



*Università degli Studi di Napoli “Parthenope”
Dipartimento di Scienze e Tecnologie*

Corso di Topografia e Idrografia

Lezione 16

Modellazione batimetrica Utilizzo della geostatistica

Claudio Parente

Geostatistica – Applicazioni per DTM e modellazione batimetrica

In *Idrografia* occorre occuparsi anche della restituzione del rilievo, ovvero costruire il modello dei fondali marini (o anche fluviali o lacustri) a partire dai dati di profondità acquisiti.

Qualora questi dati non siano già disposti secondo un grigliato regolare, come nel caso dei dati acquisiti con *single beam*, è necessario effettuare delle interpolazioni e quindi ricorrere alla Geostatistica.

In questa lezione ci occupiamo dei fondamenti della geostatistica ed in particolare dell'interpolatore *Kriging*.

Geostatistica – Applicazioni per DTM e modellazione batimetrica

La prima legge della geografia: “tutto è correlato con tutto, ma le cose più vicine sono più correlate delle cose lontane”.

(Tobler 1978).

Partendo da questo principio, già si evince che le profondità di punti vicini tendono ad assumere profondità simili tra di loro, o per lo meno profondità spazialmente correlate in misura maggiore di quanto non avvenga per punti più distanti.

Partiamo dalla definizione di Geostatistica e quindi ricordiamo cosa è la Statistica.

Statistica

Una definizione di Statistica:

- “tecnica che ha per scopo la conoscenza quantitativa dei fenomeni collettivi”.

Lo studio del *fenomeno collettivo* si può fare osservando **interamente** la collettività di individui (Popolazione).

Lo studio del *fenomeno collettivo* si può fare osservando **una sua parte** (Campione).

Geostatistica

La geostatistica è quella branca della statistica che si occupa dell'analisi di dati geografici.

La geostatistica si occupa di valutare l'autocorrelazione spaziale dei dati, cercando di verificare se osservazioni effettuate su punti vicini presentano effettivamente una minore variabilità rispetto ad osservazioni effettuate tra punti distanti.

L'obiettivo è quindi valutare l'effetto della posizione del punto di misura sulla variabilità del dato osservato. Tale variabilità viene di solito elaborata con lo strumento del semivariogramma che valuta la variazione del grado di variabilità di punti a distanze crescenti.

Mediamente i punti vicini hanno valori “simili”
 (“più simili” rispetto a punti lontani):

Ad esempio:

- modelli digitali delle altezze: punti vicini hanno quote simili;
- mappe vegetazionali: punti vicini hanno composizione specifica simile.

Ci sono eccezioni  discontinuità del campo

L'interpolazione è stata oggetto del Corso di Cartografia Numerica e GIS. Richiamiamo alcuni concetti che utilizzeremo in questo corso.

Si deve potere assegnare un valore anche ai punti non campionati

Si cerca di “ricostruire” la variabilità persa con il campionamento.

Si assegna un valore ai punti attraverso:

- **interpolazione**, per i punti all'interno dell'area coperta dai punti noti;
- **estrapolazione**, per i punti al di fuori.

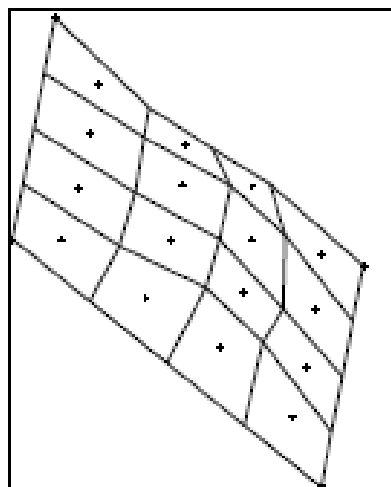
Interpolazione o approssimazione?

Spesso quello che è indicato come interpolazione in realtà è:

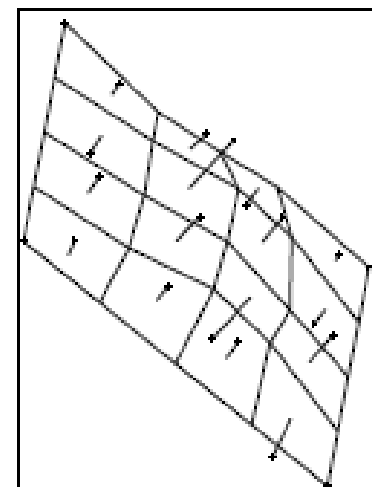
- **interpolazione**, la funzione passa per i valori nei punti noti;
- **approssimazione**, minimizza le differenze rispetto ai valori nei punti noti secondo un criterio prefissato, ma in generale **non** passa per i punti.

Esempi

interpolazione



approssimazione



Il problema dell'interpolazione consiste in:
trovare un modello plausibile per esprimere il fatto che
punti vicini hanno mediamente valori simili

Tipi di modelli di interpolazione:

Deterministici

il legame fra punti vicini è
espresso da una legge
esplicita i cui parametri hanno
significato fisico.

Stocastici

il legame fra punti vicini è
espresso da un legame
statistico (covarianza) che può
non avere significato fisico.

L'interpolazione può essere effettuata globalmente o localmente

Metodi globali

tutte le informazioni (valori) sono usate per costruire il modello.

Metodi locali

solo le informazioni (valori) dei punti vicini sono usate per costruire il modello.

Geostatistica

Oltre a valutare la variabilità spaziale la geostatistica offre delle tecniche di interpolazione spaziale che possono fornire delle stime sul valore assunto da una variabile in una posizione in cui la misurazione non è stata effettuata in base a dei dati rilevati su punti vicini. Tra questi metodi abbiamo (tra gli altri):

Interpolazione «al più vicino» (Nearest neighbor);

interpolazione con media ponderata;

interpolazione bilineare;

interpolazione con superficie del 2° ordine;

interpolazione con superficie del 3° ordine;

Kriging;

spline.

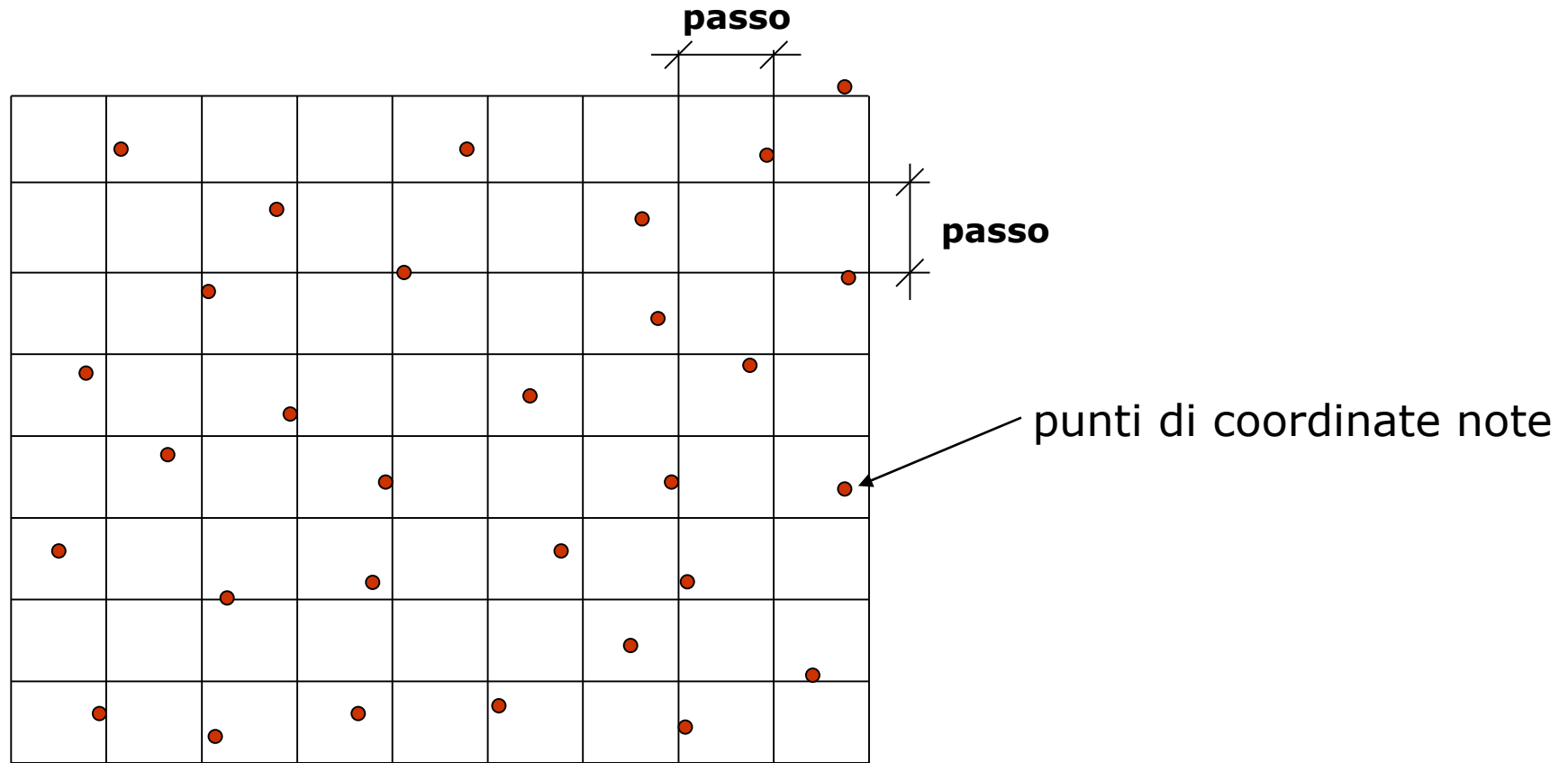
GRID per DTM

I metodi suddetti permettono di realizzare un **grid**: viene attribuito a ciascun pixel il valore di quota più probabile in base ai valori assunti da una rosa di punti prossimi.

