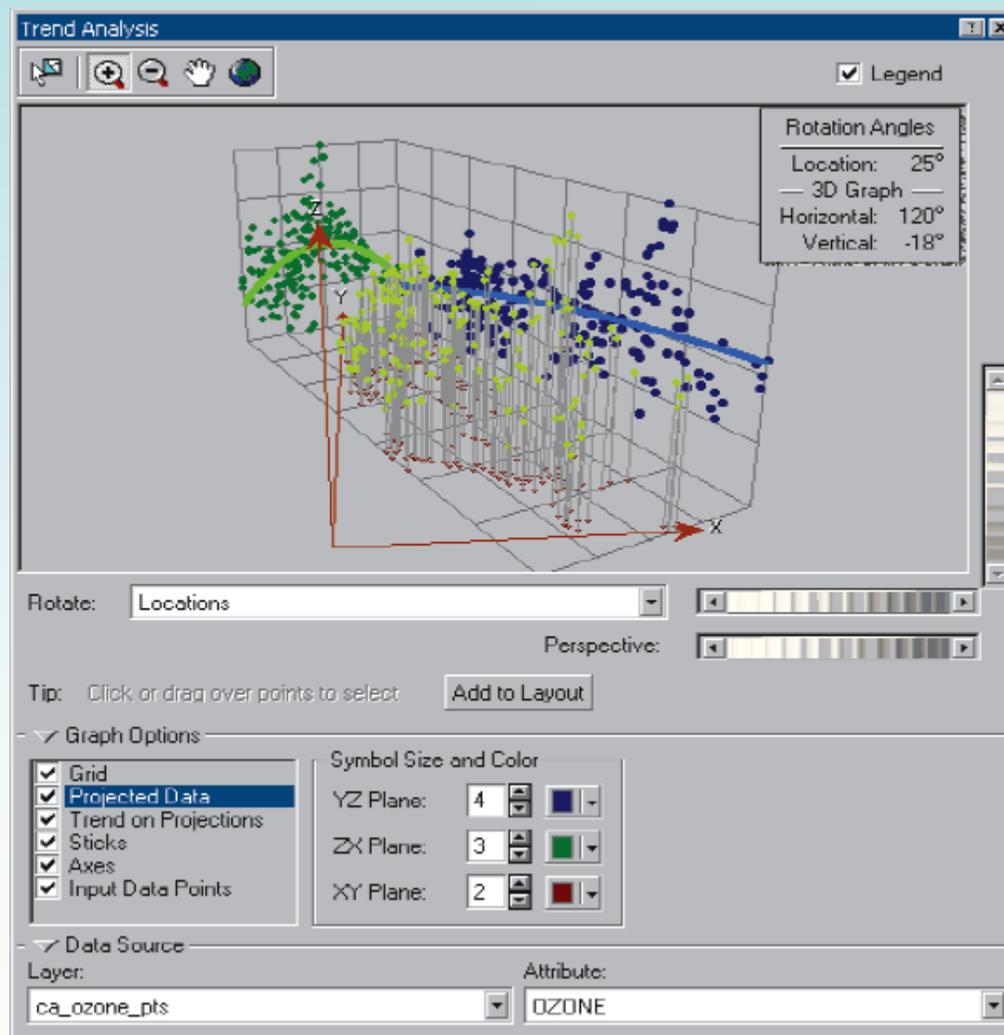
Selected attribute

QQPlot

# Explanatory Data Analysis

Analizza le proprietà spaziali e statistiche dei dati: esame distribuzione, ricerca outliers, ricerca trend e anisotropie, esame autocorrelazione spaziale, ecc.

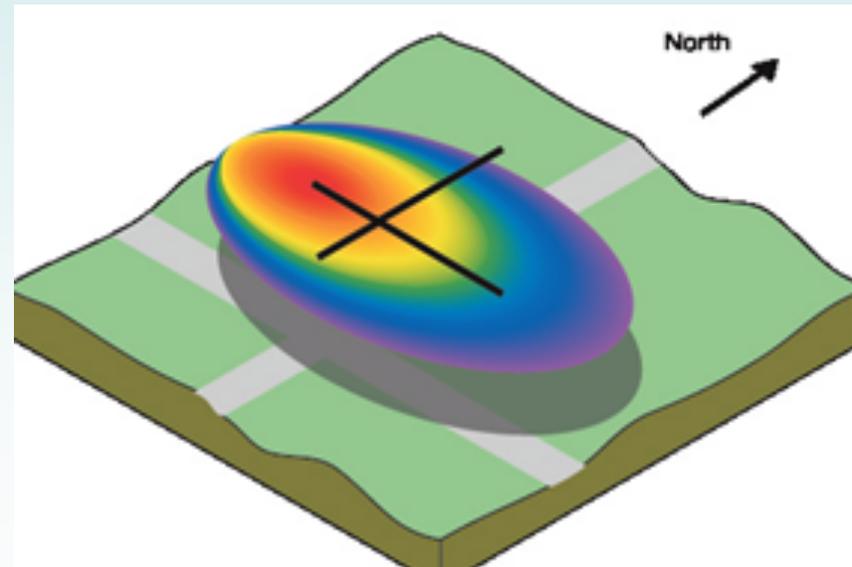
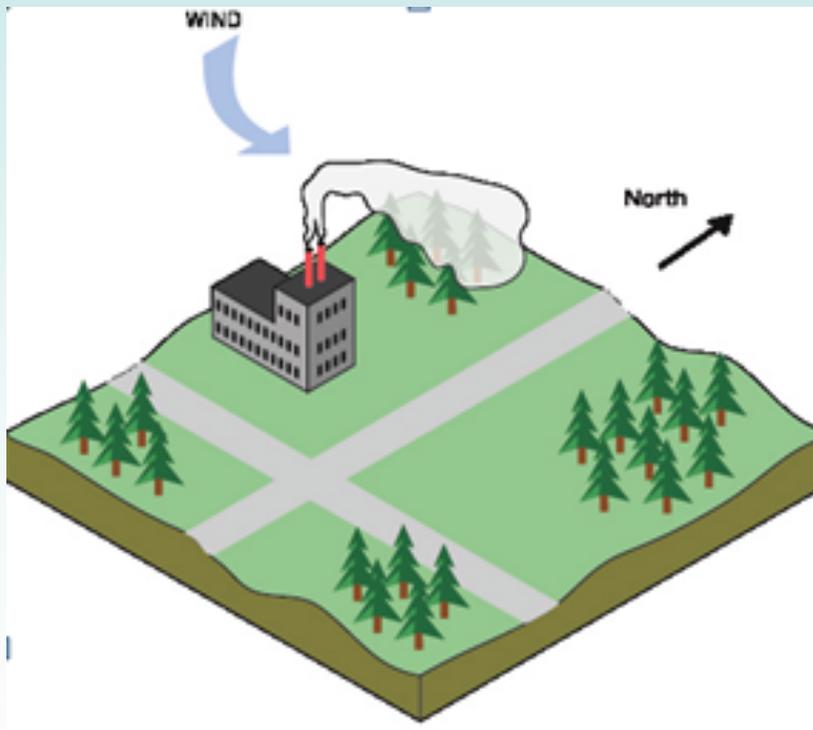
## Analisi Trend



# Explanatory Data Analysis

## Anisotropia

Spesso si nota la presenza di anisotropia (l'autocorrelazione dipende dalla direzione nei dati). Esempi: inquinanti atmosferici nella direzione del vento prevalente, flussi idrici sotterranei o superficiali, ecc.



# Geostatistica: flusso di lavoro

Campionamento  
(progettazione, acquisizione dati)

**Analisi statistica (explanatory data analysis):** Analizza le proprietà spaziali e statistiche dei dati esame distribuzione, identificazione ed eliminazione trend, identificazione anisotropie, ecc.

**Stima del variogramma sperimentale e scelta del modello da utilizzare.**

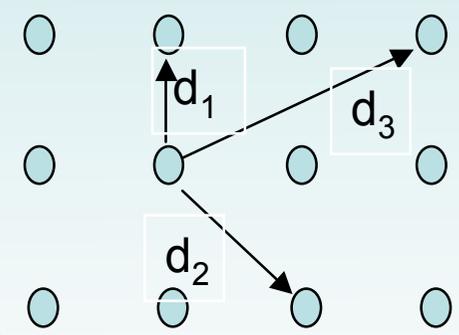
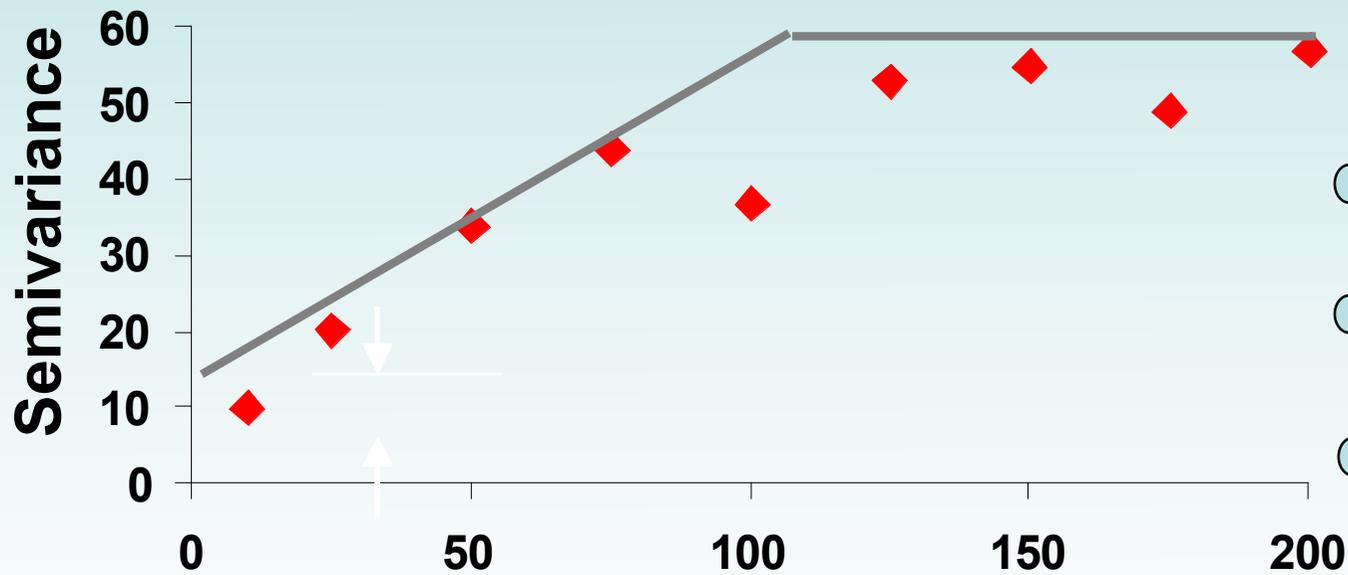
Interpolazione **Calcolo superficie statistica** di predizione e **Carta degli errori**. La predizione dei valori incogniti avviene attraverso il modello del variogramma, la configurazione spaziale dei dati e il valore dei punti misurati.

**Diagnostica:** comprendere “quanto bene” il modello predice i valori incogniti. Analisi di errore: cross validation, validation points, ecc.

# Semivariogramma

Il **semivariogramma** è una tecnica di analisi statistica che quantifica il grado di variabilità di punti a distanza crescente. Rappresenta la distanza tra i dati di input (**lag**) sull'asse orizzontale e la **semivarianza** sull'asse verticale.

La **semivarianza** è la metà della varianza (quadrato della deviazione standard) tra ogni valore  $Z$  e ognuno dei punti alla stessa distanza; **è una misura della interdipendenza dei valori di  $Z$  basata su quanto vicini essi sono (una misura del grado di dipendenza spaziale tra i campioni)**



$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n \{z(x_i) - z(x_i + h)\}^2$$

**Lag (m)**

- $\hat{\gamma}$ : semivariance
- $z$ : value of attribute  $z$
- $x$ : geographic location ( $i = 1, 2, 3, \dots, n$ )
- $n$ : count of pairs of points separated by the lag  $h$

(Chiles and Delfiner, 1999)

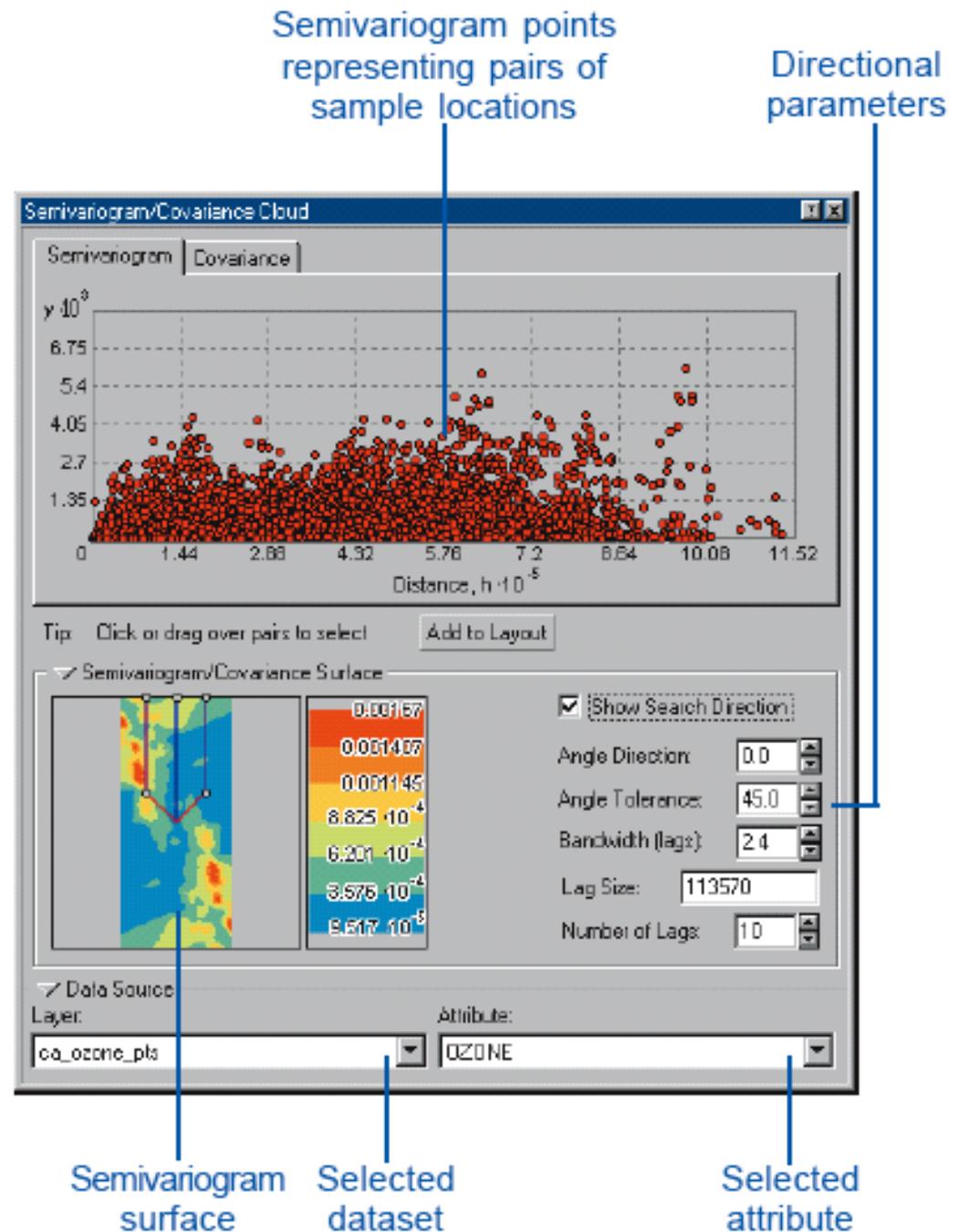
Ogni distanza tra i punti è chiamata lag (m) La distanza è calcolata tra tutte le possibili coppie di punti

## Nuvola del variogramma (variogram cloud)

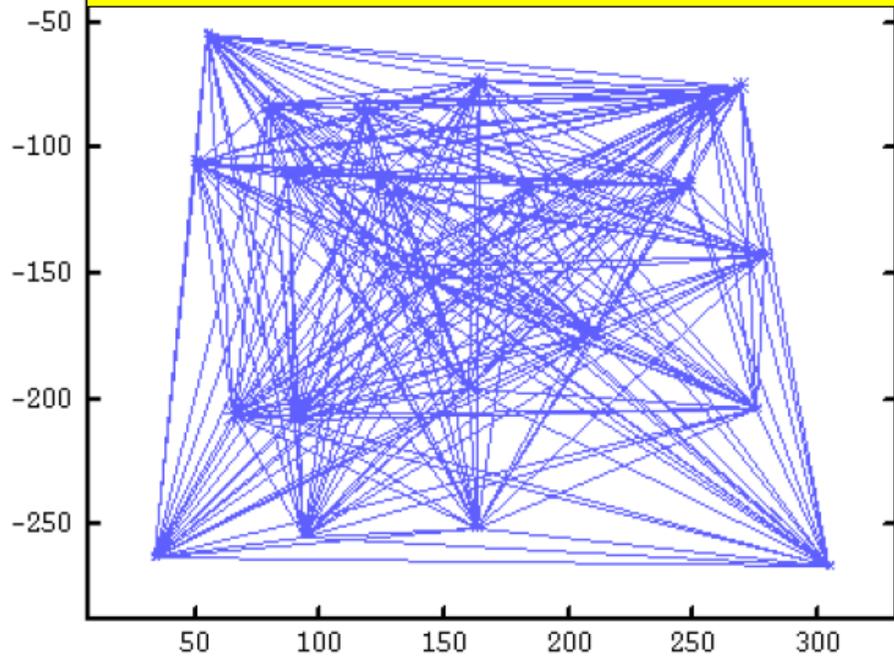
La nuvola si ottiene graficizzando le semivarianze di ogni singola coppia con i rispettivi lag.