Il grid può essere utilizzato per costruire modelli digitali del terreno (DTM, Digital Terrain Model) o dei fondali marini (Digital Bathymetry/Batimetric Model), ma anche per rappresentare la variabilità di una qualsiasi grandezza georiferibile e variabile sul territorio (esempio: temperatura, inquinante, ecc.).

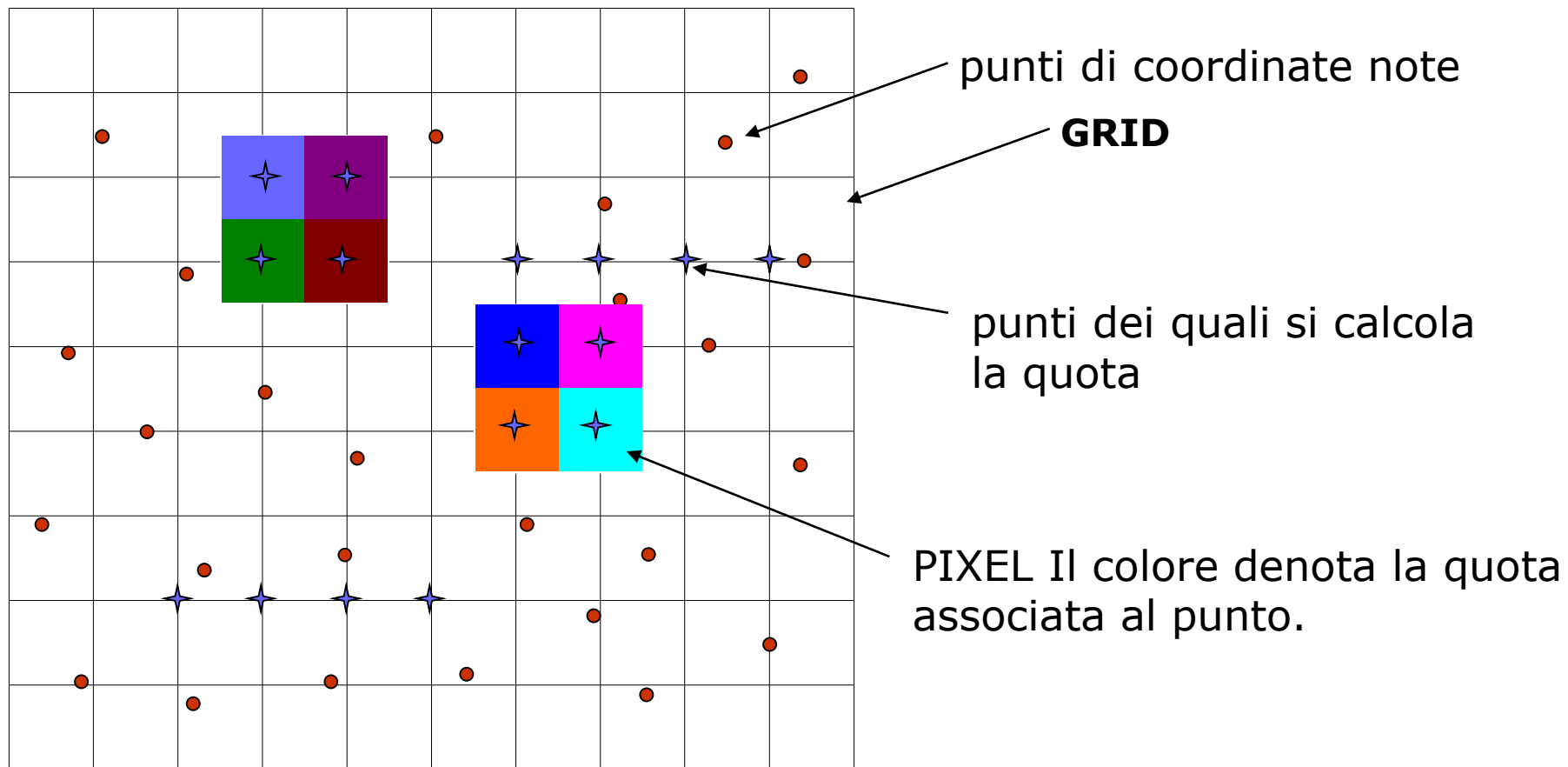
Molti dei metodi elencati nella slide precedente sono stati oggetto del precedente Corso di Cartografia numerica e GIS. In questa lezione analizziamo il kriging.

Disponiamo di una serie di punti di cui sono noti x, y, z .



Il primo passo è la costruzione di un grigliato (GRID) formato da maglie regolari di passo p . La scelta del passo deve essere compatibile con le distanze tra i punti noti. Per esempio, se il rilievo è stato condotto per una scala 1:5000, il passo viene solitamente posto pari a 4-6 m.

Nei punti di intersezione della griglia, si applicano i pixels del formato raster, ai quali si attribuisce la quota, rappresentata da un diverso colore.

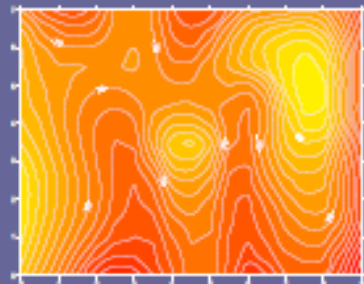


XYZ DATA FILE (*.xls, *.dat, *.slk, *.bln, *.bna)

GRID DATA COMMAND

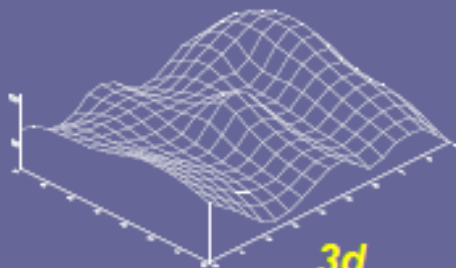
GRID FILE (*.grd)

CONTOUR



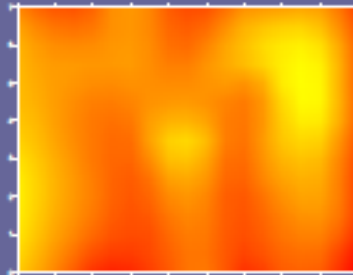
2d

WIREFRAME



3d

IMAGE

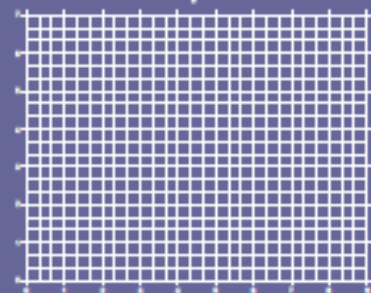
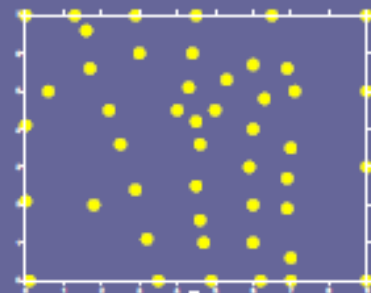