La nube di punti contiene quindi tutte le relazioni spaziali nei dati per tutte le possibili distanze tra i campioni, ma **non è una funzione continua**.

E' praticamente impossibile interpretarla e comprendere l'esistenza di correlazioni spaziali. Si rende necessario **prendere in considerazione un numero ristretto di lag**. In questo modo sono più riconoscibili gli outlier ed è possibile modellizzare la distribuzione spaziale.

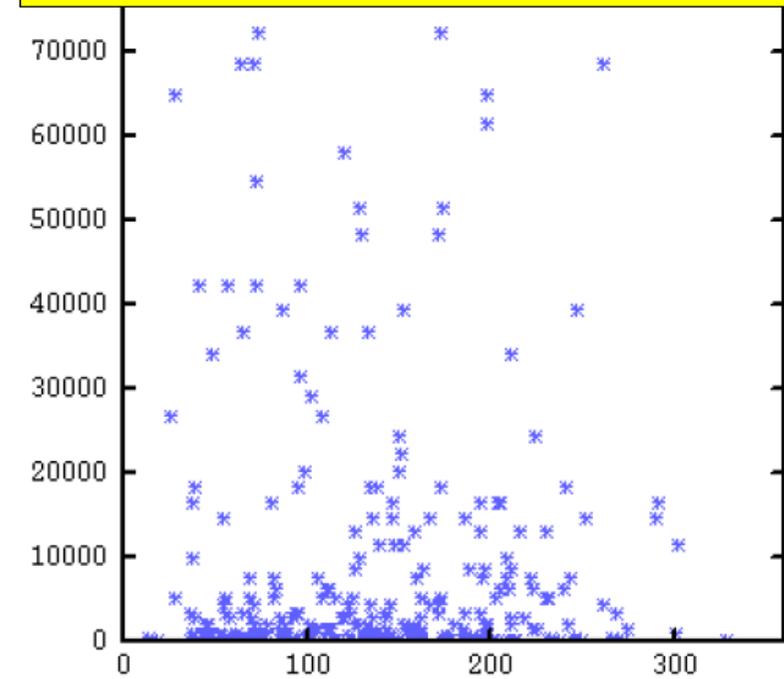


tutte le possibili coppie di campioni

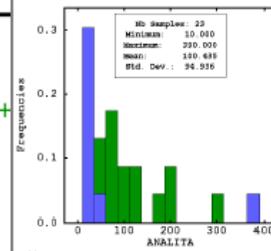
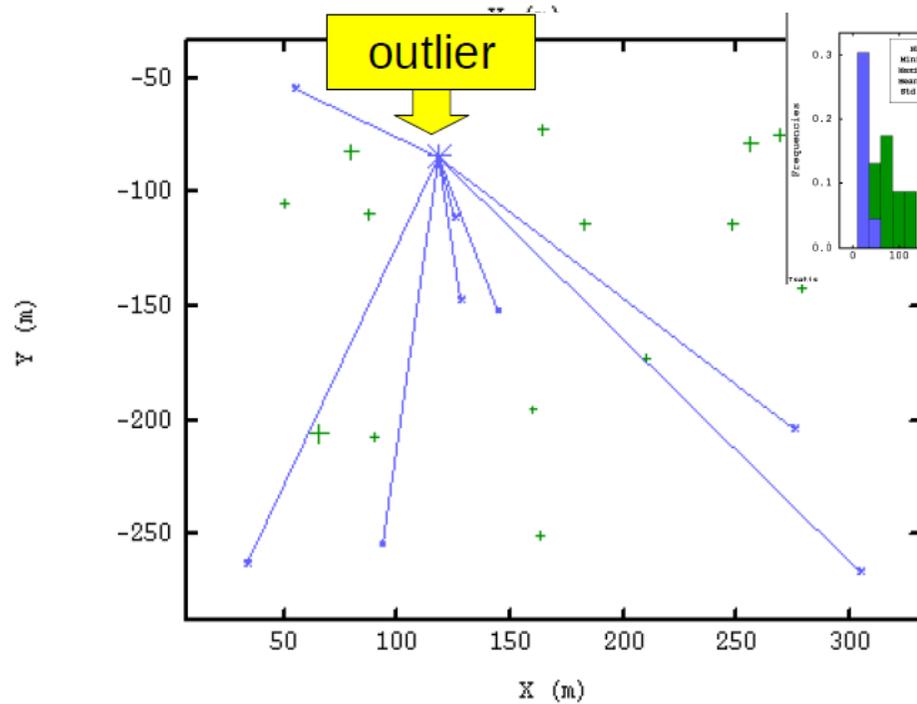


tutti i possibili valori di semivarianza

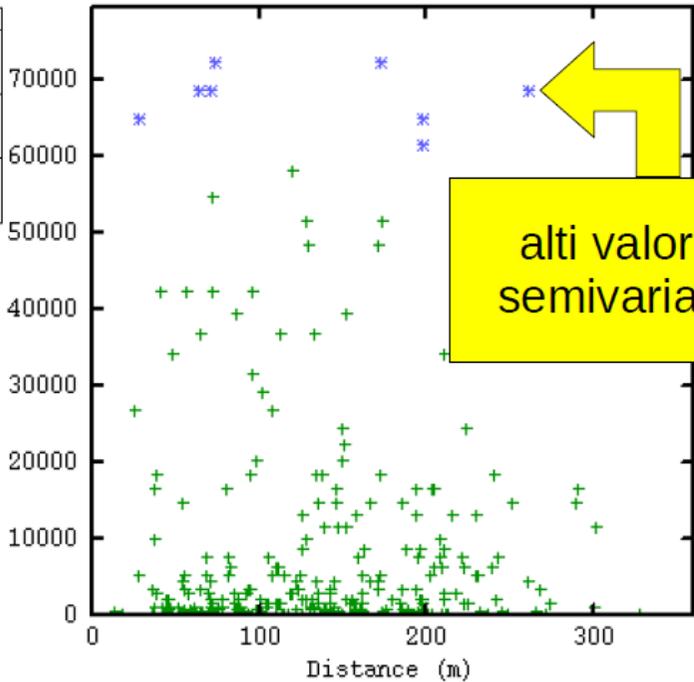
Variogram : ANALITA



outlier



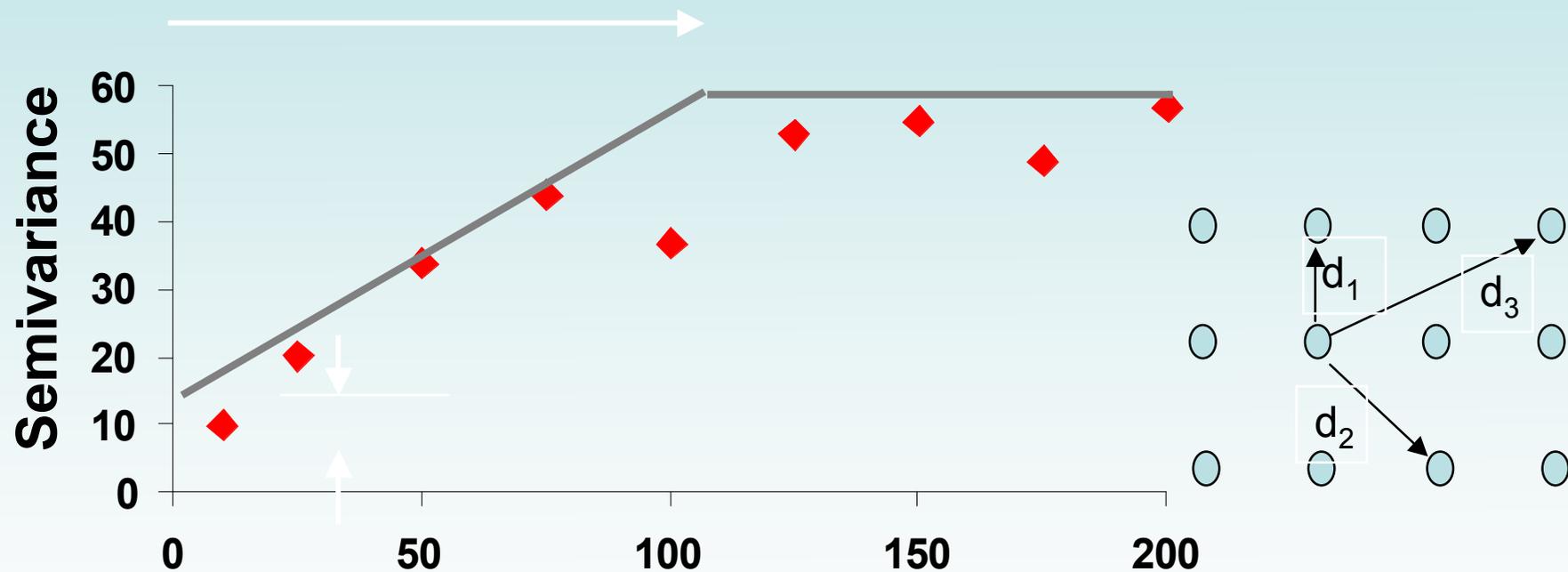
Variogram : Al



alti valori di semivarianza

# Semivariogramma sperimentale

La nuvola del variogramma viene trasformata in variogramma sperimentale, in cui si graficizzano le differenze medie nei valori di varianza a determinate distanze (lag prefissati).



$$\hat{\gamma}(h) = \frac{1}{2n} \sum_{i=1}^n \{z(x_i) - z(x_i + h)\}^2$$

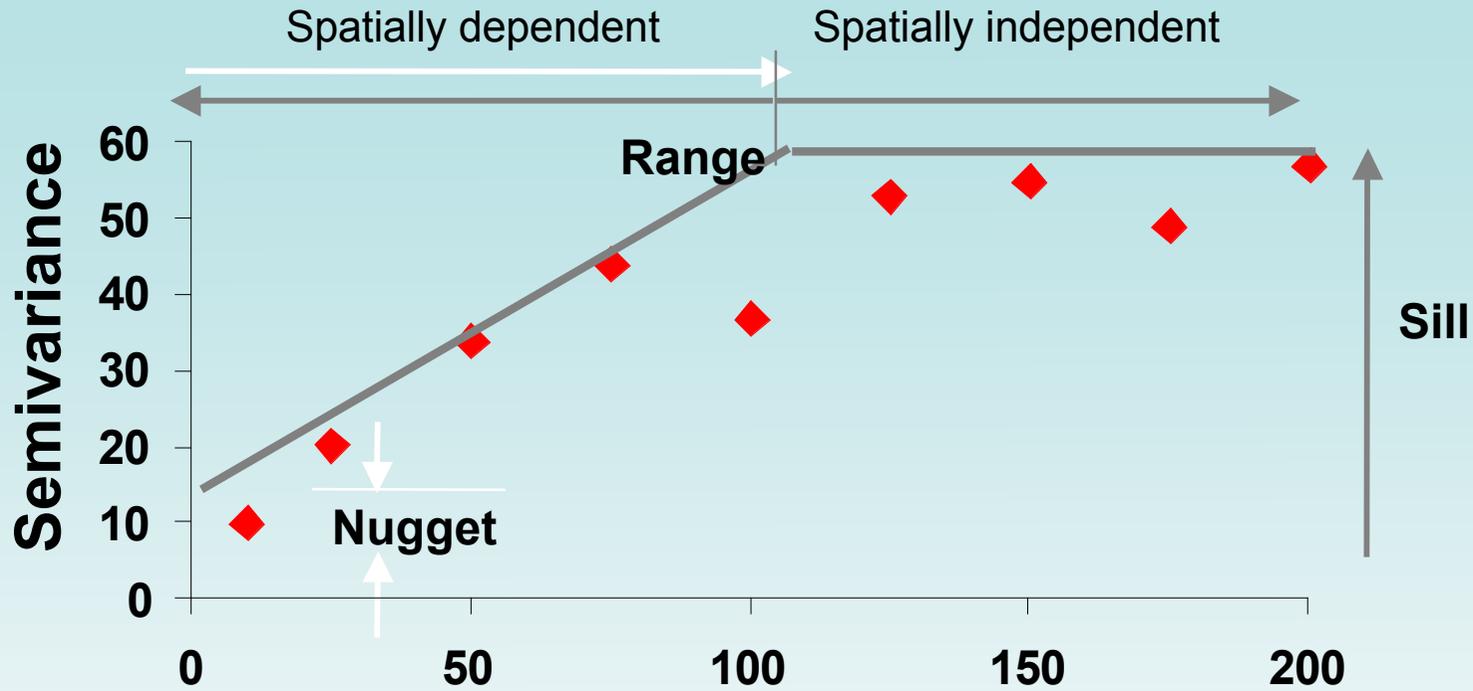
**Lag (m)**

- $\hat{\gamma}$ : semivariance
- $z$ : value of attribute  $z$
- $x$ : geographic location ( $i = 1, 2, 3, \dots, n$ )
- $n$ : count of pairs of points separated by the lag  $h$

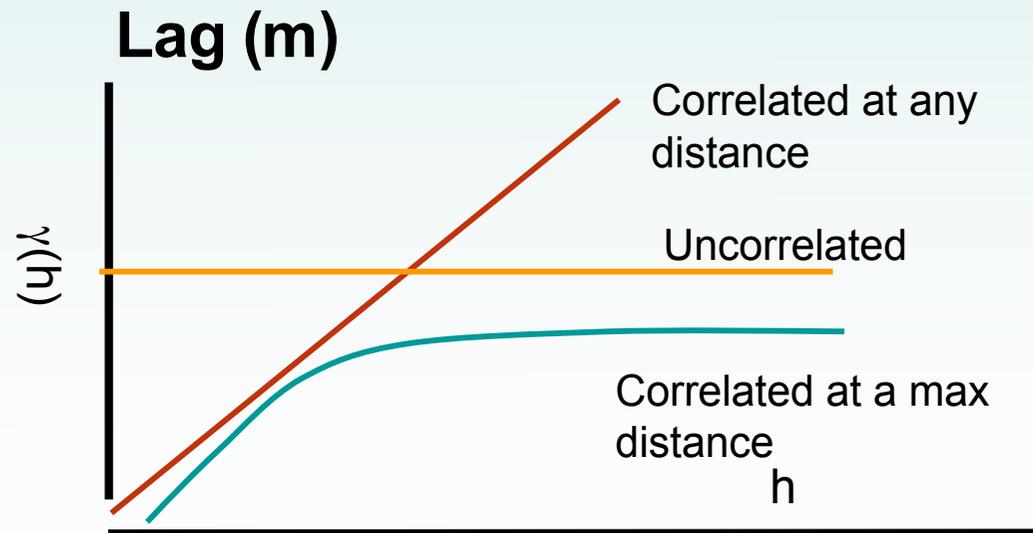
(Chiles and Delfiner, 1999)

Ogni distanza tra i punti è chiamata lag (m) La distanza è calcolata tra tutte le possibili coppie di punti

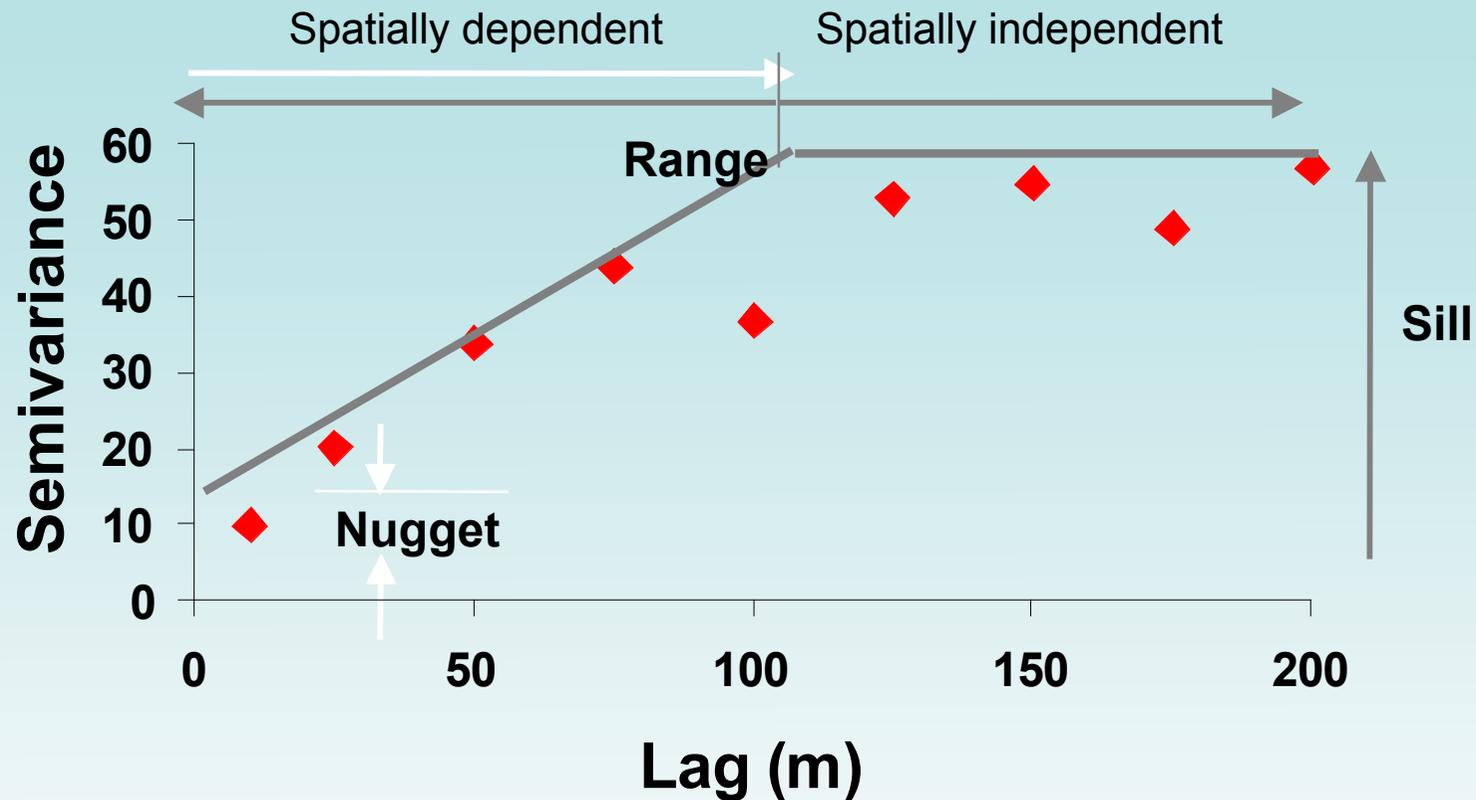
# Semivariogramma



La parte più importante del variogramma è la sua forma nei pressi dell'origine poichè ai punti più vicini verrà dato un peso maggiore durante l'interpolazione



# Semivariogramma



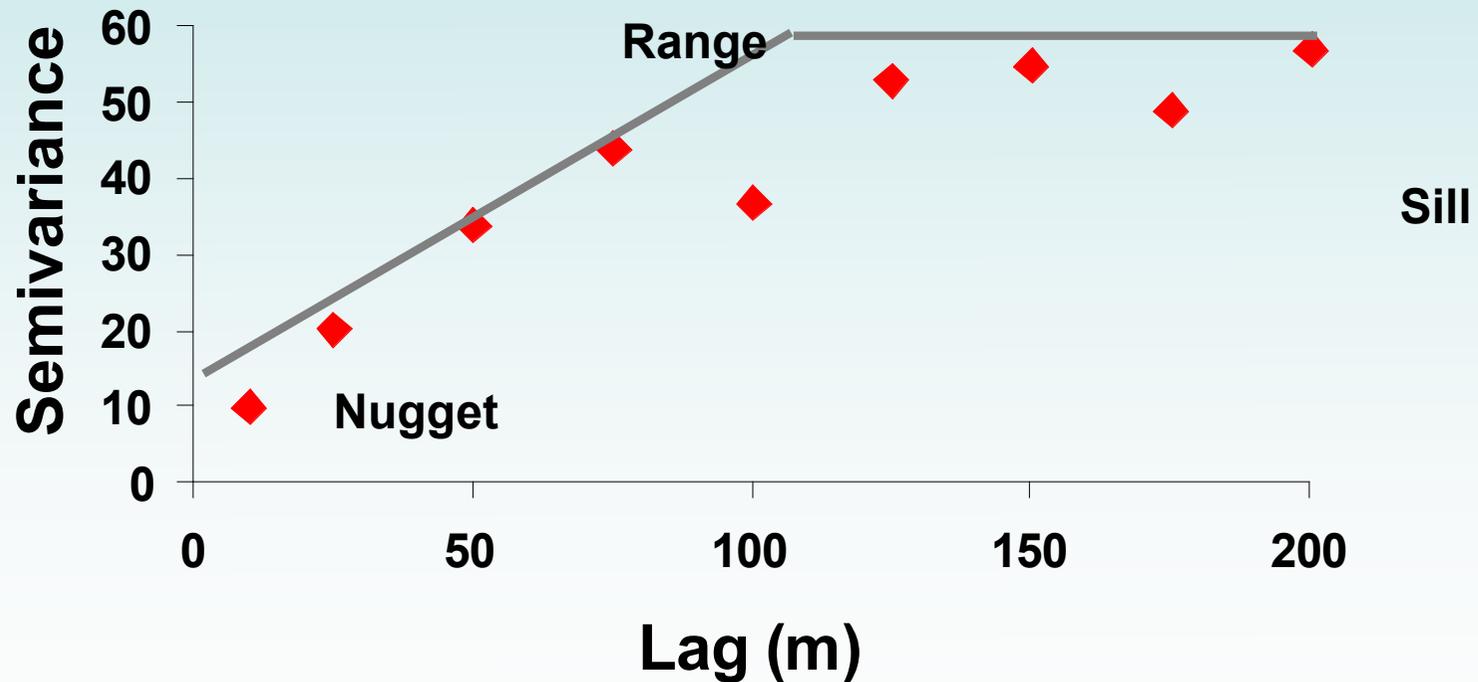
- **Nugget Variance** – E' una stima del residuo, **rumore** spazialmente non correlato. Combina le variazioni residue degli errori di misurazione con le variazioni spaziali che si verificano su distanze più brevi del passo di campionamento.

# La scelta del lag

La scelta del lag influenza fortemente il variogramma così questa distanza va definita con cura.

Lag troppo piccoli = troppe semivarianze medie (una per ogni lag) e alta probabilità di variogramma inesatto.

Lag troppo grandi = poche stime di varianza, perdita di dettaglio ed eccessivo smoothing del variogramma.

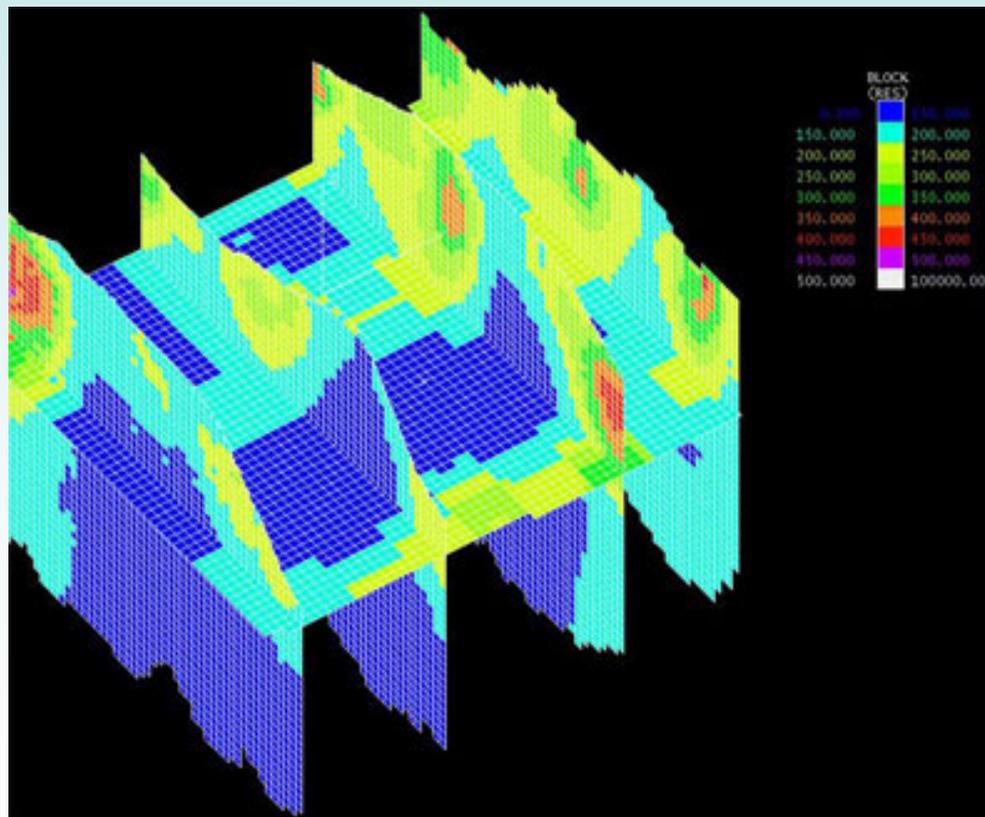


# Direzionalità del variogramma

**Il variogramma può essere calcolato in diverse direzioni dello spazio.**

In molti software l'utente può definire una direzionalità di variogramma in base alla disposizione di determinati parametri ambientali (es. direzione di faglie geologiche, reticolo idrografico, ecc.)

Sfortunatamente i dati ambientali raramente sono disponibili in maglia quadrata o rettangolare o cmq allineata.



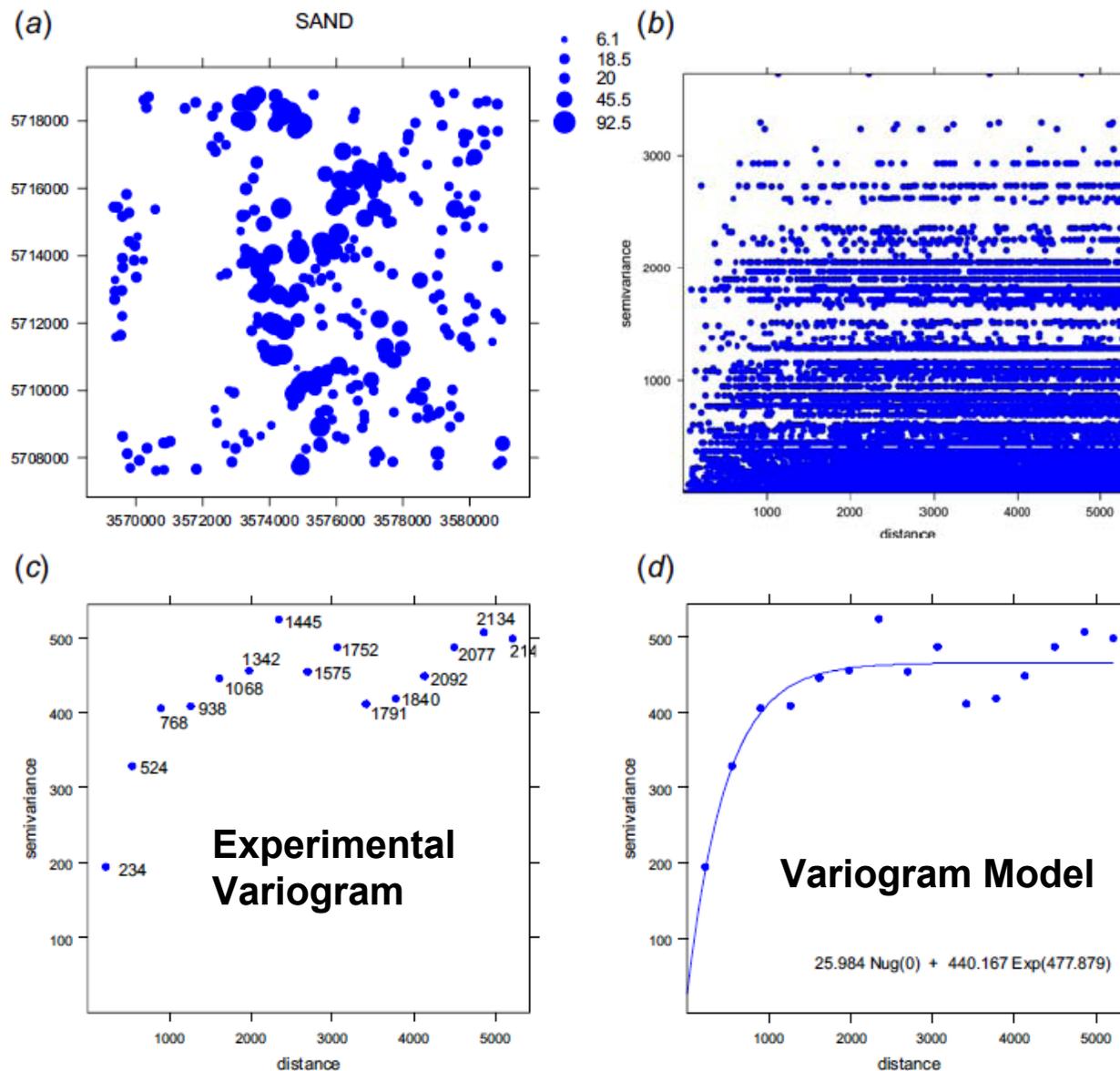
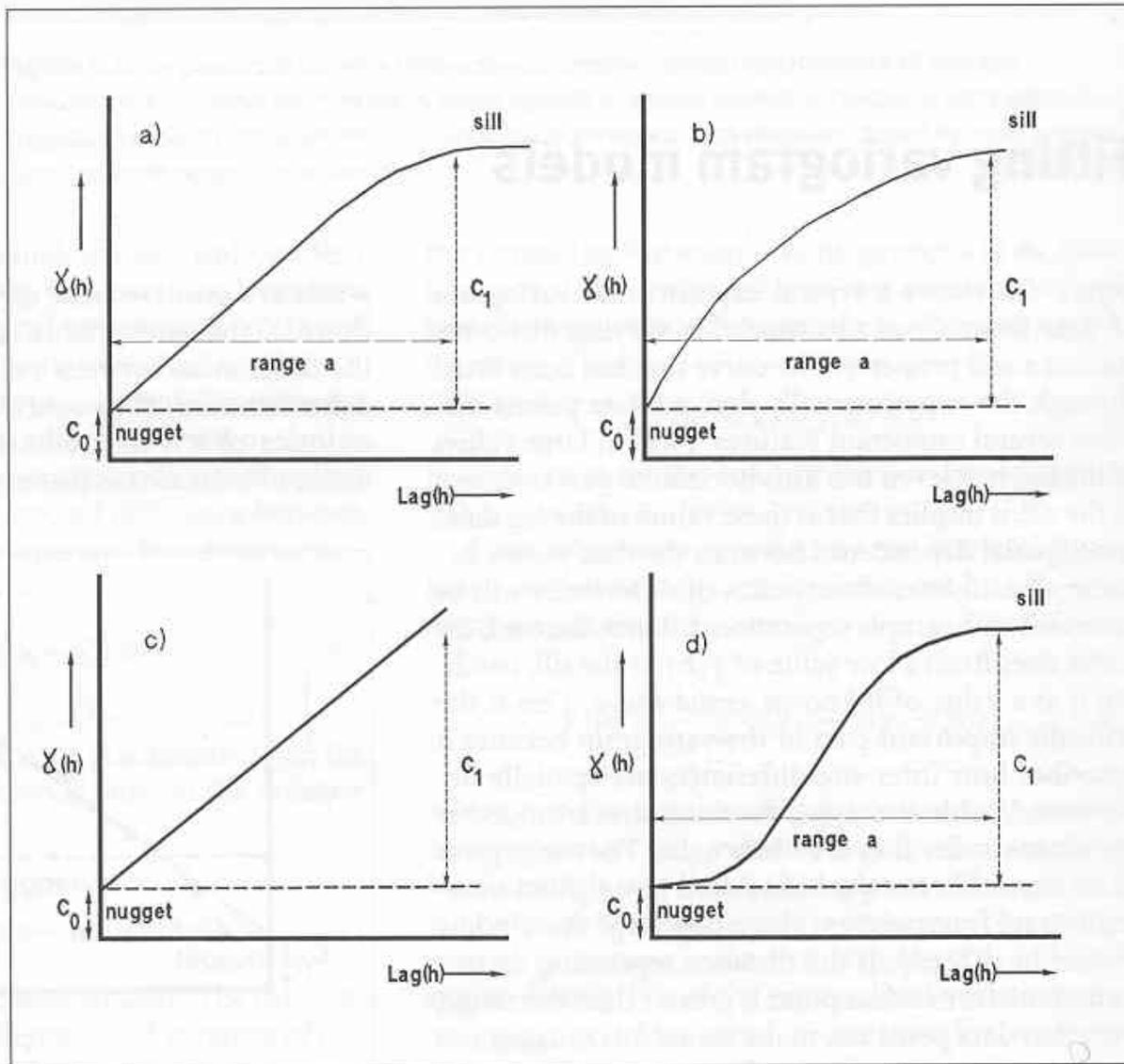


Fig. 1.7: Steps of variogram modelling: (a) location of points (300), (b) variogram cloud showing semivariances for 44850 pairs, (c) semivariances aggregated to lags of about 300 m, and (d) the final variogram model fitted using the default settings in *gstat*.