2d

SURFACE



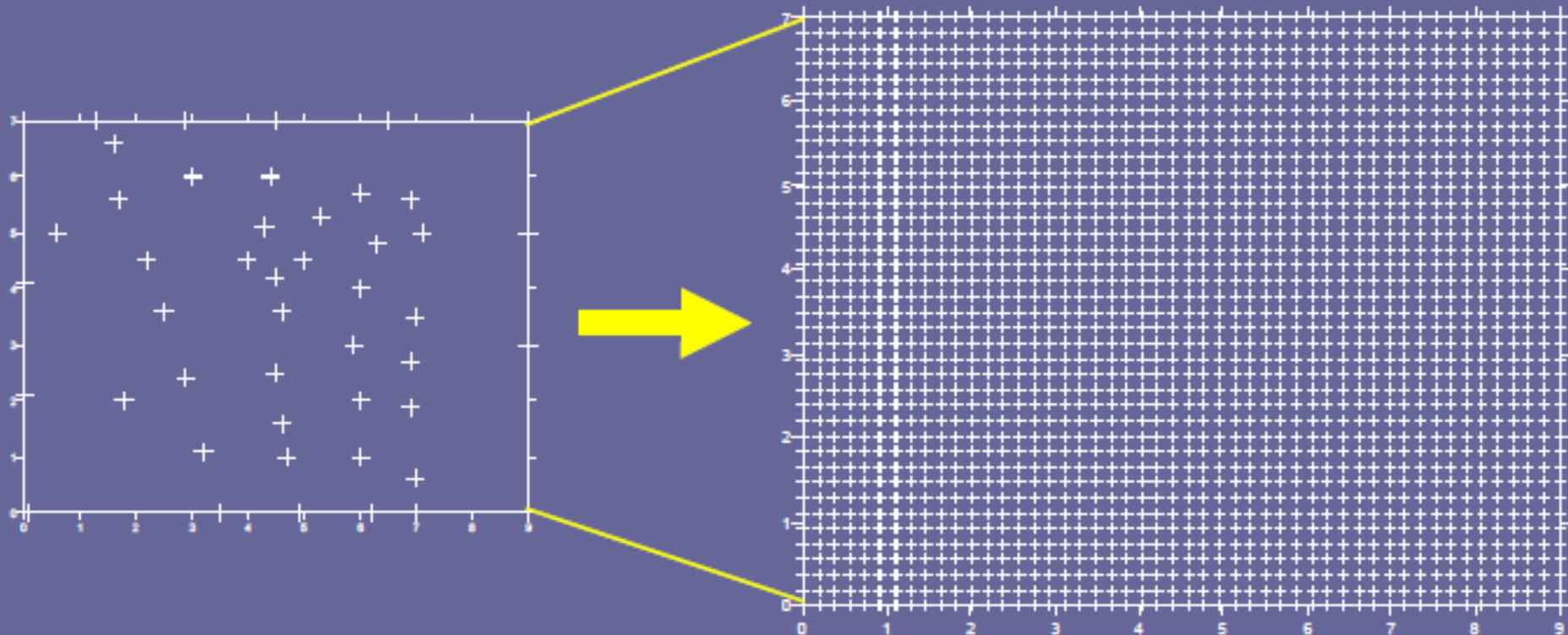
3d



INTERPOLAZIONE SPAZIALE CON SURFER 8 (GRIDDING)

Con il processo di *gridding* viene ricostruito l'andamento di una determinata variabile secondo una griglia regolare di nodi, equamente spaziata, a partire da una serie di valori discreti ed irregolarmente distribuiti nello spazio

La griglia regolare di nodi prende il nome di GRID



CONTOUR MAP (mappa a isolinee o linee isovalore)

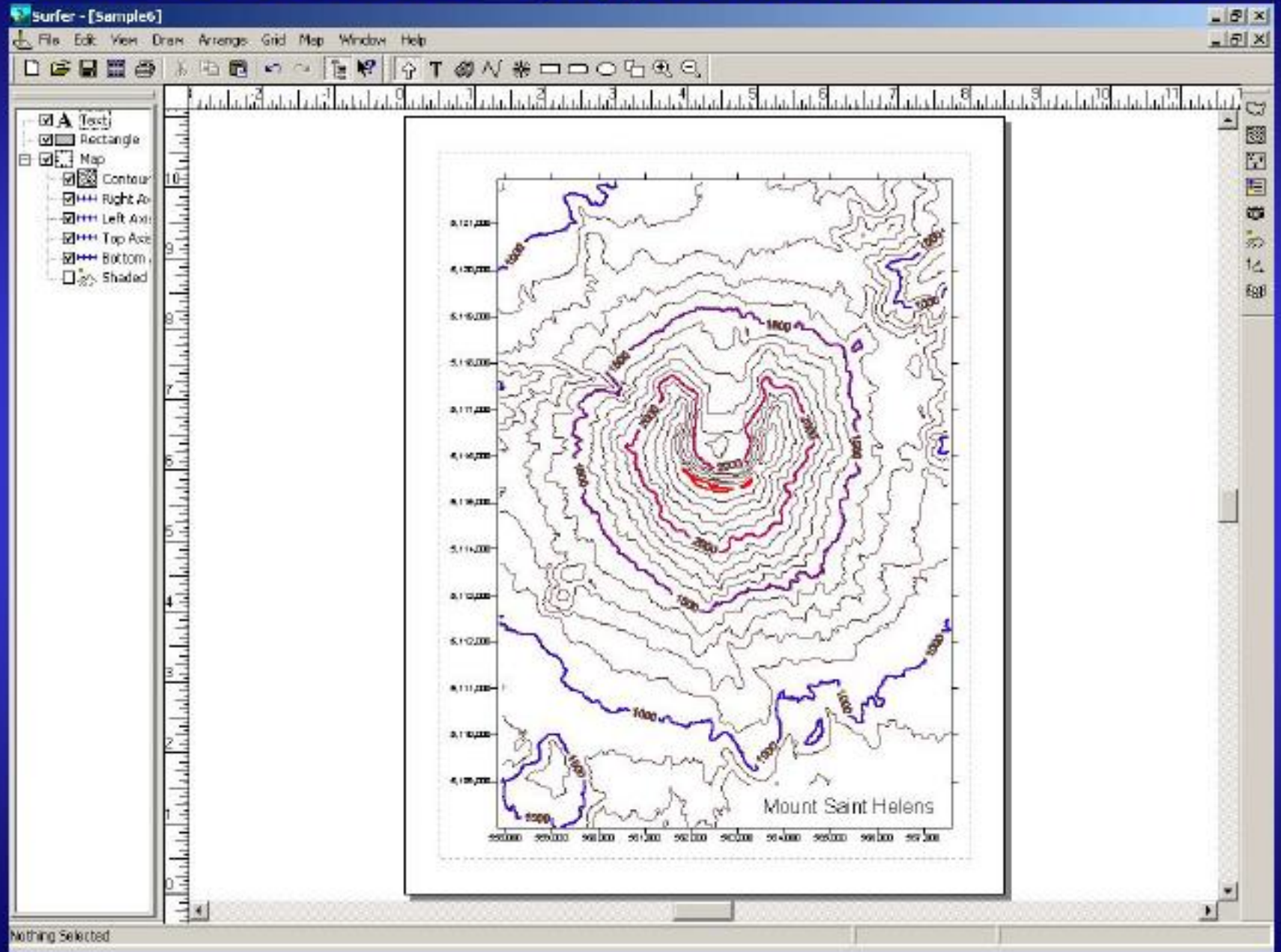
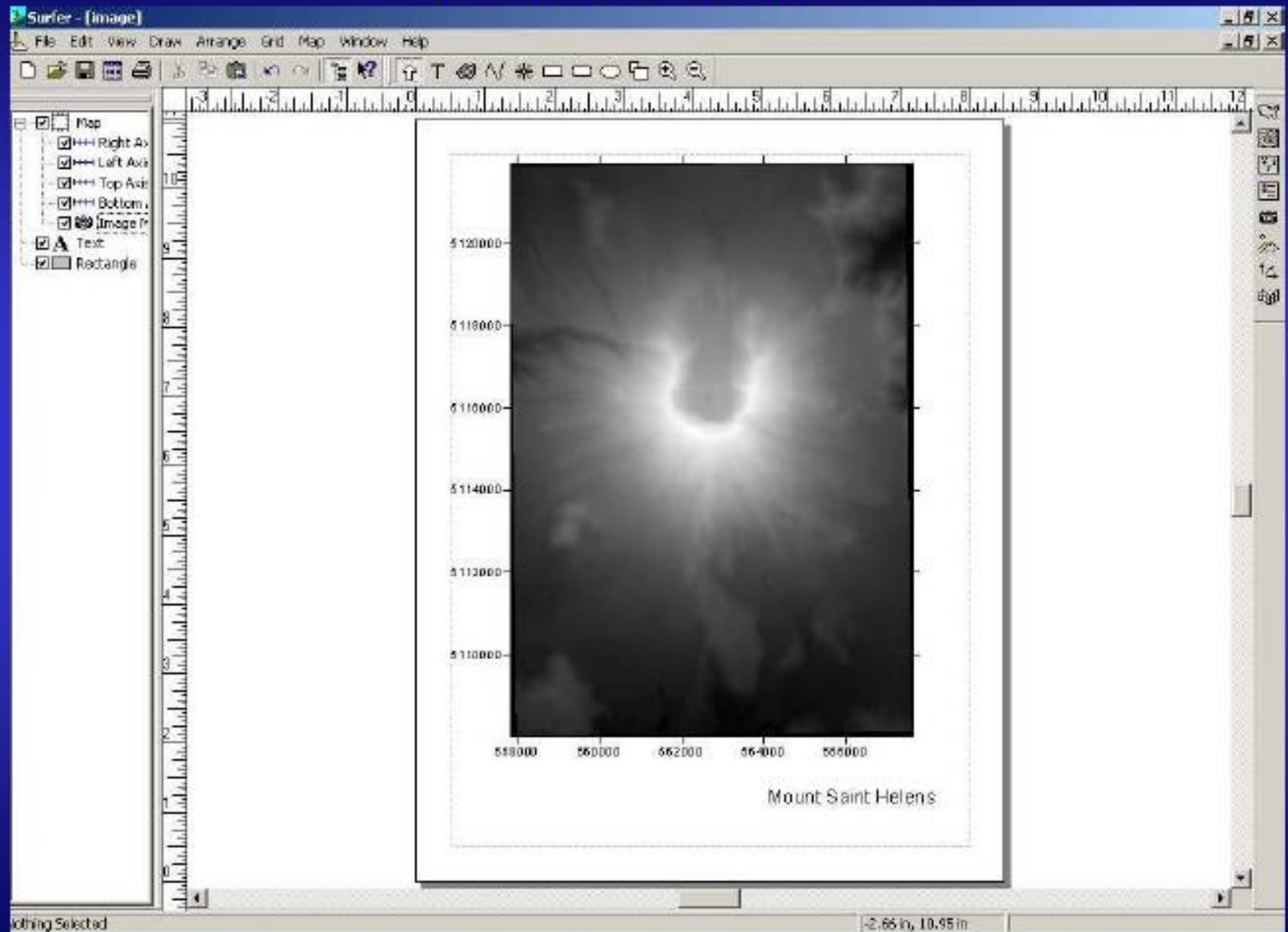
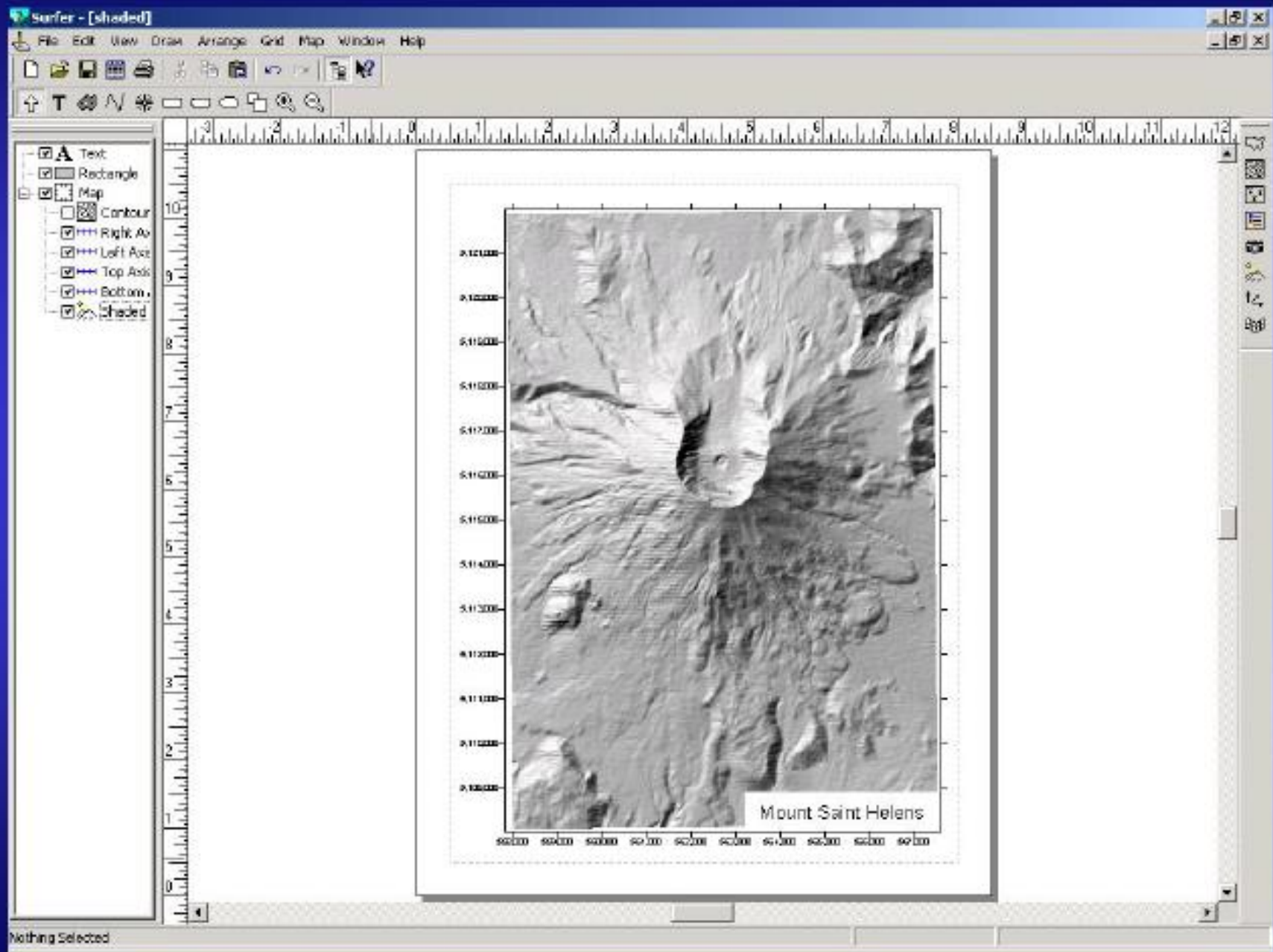


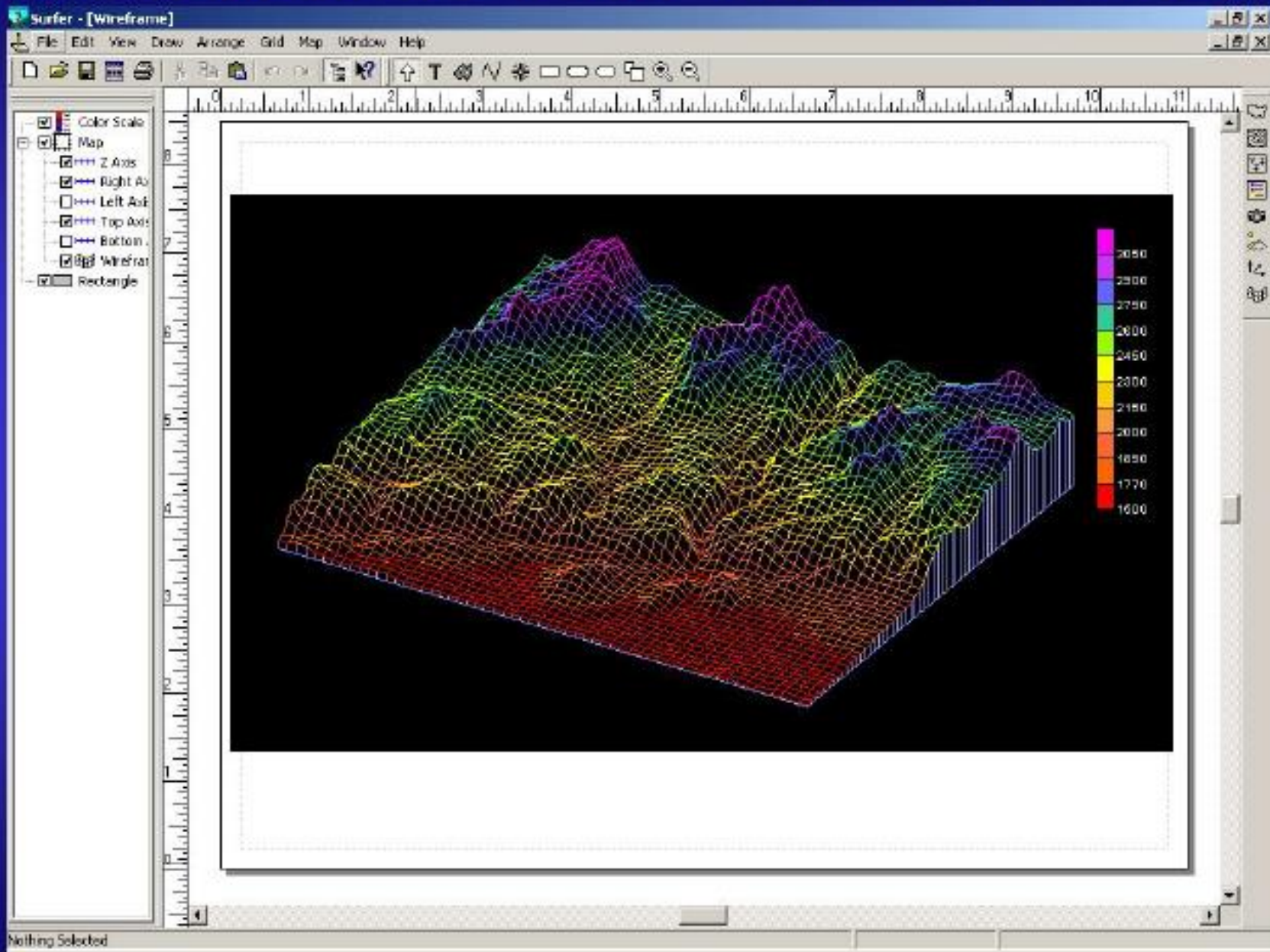
IMAGE MAP



SHADED RELIEF MAP (ombreggiatura del rilievo)