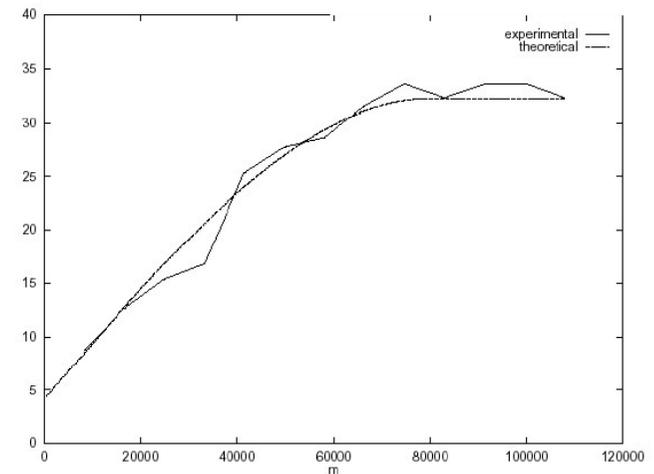
# Modelli di semivariogramma



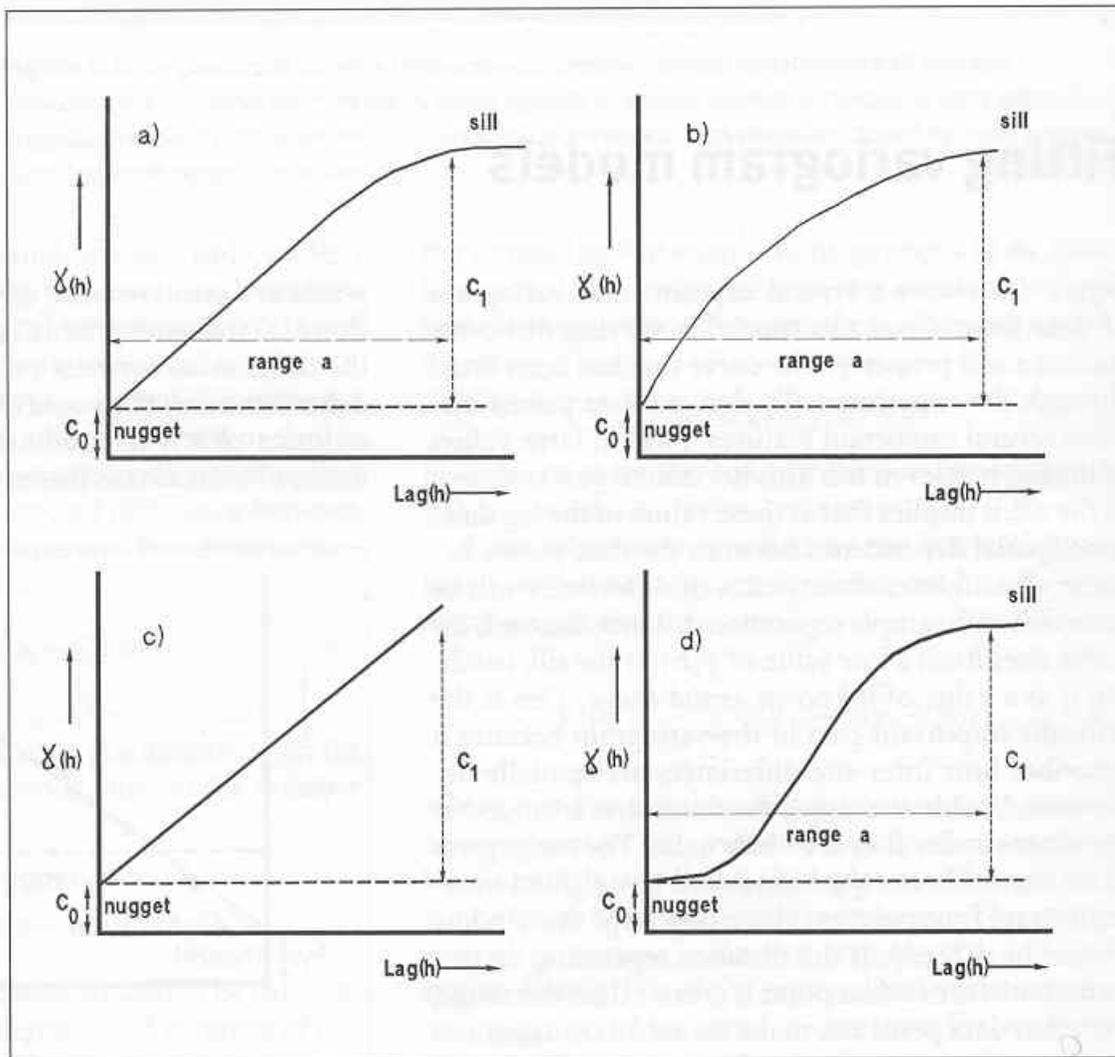
**Figure 6.3.** Examples of the most commonly used variogram models: (a) spherical; (b) exponential; (c) linear; and (d) Gaussian

Viene adattato un modello di variogramma al variogramma sperimentale.

La scelta del modello rimane un processo controverso e discusso nella geostatistica



# Modelli di semivariogramma



**Figure 6.3.** Examples of the most commonly used variogram models: (a) spherical; (b) exponential; (c) linear; and (d) Gaussian

## Sferico:

Si usa se il nugget è presente ma piccolo. E' il modello più utilizzato come default in molti software.

## Esponenziale:

Si usa se il nugget è rilevante e la crescita verso il sill è poco ripida.

# Geostatistica: flusso di lavoro

Campionamento  
(progettazione, acquisizione dati)

**Analisi statistica (explanatory data analysis):** Analizza le proprietà spaziali e statistiche dei dati esame distribuzione, identificazione ed eliminazione trend, identificazione anisotropie, ecc.

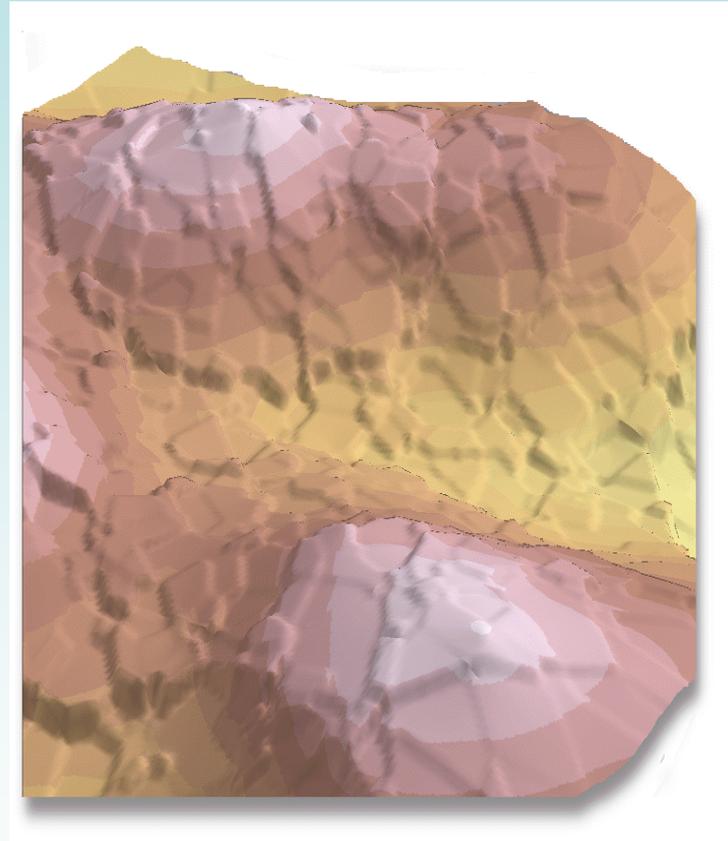
**Stima del variogramma sperimentale e scelta del modello da utilizzare.**

Interpolazione **Calcolo superficie statistica** di predizione e **Carta degli errori**. La predizione dei valori incogniti avviene attraverso il modello del variogramma, la configurazione spaziale dei dati e il valore dei punti misurati.

**Diagnostica:** comprendere “quanto bene” il modello predice i valori incogniti. Analisi di errore: cross validation, validation points, ecc.

# Kriging

- **Kriging** è una famiglia di procedure geostatistiche che assumono la presenza di una correlazione spaziale tra i valori misurati. Questa correlazione è stimata attraverso la costruzione dei **variogrammi**.
- **Di fatto la parola “kriging” è considerata un termine generico sinonimo di geostatistica**
- Fornisce non solo i valori interpolati ma anche una **stima della quantità di errore** potenziale dell'output.
- Richiede una **elevata interazione** da parte dell'operatore per poter essere usato correttamente
- Il nome viene da D. B. **Krige**, un ingegnere minerario sudafricano che ha definito il metodo (1951) insieme a H. S. Sichel come strumento di indagine dei giacimenti minerari. 10 anni più tardi, prendendo spunto dal loro lavoro, G. Matheron formalizzò la Teoria delle V.R.

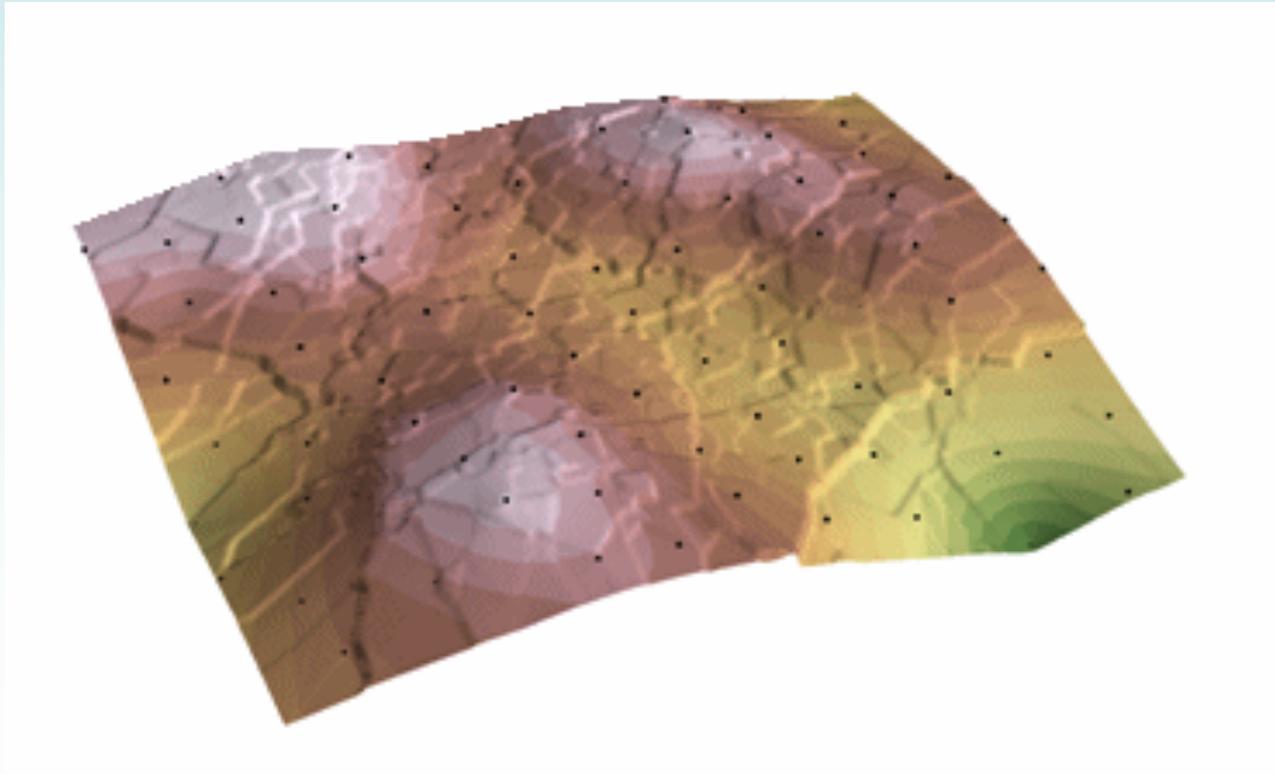


# Kriging

La versione standard di Kriging è chiamata Ordinary Kriging (OK)

Può essere visto come un Inverse Distance Weighted molto più sofisticato: **i pesi anzichè determinati in modo arbitrario, riflettono la struttura di autocorrelazione spaziale definita dalla funzione di variogramma.**

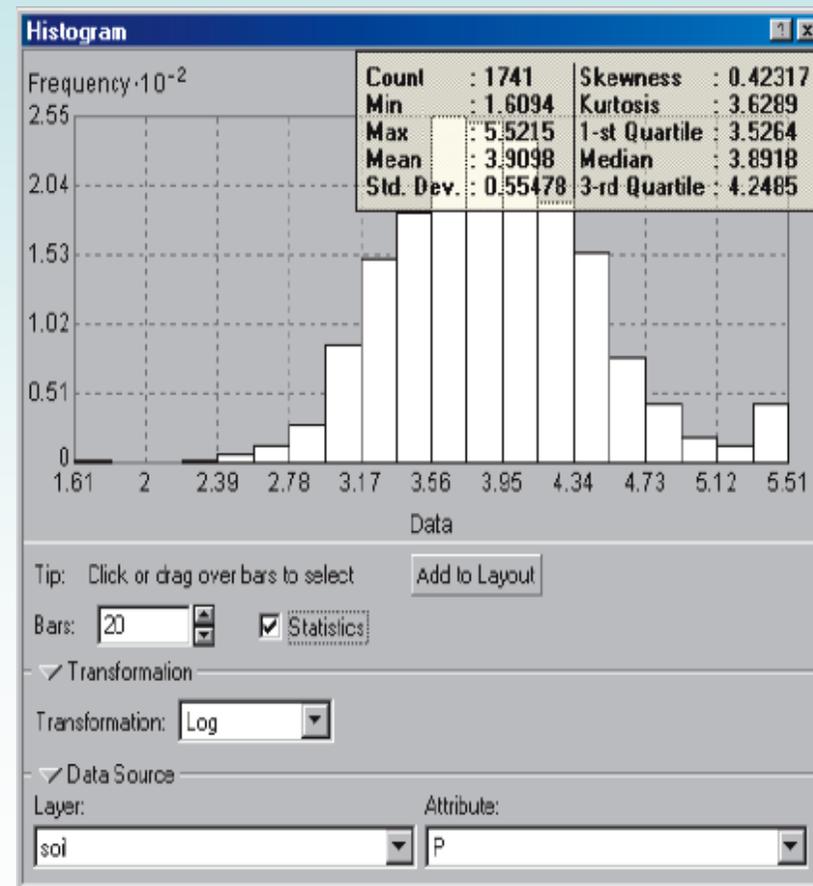
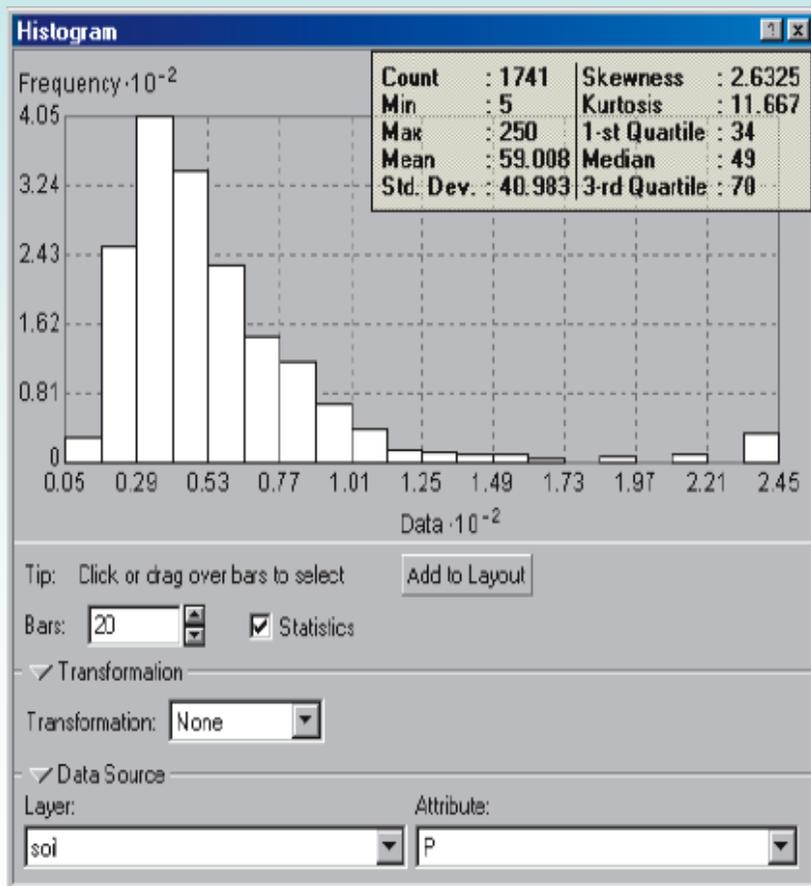
Una volta definito il modello di variogramma, possiamo usarlo per derivare le semivarianze in tutte le posizioni e risolvere i pesi di Kriging.



# Alcuni tipi di Kriging richiedono particolari condizioni di utilizzo

## Distribuzione normale dei dati

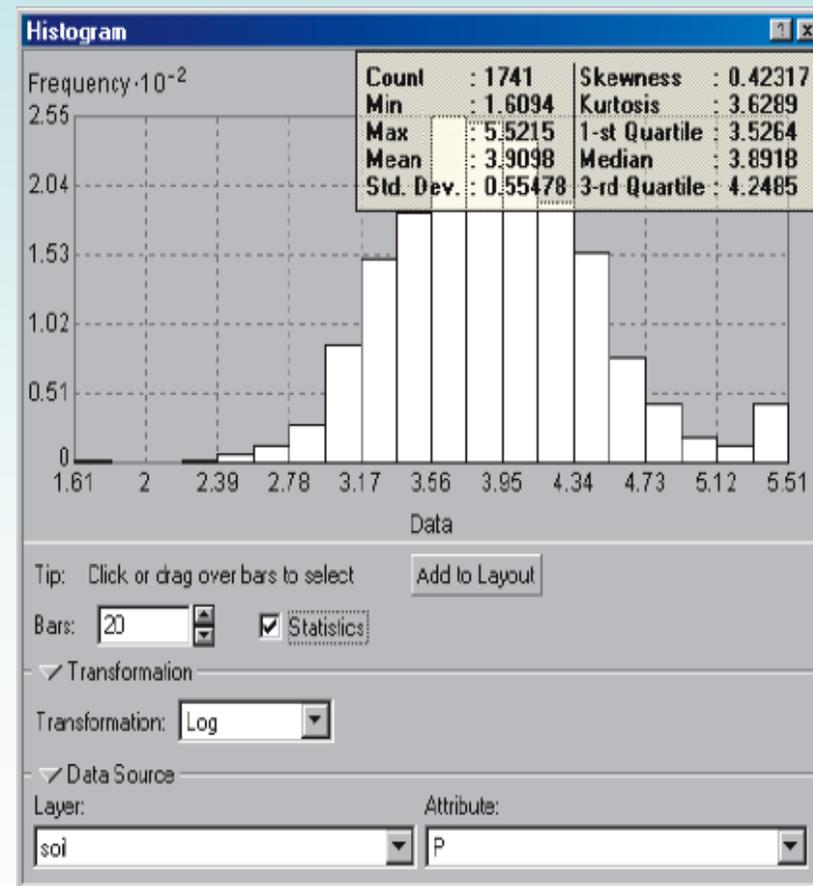
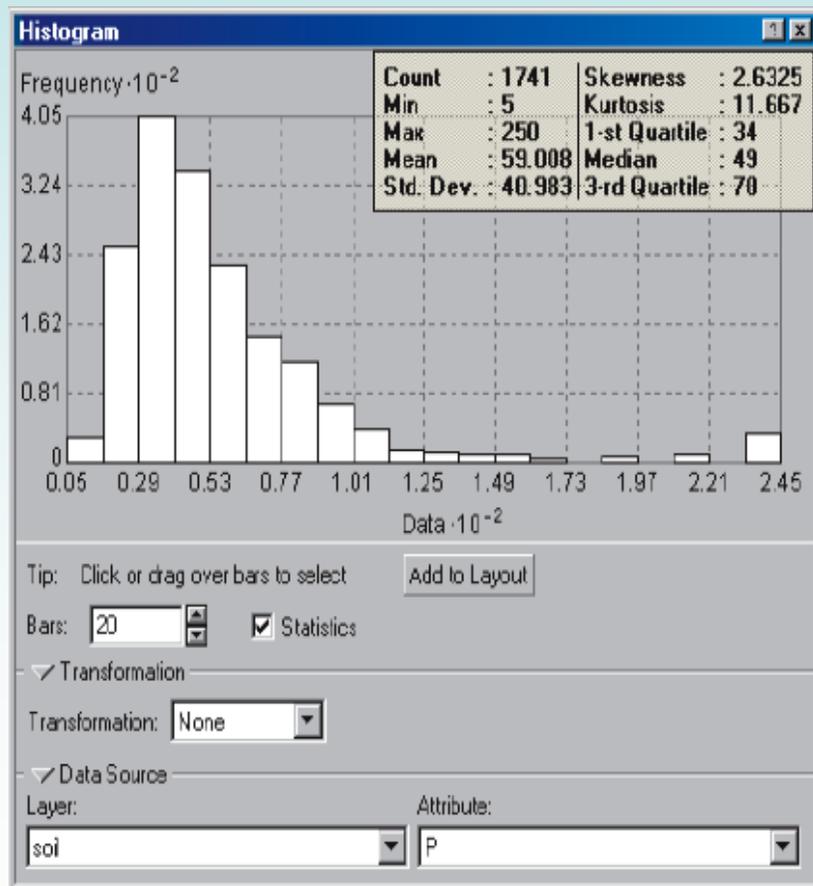
In molti casi i dati sono marcatamente asimmetrici e non-normali. Il variogramma è sensibile alle asimmetrie positive; spesso queste asimmetrie possono essere rimosse con stabilizzazione delle varianze. Se la trasformazione logaritmica porta ad una distribuzione quasi-normale si parla di **lognormal kriging**



# Alcuni tipi di Kriging richiedono particolari condizioni di utilizzo

## Distribuzione normale dei dati

In molte situazioni geologiche ed ambientali le distribuzioni tendono ad essere asimmetriche positivamente. Spesso la **trasformazione in logaritmi** dei valori originali porta ad una distribuzione normale (coincidenza di media, moda e mediana; bassi valori di skewness ed errore standard).



Lez02 (PROTETTO) - Adobe Reader

File Modifica Vista Finestra ?

16 / 43 120%

Commento Condividi

Università degli Studi di Siena

Corso di Laurea Specialistica in Geologia Applicata - Master Il Livello in Geotecnologie Ambientali (MSGa) - Professional Course in Geostatistica Ambientale  
A.A. 2008/2009

**GEOSTATISTICA AMBIENTALE**

CGT / Centro di GeoTecnologie  
Via Verrì Vecchi, 34 - 53007, San Giovanni Valdarno (AR)

1. Statistica univariata (1.5. Distribuzioni di probabilità)

## 1.5.2. La distribuzione lognormale

- ▶ Esistono vari metodi per testare la bontà del fitting del modello lognormale sulla distribuzione dei valori campionati,
- ▶ uno dei quali è per via grafica, utilizzando un grafico con un asse logaritmico e l'altro di probabilità, chiamato **log-probability plot**.
- ▶ Questo grafico funziona in modo tale che la curva di frequenza cumulativa dei dati plottati viene disegnata come una linea retta. Il metodo prevede le seguenti fasi:
  - ▶ Raggruppare i dati in appropriate classi;
  - ▶ Cumulare le classi per ottenere la freq cum.
  - ▶ Calcolare la frequenza cumulativa %;
  - ▶ Plottare i dati su scala logaritmica contro la frequenza cumulativa % sull'asse di probabilità.

**Teoricamente nessuna distribuzione campionaria segue un modello di distribuzione logaritmica a due parametri, dato che nessuna serie si dispone perfettamente lungo una retta.**

Tuttavia, va ricordato che si sta operando nello spazio campionario, e che che dobbiamo cercare è il modello che approssimi meglio la distribuzione.

Analizzando di nuovo le curve, tutte e tre potrebbero essere allineate rispettivamente su una retta per la maggior parte dei dati.

Soltanto le code della distribuzione si discostano leggermente dalla retta teorica.

Valori originali [grado %]

Frequenza cumulativa percentuale, su asse di probabilità

Legend: Al (red circles), Si (blue squares), Fe (green diamonds)

Lunedì 16 Marzo 2009

Lezione 2 - Modulo 1: Analisi Statistica

(Queste lezioni sono parte integrante del corso di Geostatistica Ambientale, e sono fruibili per uso interno al CGT, e secondo la licenza seguente: Copyright © Enrico Guastaldi 2009 - Licenza Creative Commons Attribuzione - Non Commerciale - Condividi allo stesso modo 2.5 Italia)

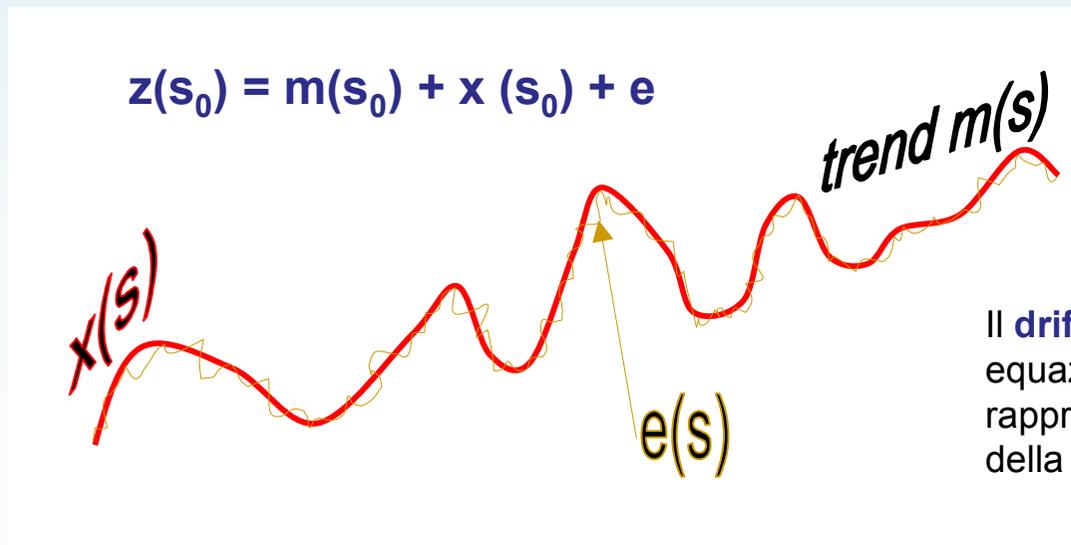
16

# Universal Model of Variation

Secondo Matheron il valore di una variabile può essere modellato come una somma di componenti deterministiche e stocastiche (MODELLO UNIVERSALE DELLA VARIAZIONE SPAZIALE).

3 variabili separate (valutate individualmente):

1. **Drift o struttura** (funzione deterministica che descrive la componente strutturale): tratta la superficie come un trend generale in una particolare direzione.
2. **Piccole variazioni** (funzione stocastica, **residuo spazialmente correlato**) rispetto al trend generale (piccoli dossi o depressioni) che sono distribuite in modo casuale ma sono correlate spazialmente.
3. **Random noise** che non è associato né al trend generale né ai dati autocorrelati spazialmente (errori di misura)



Il **drift** è stimato usando una equazione matematica che meglio rappresenta l'andamento generale della superficie (trend surface)

# Alcuni tipi di Kriging richiedono particolari condizioni di utilizzo

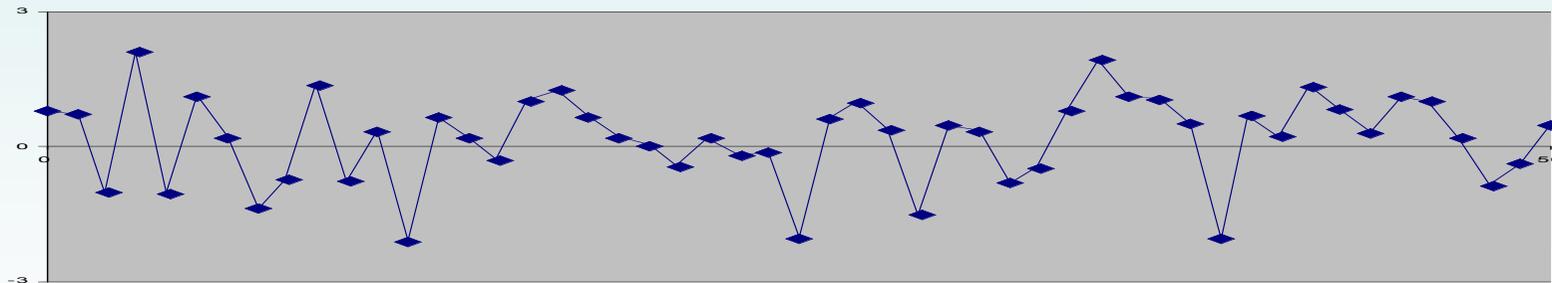
L'Ordinary Kriging assume la presenza di stazionarietà:

## Processo stazionario

é un processo in cui la funzione di densità di probabilità, la media e la varianza di una variabile casuale non cambiano nel tempo e nello spazio.

In pratica la distribuzione del processo casuale è indipendente dalla posizione. Se viene modificata la disposizione dei campioni in una determinata direzione dello spazio, i parametri di distribuzione non cambiano.