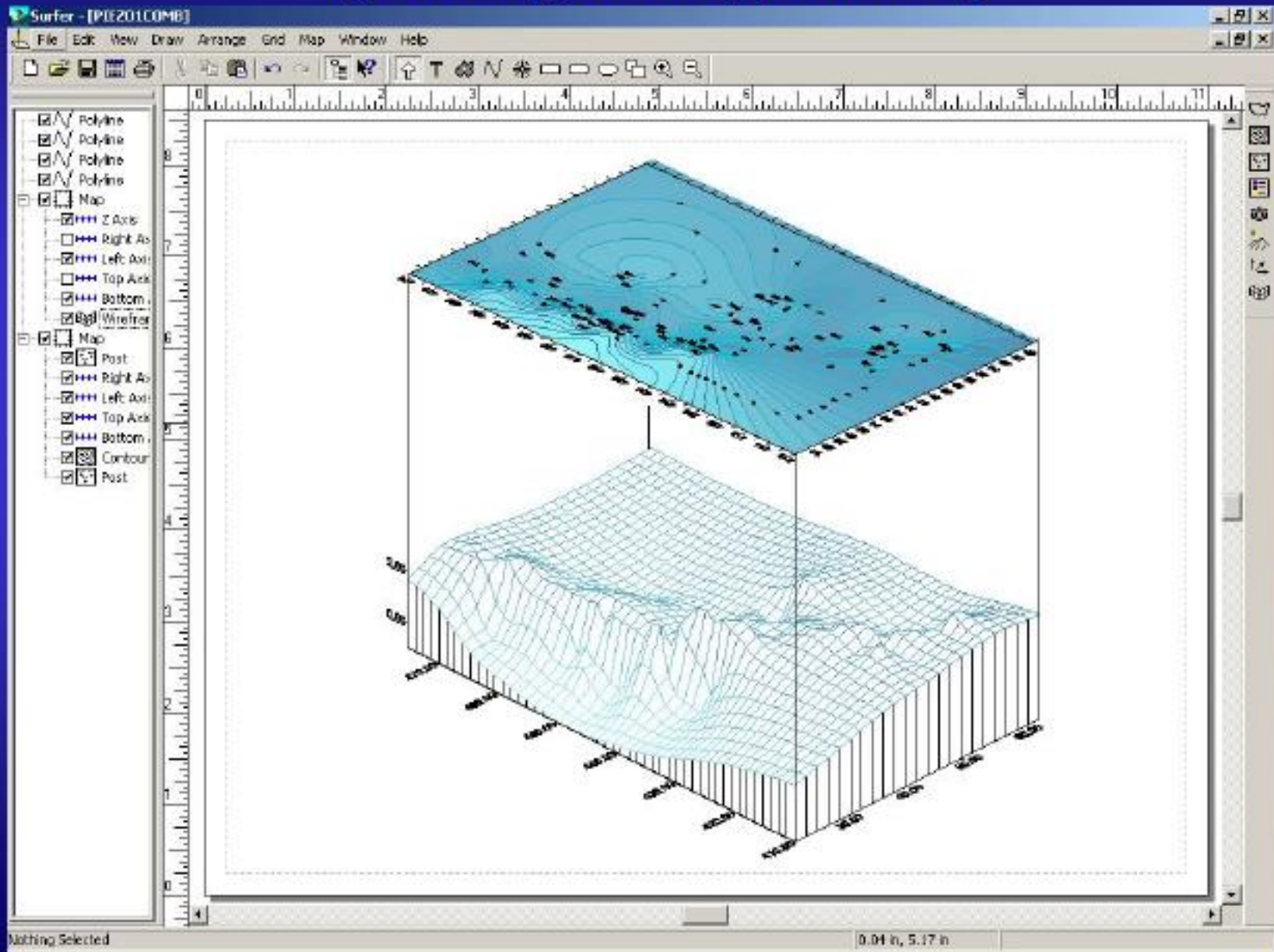


WIREFRAME MAP (vista 3D con reticolo)

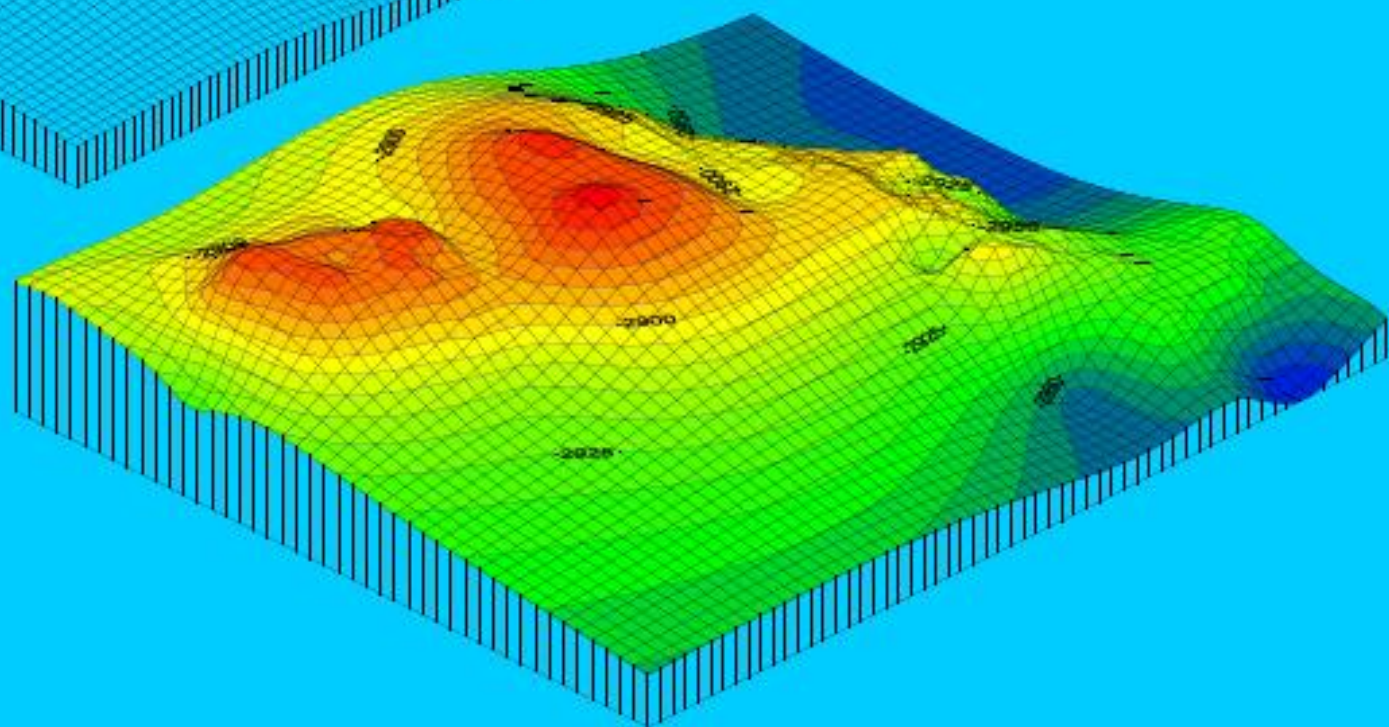
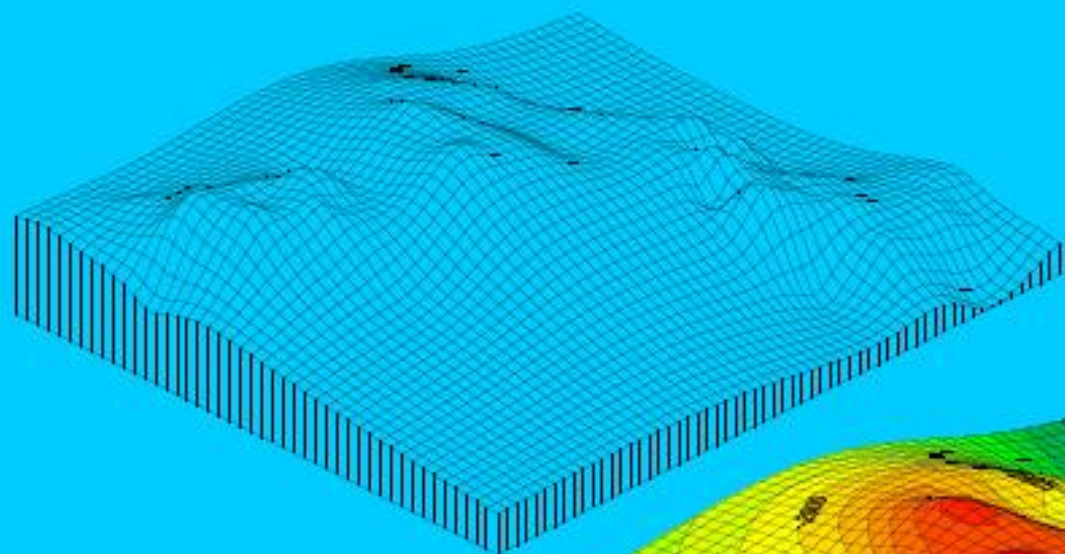


WIREFRAME/CONTOUR MAP

(stacking; sovrapposizione)



Drappeggio



GEOSTATISTICA

Tutte le discipline che studiano l'ambiente, in particolare la scienza del suolo, per ottenere stime affidabili delle proprietà indagate, devono affrontare il

PROBLEMA DELLA VARIABILITÀ DEGLI ATTRIBUTI NELLO SPAZIO.