**Stationary data series : mean independent of position**



# Alcuni tipi di Kriging richiedono particolari condizioni di utilizzo

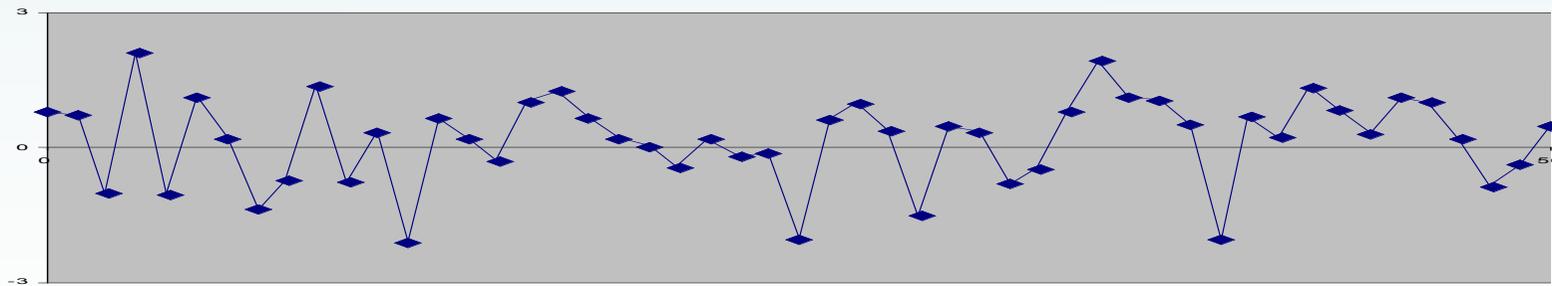
## Stazionarietà

La struttura spaziale della variabile è consistente sull'intero dataset

**1° ordine (in senso stretto):** la media è costante sull'intera regione e non dipende dalla posizione.

**2° ordine:** la varianza delle differenze tra due campioni è indipendente dalla posizione e dipende solamente dalla loro distanza

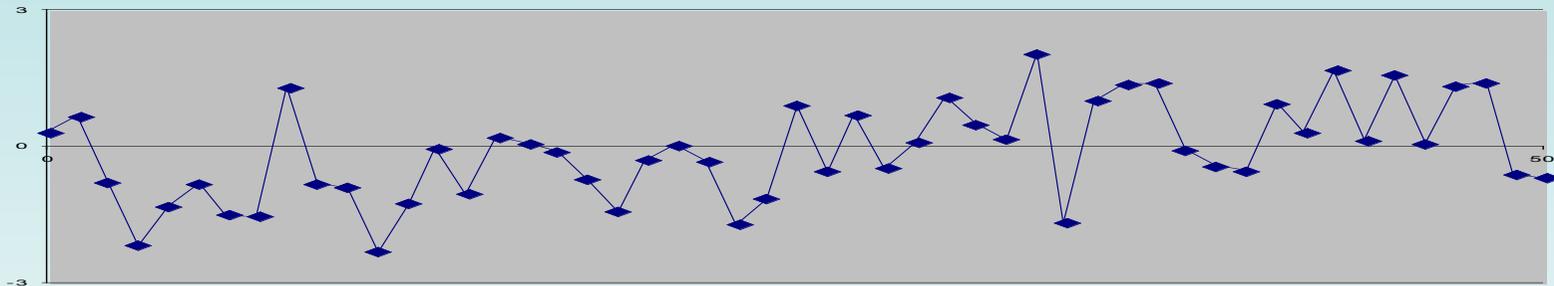
**Stationary data series : mean independent of position**



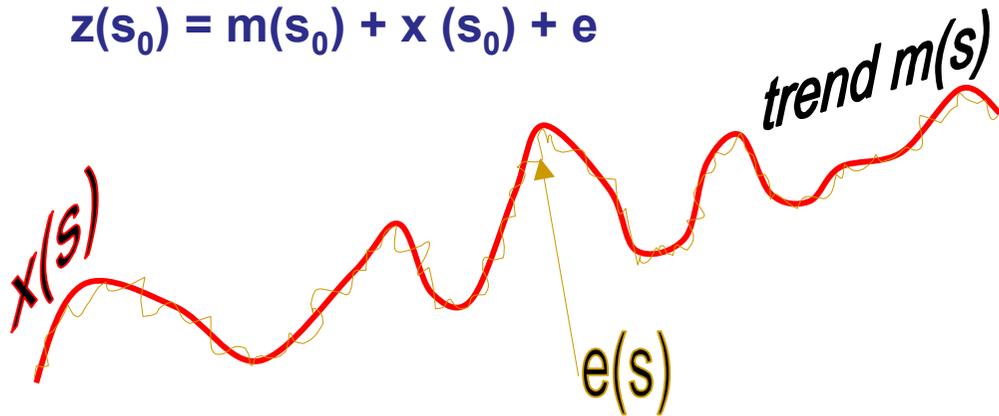
## Alcuni tipi di Kriging richiedono particolari condizioni di utilizzo

**In molti casi tuttavia i dati non sono stazionari a causa della presenza di importanti trends.**

Data series with trend: mean changes with position



$$z(s_0) = m(s_0) + x(s_0) + e$$



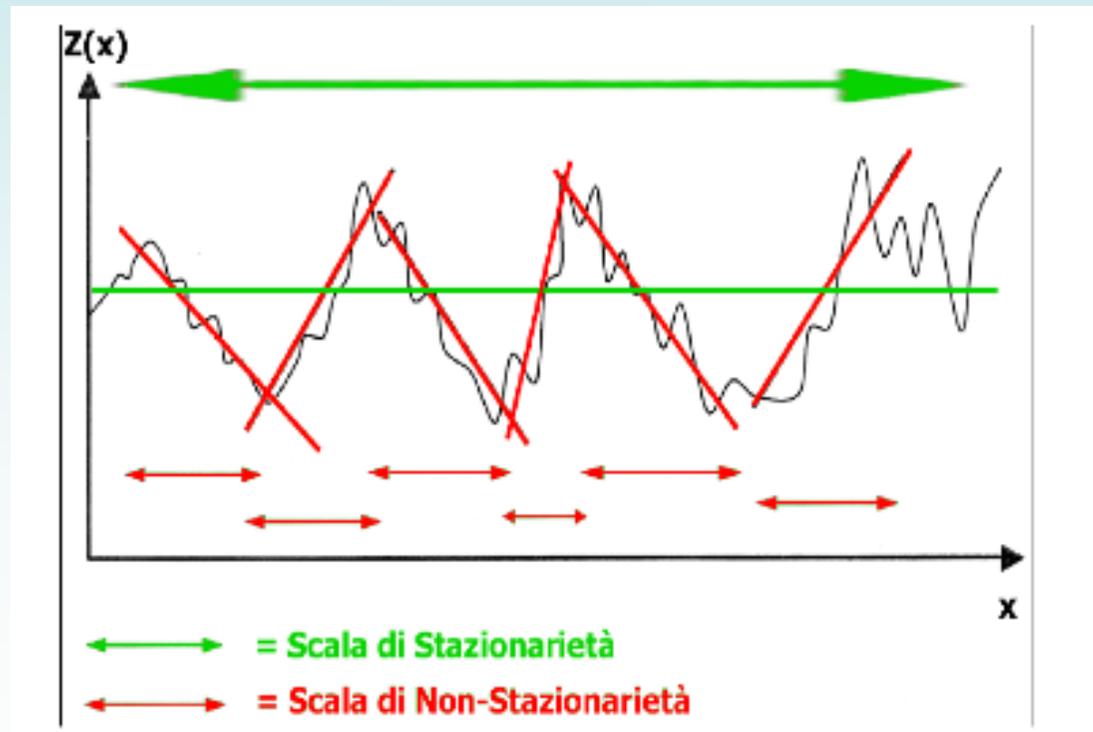
Se è presente un trend la media (valore atteso) non è costante ma è funzione della posizione (variogramma sperimentale concavo verso l'alto).

Il dato deve essere **“detrendizzato”** prima dell'analisi

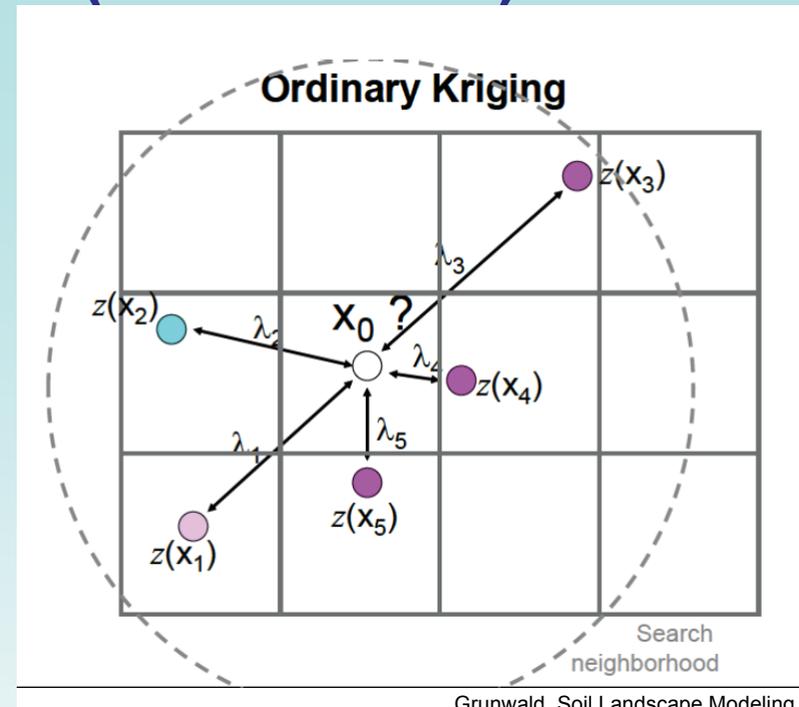
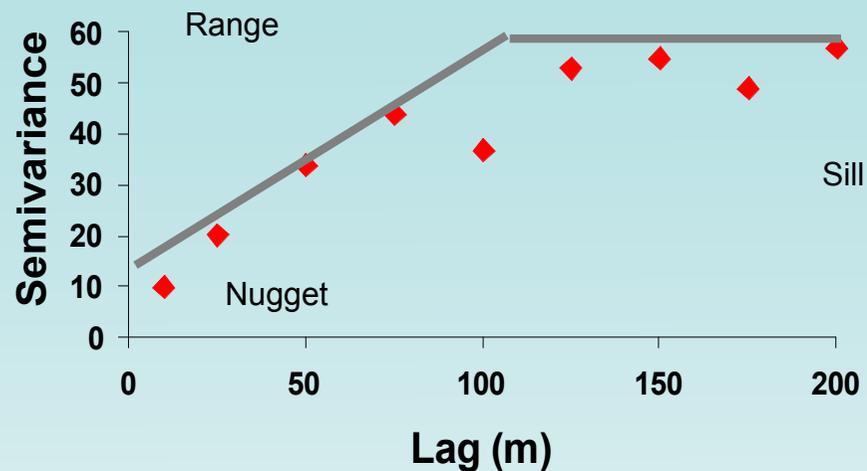
# Condizione di non-stazionarietà = drift

Il drift (condizione di non-stazionarietà) è un modello dipendente dalla scala.

Nella geostatistica la scala di lavoro (nominale) entra in gioco come fattore fondamentale nell'analisi delle correlazioni spaziali. Di solito è sempre possibile trovare una scala in cui il drift si annulla o diventa non importante.



# Kriging: attribuzione dei pesi (coefficienti)



Grunwald, Soil Landscape Modeling

$$\hat{Z}(x_o) = \sum_{i=1}^N \lambda_i * z(x_i) \quad \sum_{i=1}^N \lambda_i = 1$$

$\hat{Z}(x_o)$ : Estimate of the random variable

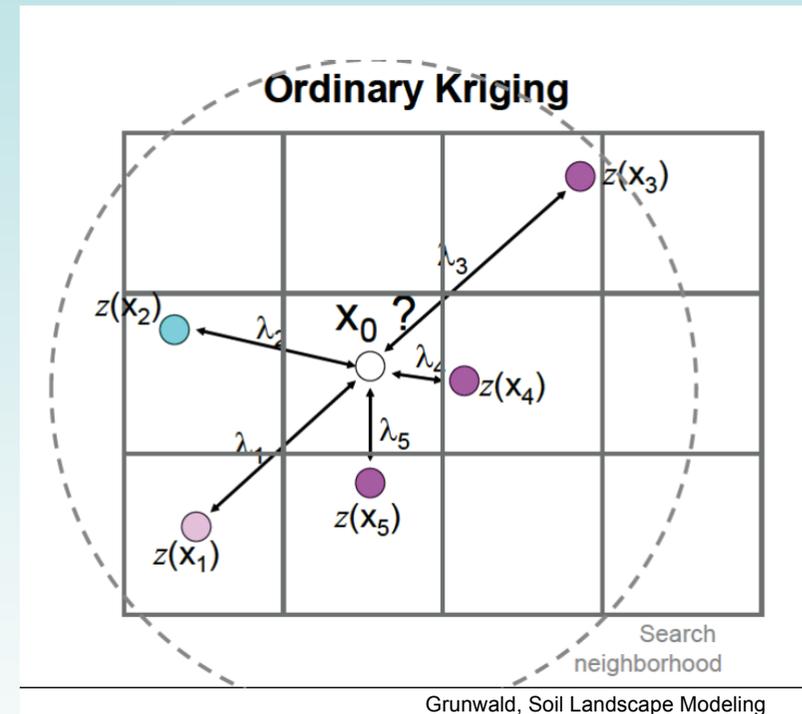
- Kriging stima i valori incogniti attraverso una **media pesata dei punti misurati nel vicinato**.
- I pesi da attribuire ai punti misurati dipendono non solo dalla distanza tra essi e il punto da predire ma anche dall'andamento dell'autocorrelazione spaziale definita dal variogramma.
- I pesi (coefficienti) sono scelti in modo tale da **rendere minima la varianza dell'errore di stima** (Minimum-variance unbiased estimator, MVUE)
- Il **raggio di ricerca** (neighborhood) è in prima analisi il valore del range.

# Kriging: attribuzione dei pesi (coefficienti)

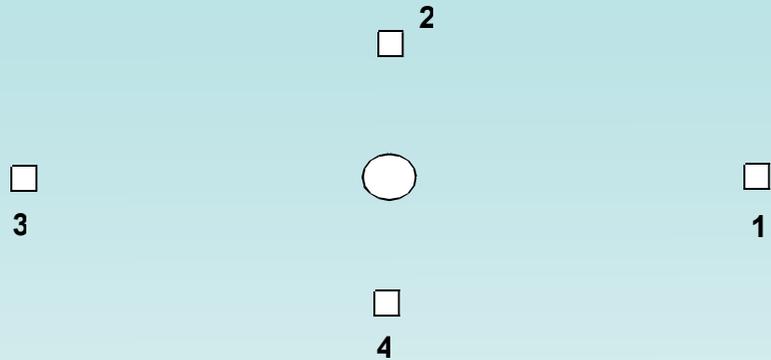
- I 4 o 5 punti più vicini contribuiscono spesso per l'80% del peso totale, i 10 successivi per la rimanente parte.

I pesi dipendono inoltre dalla configurazione spaziale delle misure:

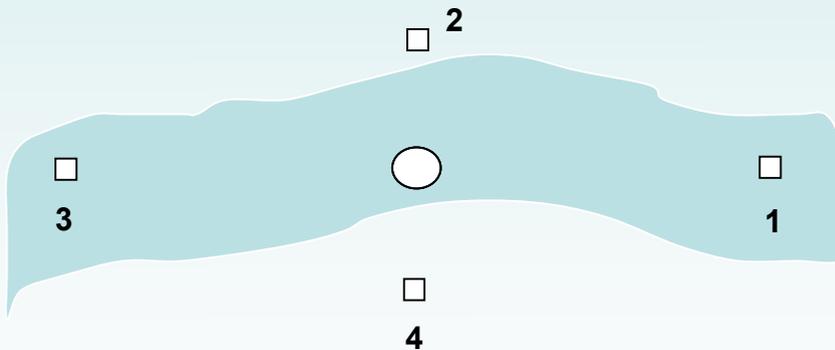
- maggiore è la *varianza di nugget*, più bassi risultano i pesi assegnati ai punti vicini.
- I pesi relativi dipendono dalla dimensione del raggio di ricerca (*neighborhood*): all'aumentare di questa i pesi dei punti vicini diminuiscono e aumentano quelli dei punti più lontani
- Ai punti distribuiti in cluster vengono assegnati pesi individuali inferiori rispetto ai punti isolati alla stessa distanza



# Kriging: attribuzione dei pesi (coefficienti)

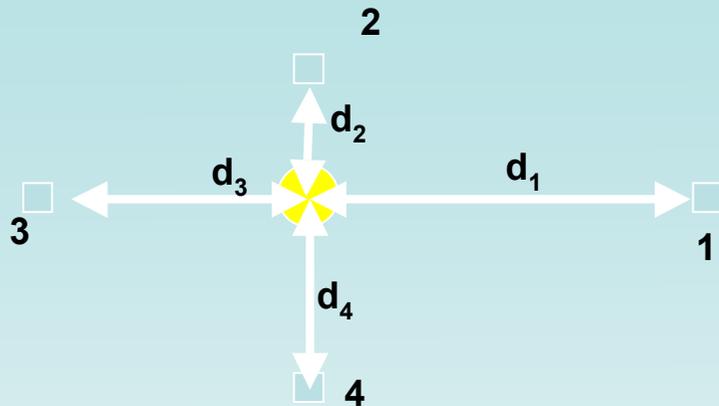


- The unsampled location is in the middle
- Which sample points will receive equal weight?
  - 2 and 4; and 1 and 3
- Which pair will receive smaller weight?
  - 1 and 3
- Kriging weights are consistent with these observations, the estimated value will fall between the four sample points

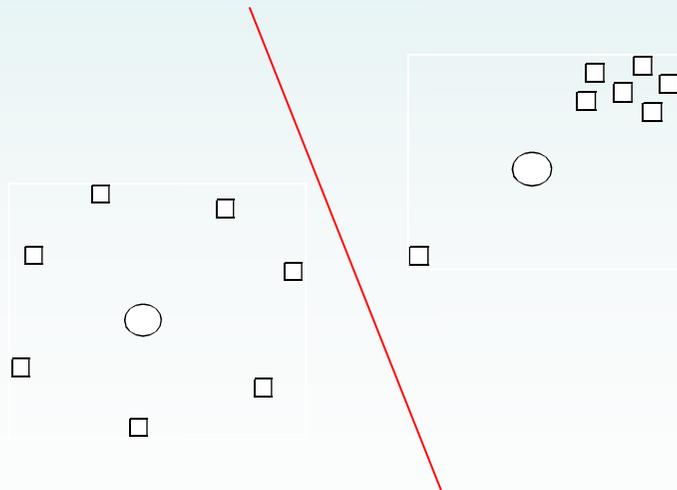


- The unsampled location is in the middle
- Points 1 and 3 are located within a channel
- Which sample points will receive equal weight?
  - 2 and 4; and 1 and 3
- Which pair will receive smaller weight?
  - 2 and 4
- **Kriging weights are dependent on the strength of spatial relationship; not just on the distance**
- The relationship is captured by variogram