Secondo l'impostazione statistica classica, che presuppone l'indipendenza spaziale dei campioni e la presenza di una variabilità casuale introdotta con procedure codificate di randomizzazione, il valore di una proprietà in una località non campionata potrà essere stimato attraverso il valore medio delle osservazioni effettuate nella regione indagata, e con una misura di dispersione (varianza, deviazione standard, limiti fiduciali, ecc.) che definisce la precisione della stima.

Misure di continuità spaziale quantificano la relazione tra il valore di una variabile in un punto e il valore della stessa variabile (o di un'altra variabile differente) in un altro punto.

Il concetto di base della **CONTINUITA' SPAZIALE** è semplice:

Mediante più due valori sono vicini più i valori delle variabili in essi misurati si somigliano

Gli strumenti per l'analisi geostatistica

Tra gli strumenti più diffusi per l'analisi geostatistica dei dati ricordiamo:

H-scatterplot;
Variogramma.

h_scatterogramma

Le relazioni spaziali tra i dati possono essere fra gli altri modi indagate usando un h_scatterogramma.

L' h_scatterogramma è un grafico che riporta tutte le coppie di valori noti dell'attributo che stanno in siti separati da una data distanza h lungo una certa direzione θ ($z(x,y), z((x,y)+h)$).

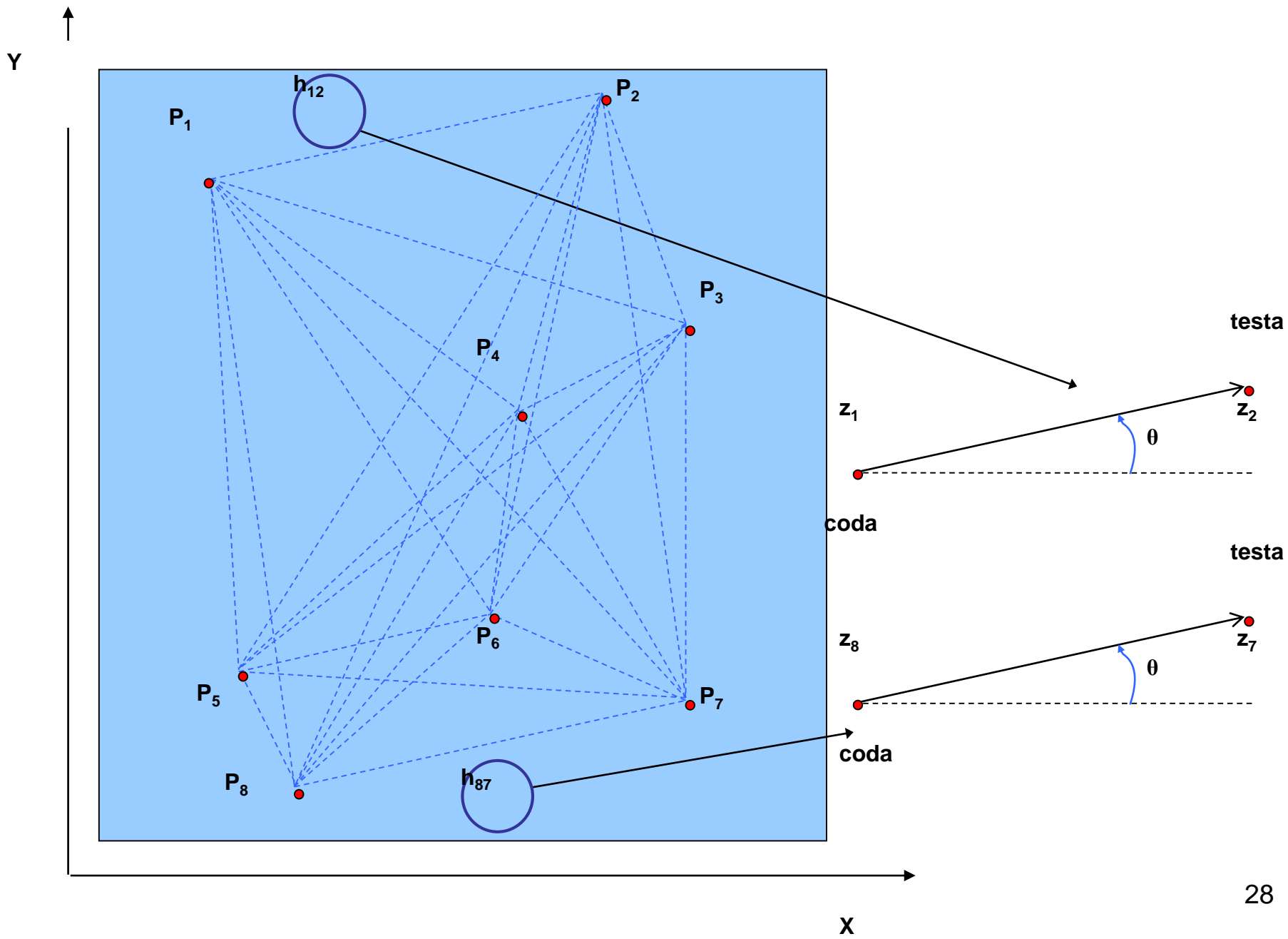
La notazione vettoriale h tiene conto sia della distanza che della direzione.

h_scatterogramma

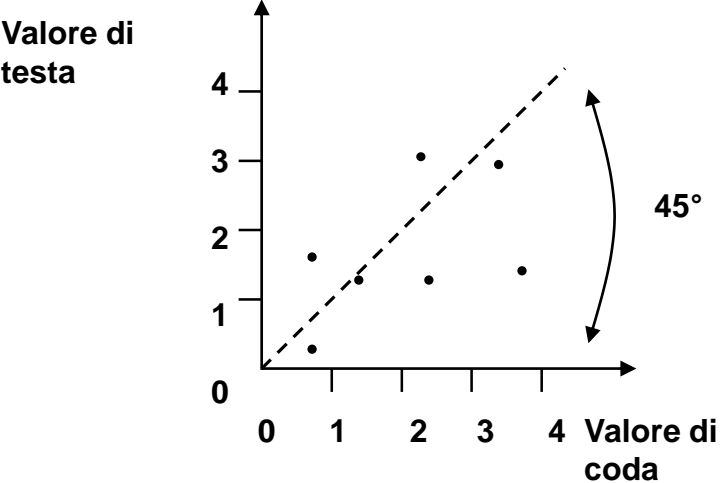
In generale la distanza h che separa ogni coppia di punti, poiché sono note le coordinate planimetriche x e y degli stessi, è data banalmente dall'applicazione del teorema di Pitagora.

La direzione nonché il verso del vettore h sono invece esplicitate dal valore dell'angolo θ che è misurato in senso antiorario a partire dal semiasse X positivo.

Per convenzione, il valore dell'attributo all'inizio del vettore h è chiamato valore di coda mentre quello alla fine dello stesso è detto valore di testa.



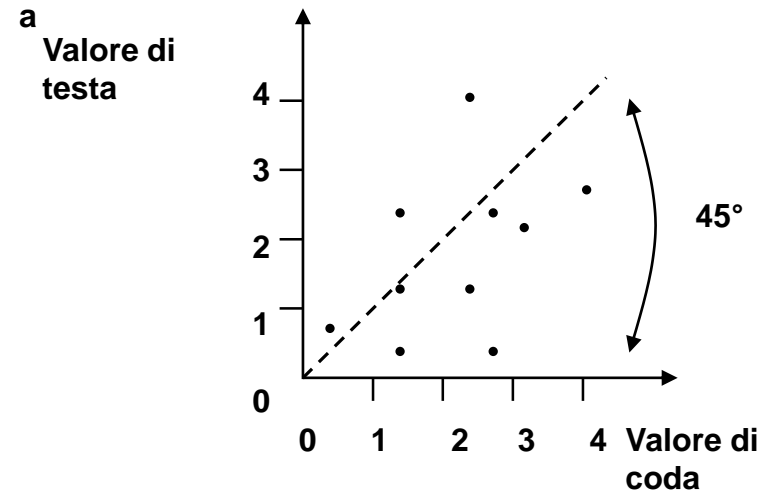
h
1_scatterogramma



.....

I valori di testa e coda di tutte le coppie sono riportate nel relativo h_scatterogramma. Per una data distribuzione di punti noti è possibile quindi costruire diversi h_scatterogrammi che graficamente mostrano per un determinato vettore h la somiglianza o dissomiglianza fra i dati.

h_n_scatterogramm
a



Semivariogramma

E' il diagramma risultante dalla rappresentazione della semivarianza in funzione della distanza tra due punti.

Matematicamente la semivarianza è data da:

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

dove:

$y^*(h)$ è il valore della semivarianza alla distanza h ;

z è il valore della variabile quota;

x_i ed x_i+h indicano le posizioni dei punti di ciascuna coppia;

n è il numero delle coppie dei punti separati da h .

Kriging

- Il metodo di interpolazione Kriging parte dal presupposto che la variazione spaziale di un qualsiasi attributo continuo è spesso troppo irregolare per essere modellata da una semplice funzione matematica.
- La variazione può essere invece meglio descritta da una superficie stocastica; l'attributo in questo caso prende il nome di variabile regionalizzata.
- Il metodo assicura quindi che la predizione dei valori dell'attributo nei punti non campionati sia ottimale tenendo conto delle assunzioni fatte. I pesi di interpolazione sono cioè scelti in maniera tale da ottimizzare la funzione interpolante.