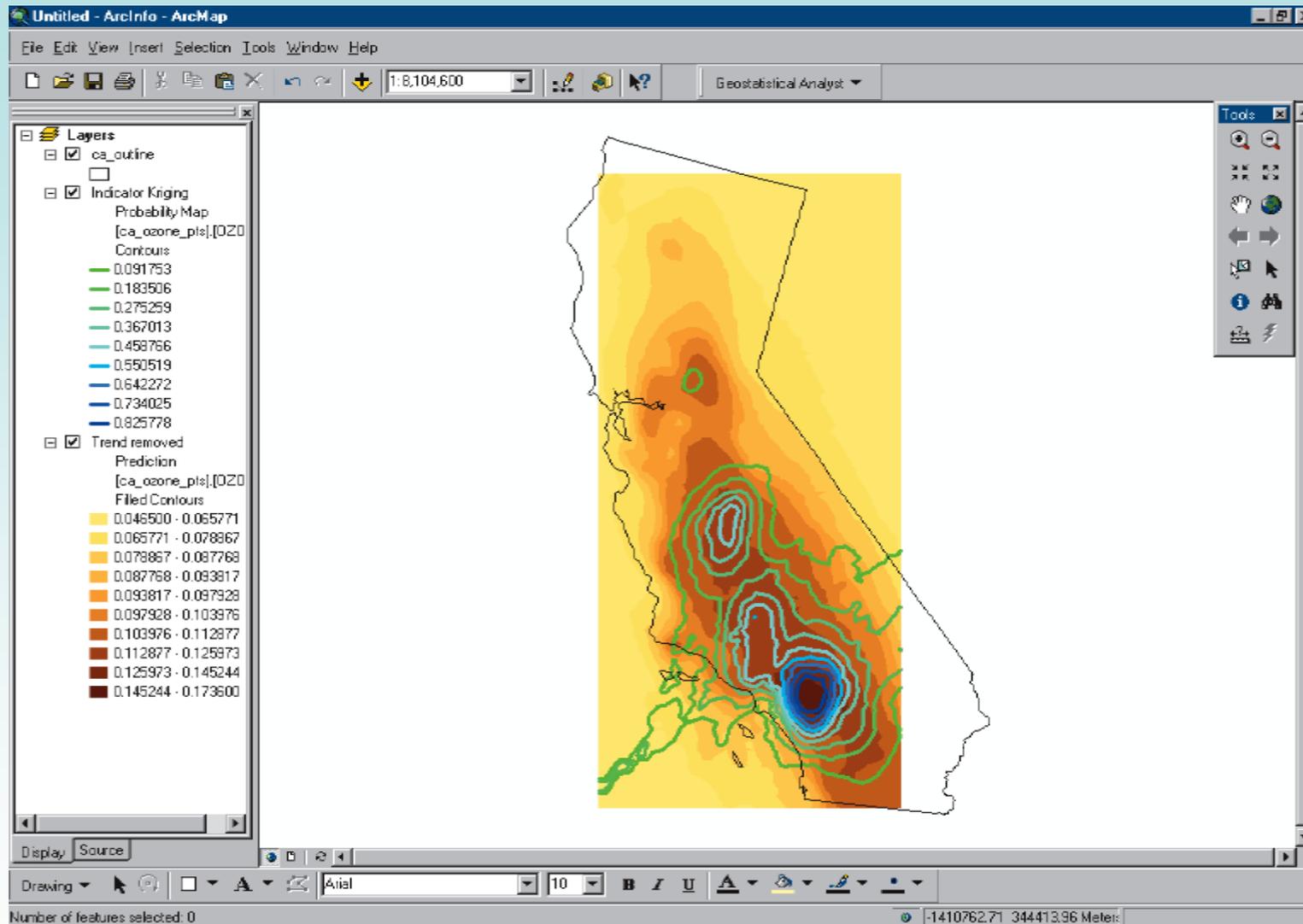
# Kriging: attribuzione dei pesi (coefficienti)



- The unsampled location is in the middle
- **All the sample points are outside the variogram range**
- Which sample points will receive equal weight?
  - **All of them**
- How the weights would be affected if the nugget component of the variogram model gets bigger?
  - The weights assigned will become progressively closer
- **Smaller the nugget, more confidence in the spatial relationship, more dependence on the strength of the relationship**
- That is why, it is important to model variogram with zero nugget unless sample measurements indicate error

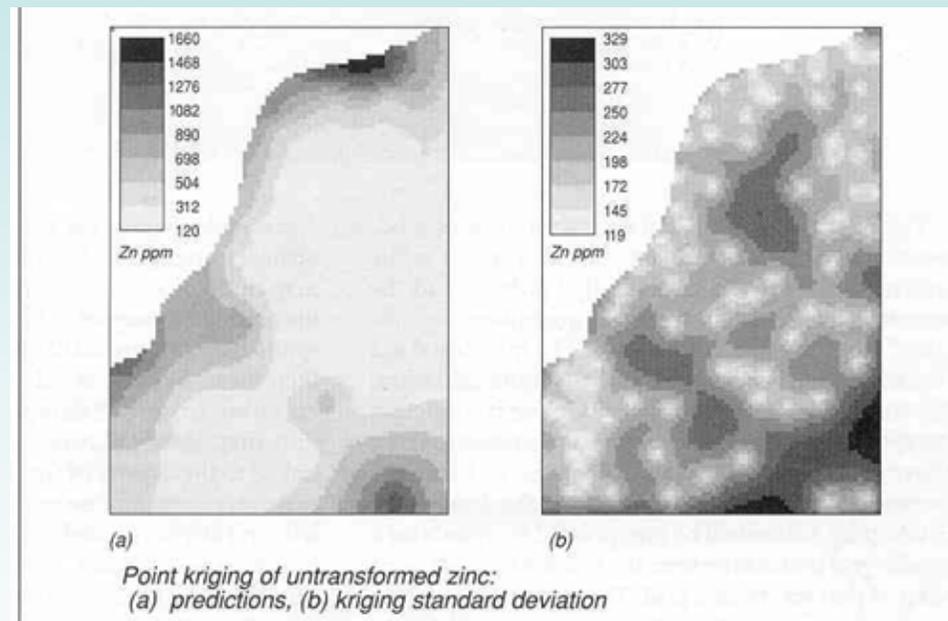
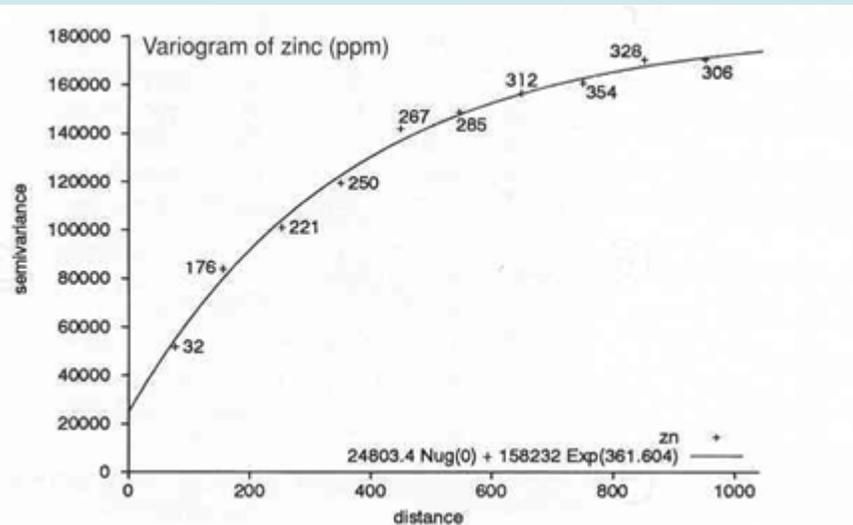


# Kriging: Carta delle previsioni



# Kriging: Carta degli errori

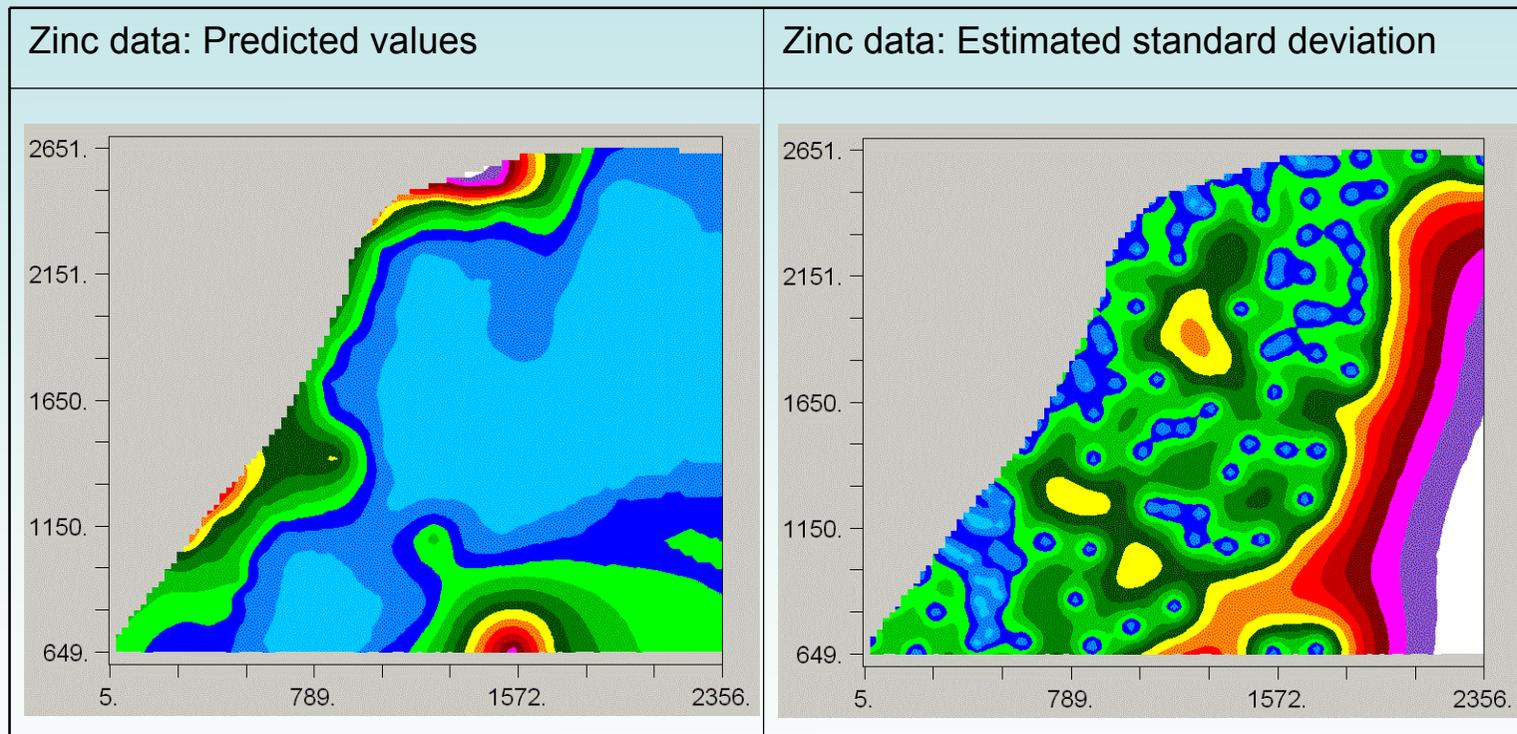
Oltre alla stima del valore, kriging può stimare l'errore associato al valore interpolato offrendo **una misura dell'incertezza** della predizione effettuata (**Prediction Variance**). La Prediction Variance esprime quindi la qualità dell'interpolazione. Solitamente è prossima allo zero in un punto di misura.



**Figure 6.6.** Results of interpolating the Maas data set using ordinary kriging. Above—variogram with fitted exponential model; (a) Kriging predictions of zinc levels, (b) Kriging standard deviations

# Kriging: Carta degli errori

## Ordinary Kriging



# Tipi di kriging

- Ordinary
- Simple
- Universal
- Lognormal
- Indicator
- Probability
- Disjunctive
- Co-kriging (multivariata)
- ecc.

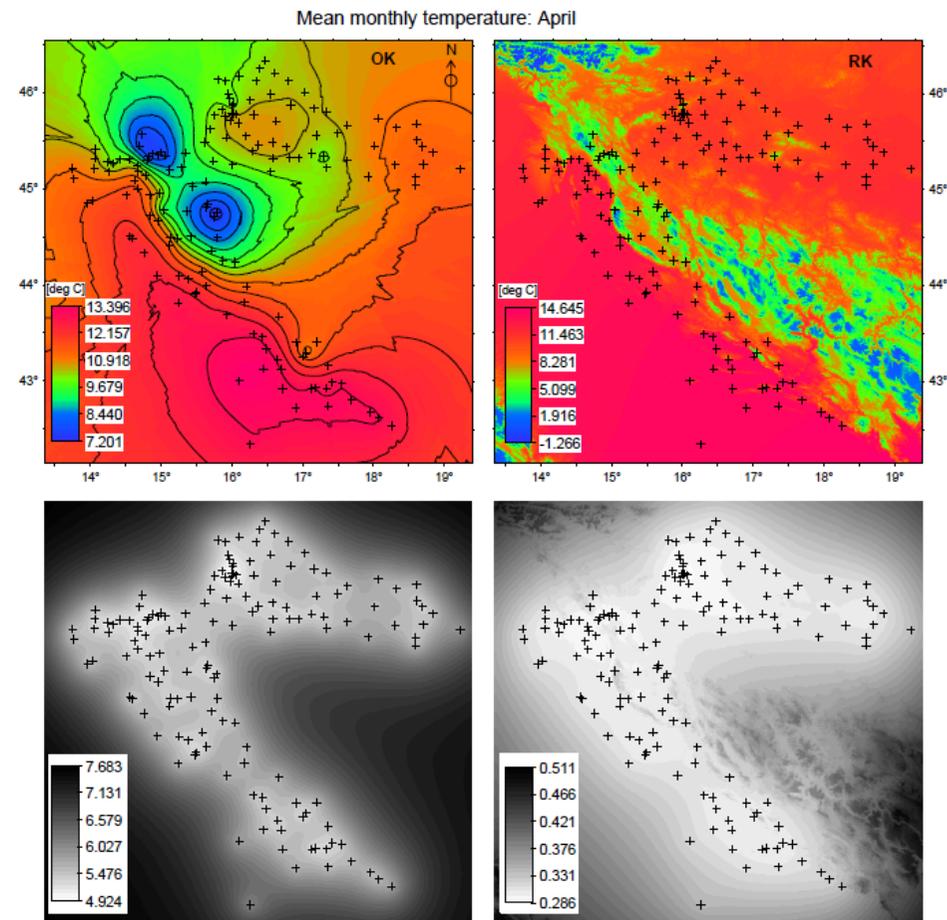
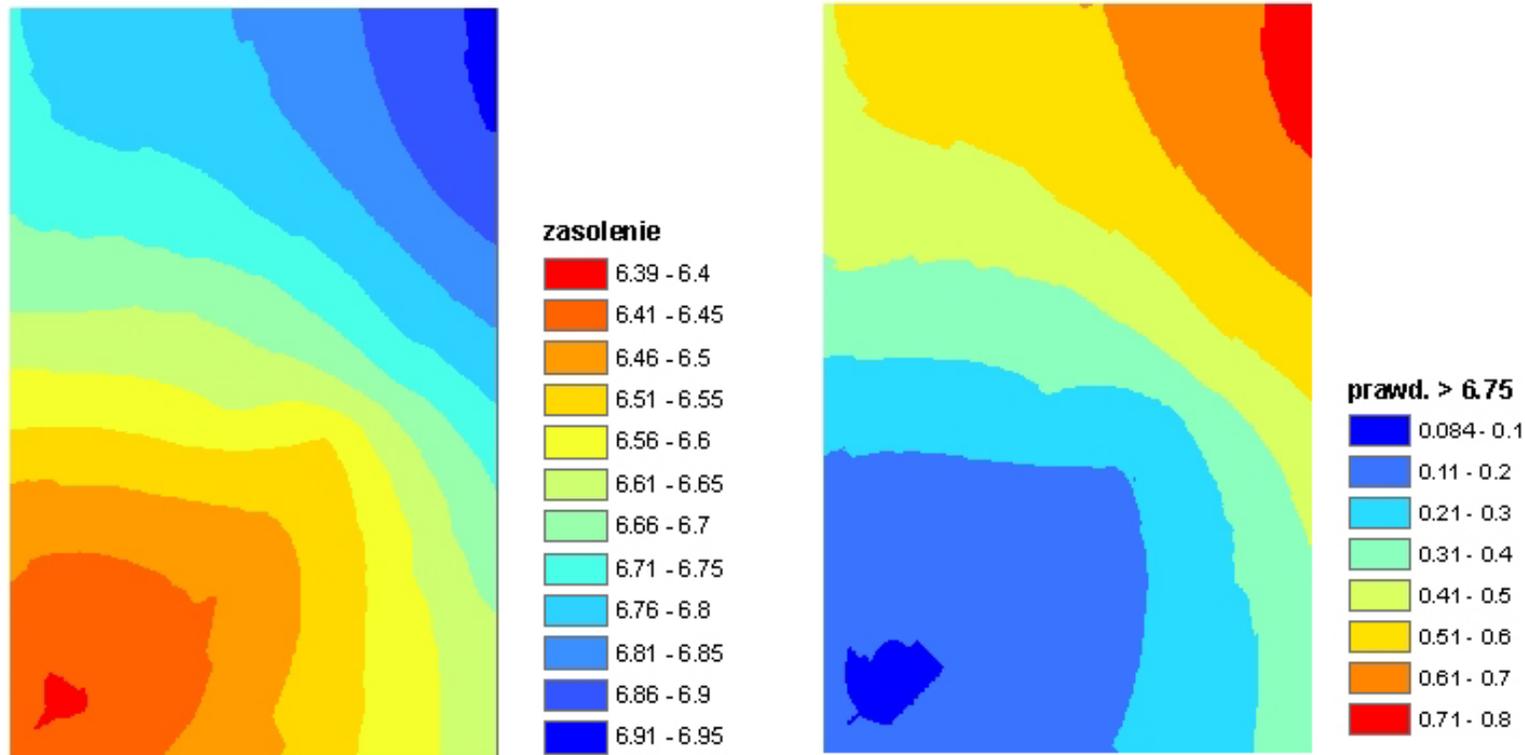


Fig. 2.11: The long-term mean monthly temperature for April interpolated using OK (left) and RK (right) at 1 km grid. Predictions (above) and prediction variances (below) in this case differ significantly. In this case, auxiliary predictors (elevation, latitude, direct annual solar insolation, distance from the coast line) explain 95.2% of variation in the original data (152 meteorological stations), hence the mapping precision is considerably higher for RK.

## Probability Kriging

Determinare la probabilità che la variabile superi un certo valore soglia (probability map).

Esempi: superamento limiti normativi nel contenuto in arsenico dell'acqua potabile, valori soglia di metalli pesanti nel suolo, ecc.



## Analisi multivariata (co-kriging)

Si osservano spesso variazioni associate (co-variazioni) dei valori di due o più parametri ambientali.

Molti fenomeni sono tra loro correlati:

Condizioni igieniche e malattie

Pressione atmosferica ed inquinamento

Associazioni Pb e Ag oppure Pb e Zn

**La covarianza è una misura del grado di associazione tra i valori misurati di due variabili.**

Le covarianze possono essere:

positive o dirette (Es: Alti valori di Pb sono spesso associati ad alti valori di Ag)

Negative o inverse (bassi valori di una variabile e contemporaneamente alti dell'altra)

Nulle: nessuna associazione

# Geostatistica: alcune Problematiche

- Strategia di campionamento
- Numero di punti campionati
- Errori di misurazione
- Scelta metodo interpolazione
- Effetto bordo
- Eccessivo smoothing di certi algoritmi (es. Kriging)
- Utilizzo dei valori di default nelle GUI (next-next-next-finish)
- Risoluzione (cell-size) non adatta del grid interpolato

# Alcune Problematiche

Risoluzione (cell-size) non adatta del grid interpolato

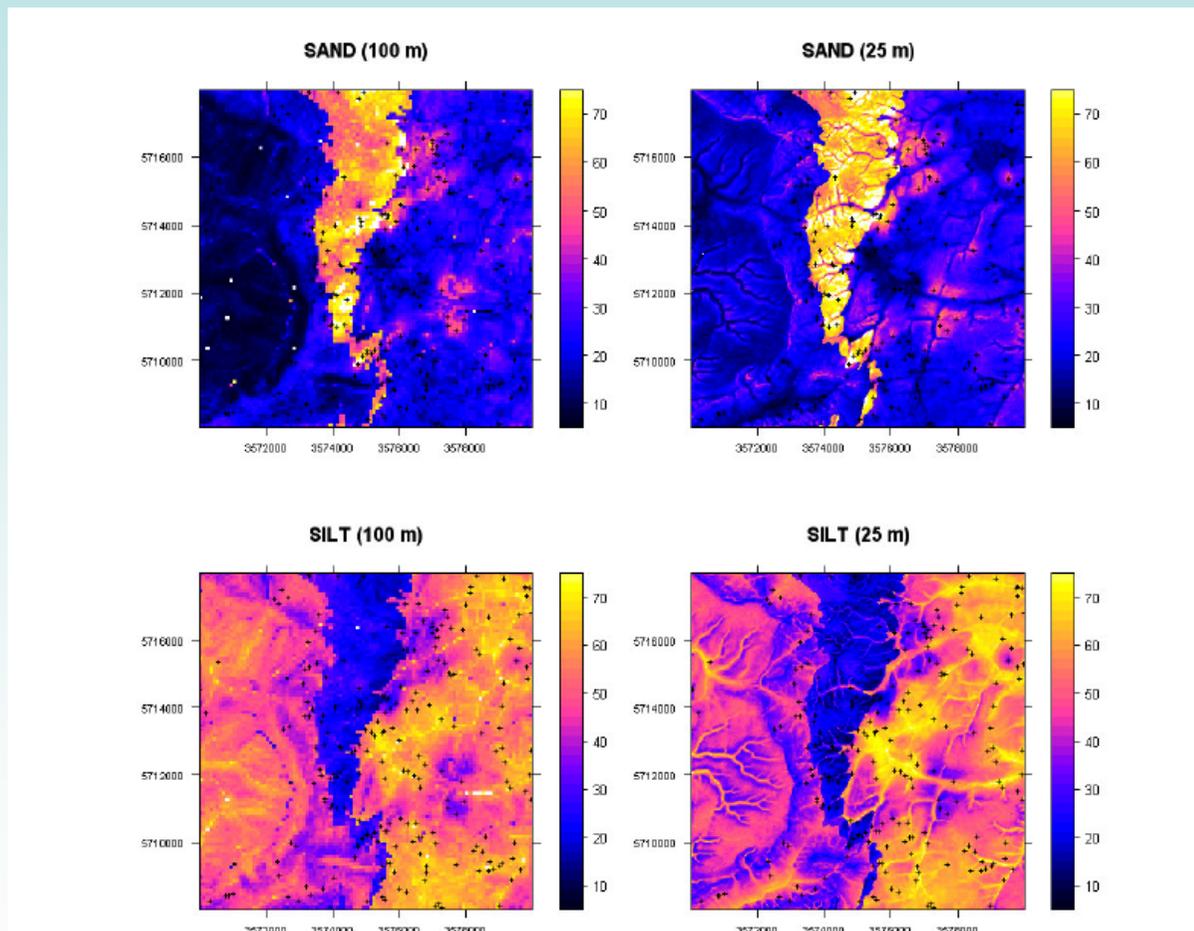
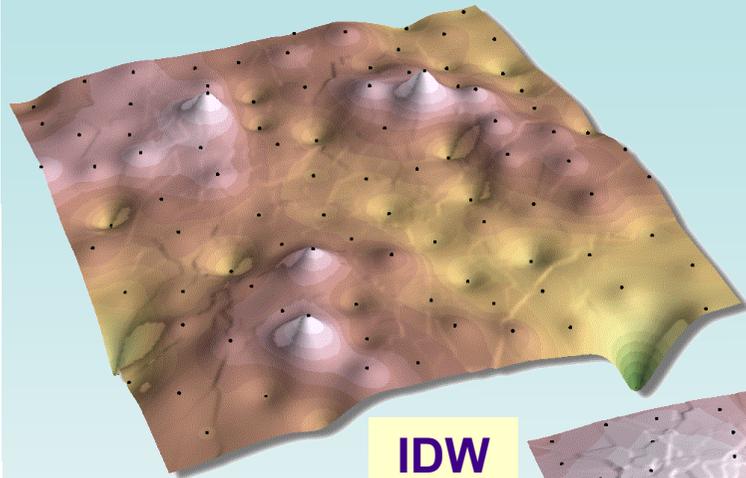
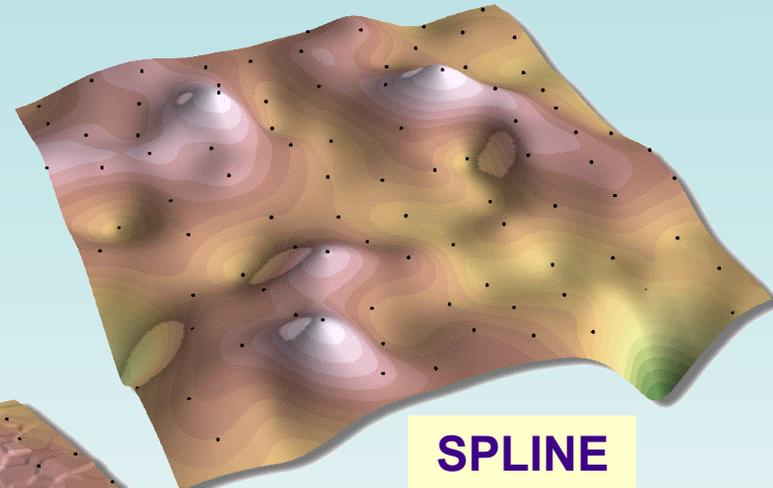


Fig. 4.23: Predictions of SAND and SILT produced using the 100 m and 25 m resolution maps.

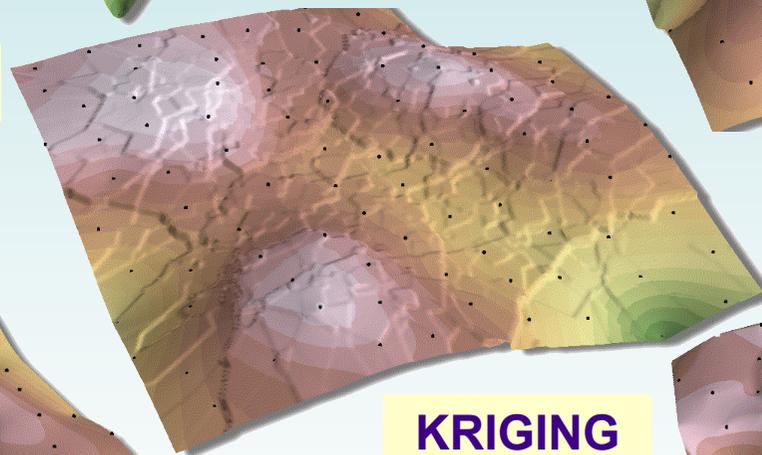
Non esiste un algoritmo di interpolazione appropriato per tutte le applicazioni. La scelta dipende dal tipo di superficie, distribuzione e qualità dei campioni, grado di accuratezza e sforzo computazionale richiesto.



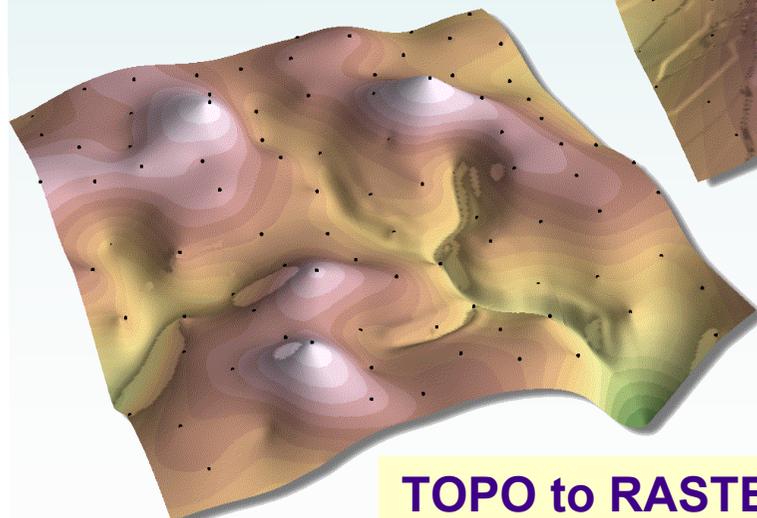
**IDW**



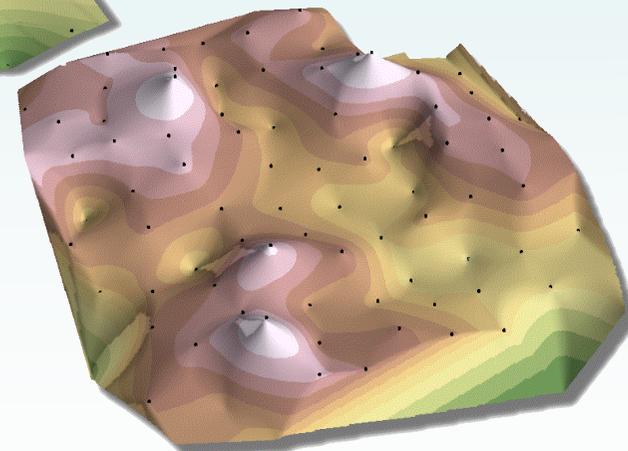
**SPLINE**



**KRIGING**



**TOPO to RASTER**



**NATURAL NEIGHBOR**

## A Comparison of the Geostatistical Analyst methods

Method	Deterministic/ Stochastic	Output Surface Types	Computing Time/ Modeling Time <sup>1</sup>	Exact Interpolator	Advantages	Disadvantages	Assumptions <sup>2</sup>
Inverse Distance Weighted	Deterministic	Prediction	Fast/Fast	Yes	Few parameter decisions	No assessment of prediction errors; produces "bulls eyes" around data locations	None
Global polynomial	Deterministic	Prediction	Fast/Fast	No	Few parameter decisions	No assessment of prediction errors; may be too smooth; edge points have large influence	None
Local polynomial	Deterministic	Prediction	Moderately Fast/Moderate	No	More parameter decisions	No assessment of prediction errors; may be too automatic	None
Radial basis functions	Deterministic	Prediction	Moderately Fast/Moderate	Yes	Flexible and automatic with some parameter decisions	No assessment of prediction errors; may be too automatic	None
Kriging	Stochastic	Prediction; Prediction Standard Errors; Probability; Quantile	Moderately Fast/Slower	Yes without measurement error; No with measurement error	Very flexible; allows assessment of spatial autocorrelation; can obtain prediction standard errors; many parameter decisions	Need to make many decisions on transformations, trends, models, parameters, and neighborhoods	Data comes from a stationary stochastic process, and some methods require that the data comes from a normal distribution
Cokriging	Stochastic	Prediction; Prediction Standard Errors; Probability; Quantile	Moderate/ Slowest	Yes without measurement error; No with measurement error	Very flexible; can use information in multiple datasets; allows assessment of spatial cross- correlation; many parameter decisions	Need to make many decisions on transformations, trends, models, parameters, and neighborhoods	Data comes from a stationary stochastic process, and some methods require that the data comes from a normal distribution

1. Computing time is computer-processing time to create a surface. Modeling time includes user-processing time to make decisions on model parameters and search neighborhoods.

2. We assume that all methods are predicting a smooth surface from noisy data.

## Alcune immagini provenienti da:

- Lectures of Applications in GIS/Geographic Data Analysis (Department of Geography and Urban Studies, Temple University)
- Lectures of 19<sup>th</sup> Advanced Summer School in Regional Science
- UB Geology GLY560: GIS
- [www.spatialanalysisonline.com](http://www.spatialanalysisonline.com)
- Tomislav Hengl- Practical Guide to Geostatistical Mapping of Environmental Variables. 2007,European Commission - Joint Research Centre - Institute for Environment and Sustainability
- Raco et. Al. (2009) - Gas emission into the atmosphere from controlled landfills: an example from Legoli landfill (Tuscany, Italy) - Environ Sci Pollut Res
- Bisson et al. LiDAR-based digital terrain analysis of an area exposed to the risk of lava flow invasion: the Zafferana Etnea territory, Mt. Etna (Italy) - Nat Hazards (2009) 50:321–334
- Baroni C., Noti V., Ciccacci S., Righini G., Salvatore M.C. (2005) - *Fluvial Origin of the Valley System in northern Victoria Land (Antarctica) from Quantitative Geomorphic Analysis* . Geological Society of America Bulletin, 117 (1-2), 212-228.