Kriging

La teoria della variabili regionalizzate assume che la variazione spaziale di una qualsiasi variabile può essere espressa come la somma di tre componenti principali:

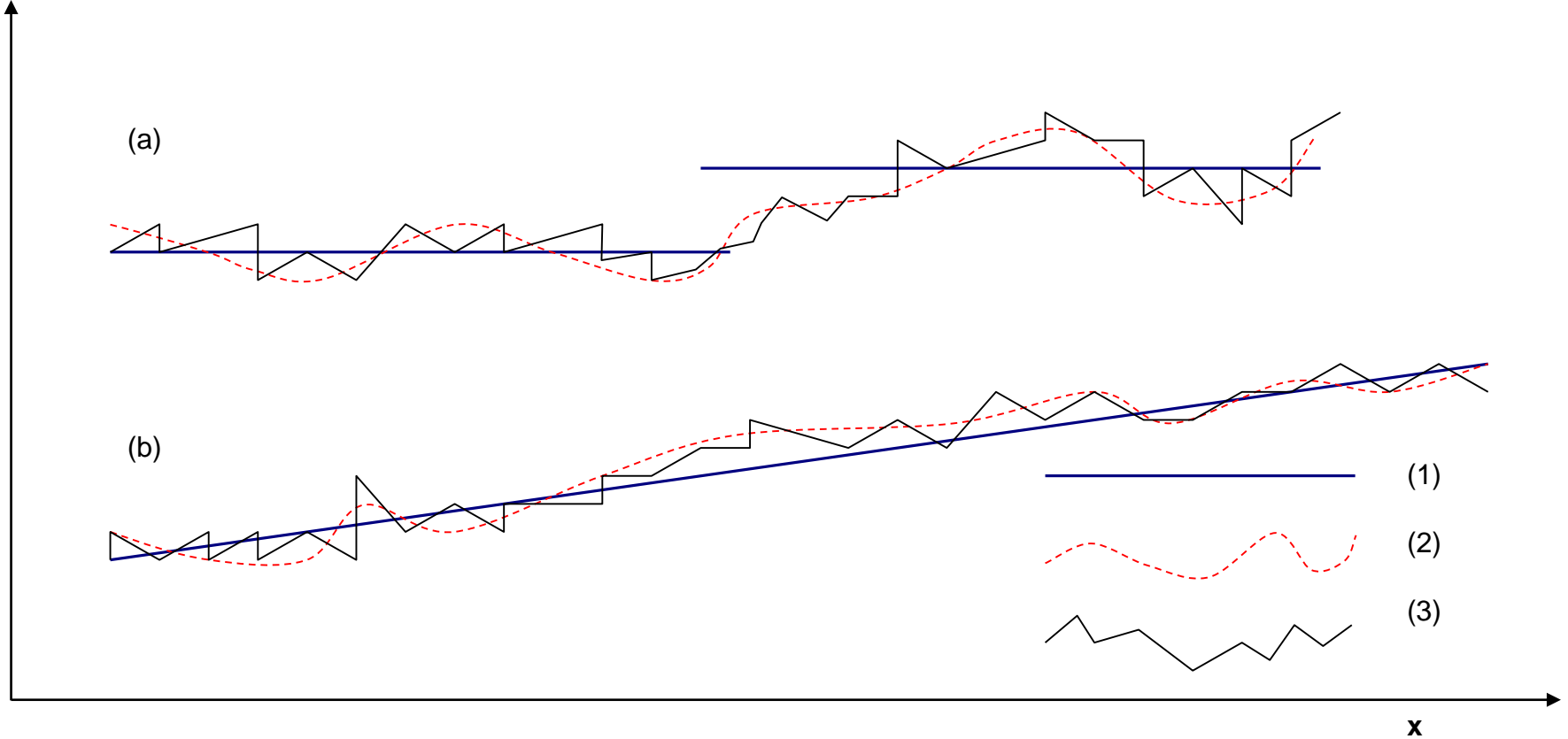
- una componente strutturale che individua l'andamento di fondo dei valori;
- una componente casuale ma spazialmente correlata nota come la variazione della variabile regionalizzata;
- un rumore di fondo o errore residuo spazialmente incorrelato.

Kriging

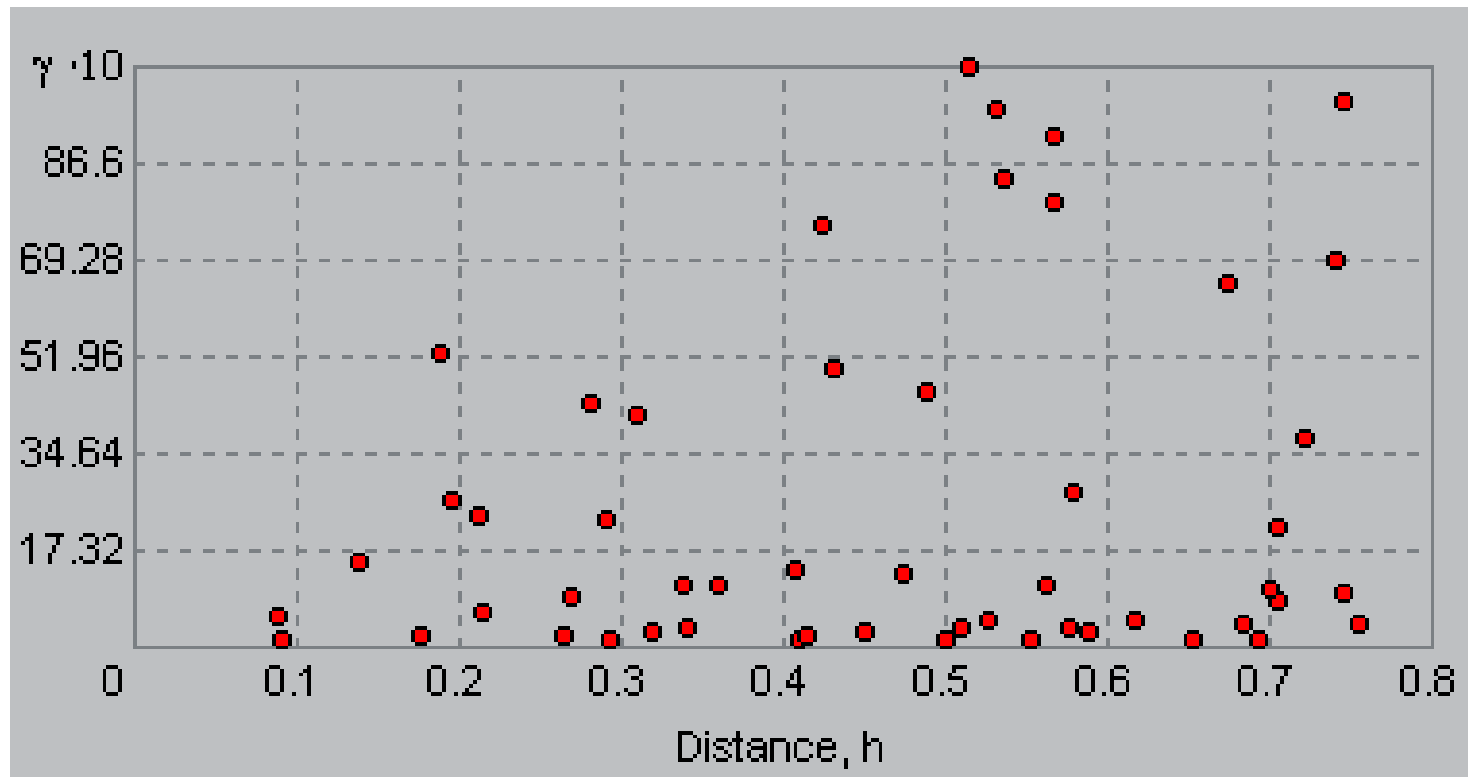
Sia \mathbf{x} il vettore che individua la posizione dei punti in una, due o tre dimensioni a seconda dei casi. Allora il valore di una variabile casuale Z nel generico punto \mathbf{x} è dato da:

$$Z(\mathbf{x}) = m(\mathbf{x}) + \varepsilon'(\mathbf{x}) + \varepsilon''$$

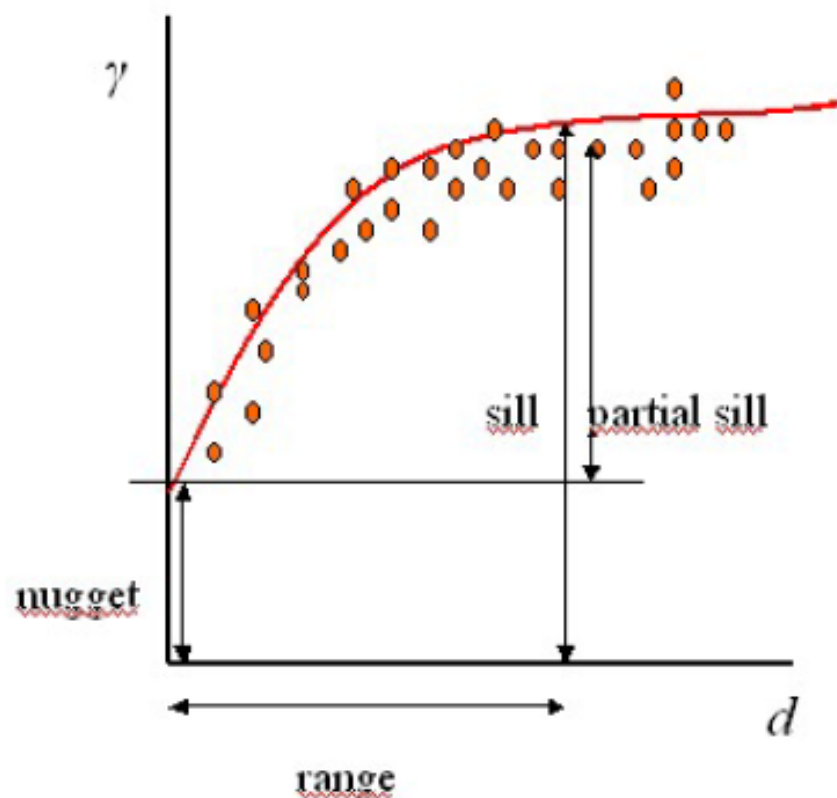
dove $m(\mathbf{x})$ è una funzione deterministica che descrive la componente strutturale di Z in funzione delle coordinate spaziali, $\varepsilon'(\mathbf{x})$ è il termine che denota le casuali ma spazialmente dipendenti differenze da $m(\mathbf{x})$ e ε'' è l'errore residuo, spazialmente indipendente.



La teoria della variabile regionalizzata divide la complessa variazione spaziale in un andamento di fondo (1), che si esplicita nel caso più semplice in differenze nei livelli medi (a) o altrimenti nella presenza di un trend (b), in una variazione spazialmente correlata (2) ma irregolare (casuale) e in una variazione locale (3) casuale ed incorrelata causata da errori di misura e variazioni spaziali su piccole distanze.



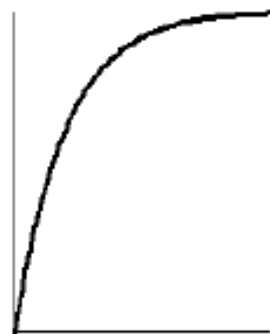
- ✖ Dall'analisi del variogramma e' possibile estrarre parametri d'interesse che ci aiutano a capire la distribuzione della variabile sul territorio:



- ✖ **Nugget:** descrive il livello di variabilità casuale;
- ✖ **Partial Sill:** descrive il livello di variabilità spaziale, cioè quella funzione della varianza che si osserva in funzione della distanza
- ✖ **Range:** distanza massima alla quale si osserva correlazione spaziale

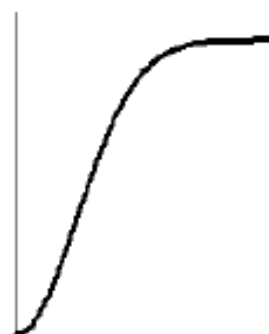
Interpolazione del variogramma

- Una volta costruito il variogramma empirico si cerca il modello che meglio lo approssima tra i modelli standard
- In questo modo otteniamo una legge che descrive l'andamento della nostra variabile sul territorio a scala globale (in tutta la zona coperta dai campioni)



Exponential Model
Cressie (1991, p. 61)

$$\gamma(h) = C[1 - e^{-h}]$$



Gaussian Model
Ponratier (1996, p. 50)

$$\gamma(h) = C[1 - e^{-h^2}]$$

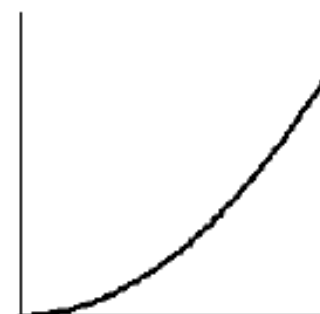


$$0 < h < 1$$

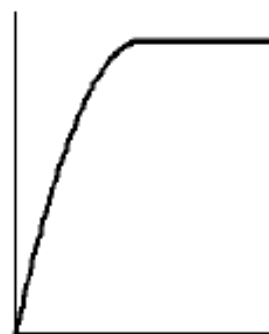
Power Models
Ponratier (1996, p. 51)

$$\gamma(h) = C|h|^n$$

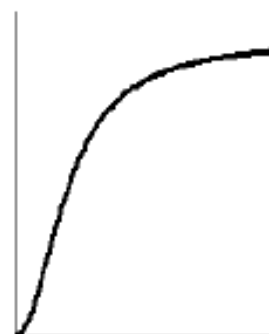
where $0 < n < 2$
where 1 is a Linear Model



$$1 < h < 2$$



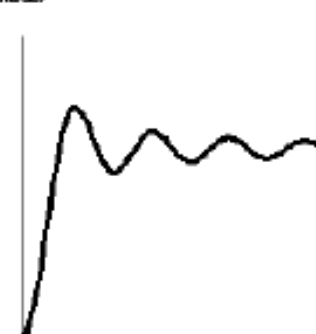
Quadratic Model
Alfaro (1980, p. 31)



Rational Quadratic Model
Cressie (1991, p. 61)

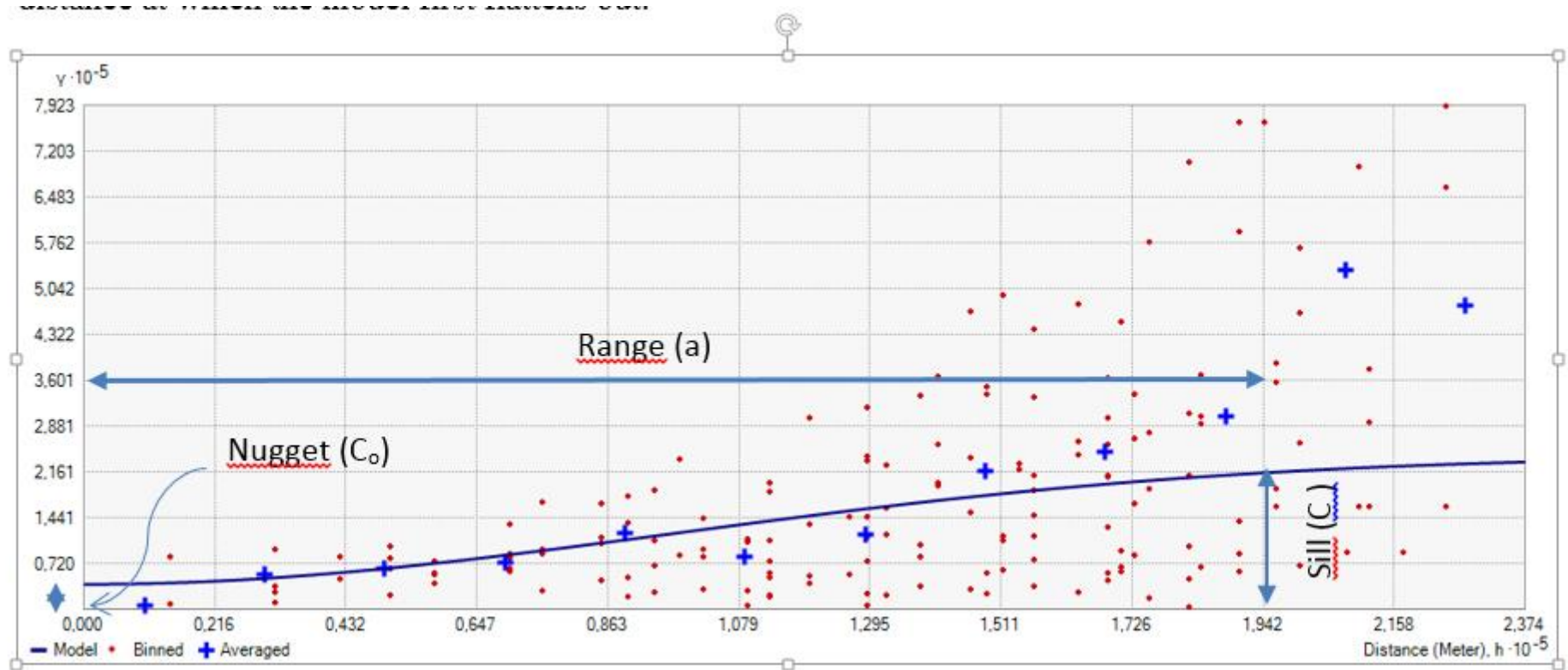


Linear Model
Kitanidis (1997, p. 61)



Wave (Hole Effect) Model
Cressie (1991, p. 62)

Esempio di semivariogramma



Esempi di alcuni modelli di variogramma

Formule

In riferimento alla figura precedente, si ha:

A linear variogram model is as follows:

$$\gamma(h) = C_0 + bh \quad (2)$$

An exponential variogram model is as follows:

$$\gamma(h) = C_0 + C[1 - e^{(-\frac{h}{a})}] \quad (3)$$

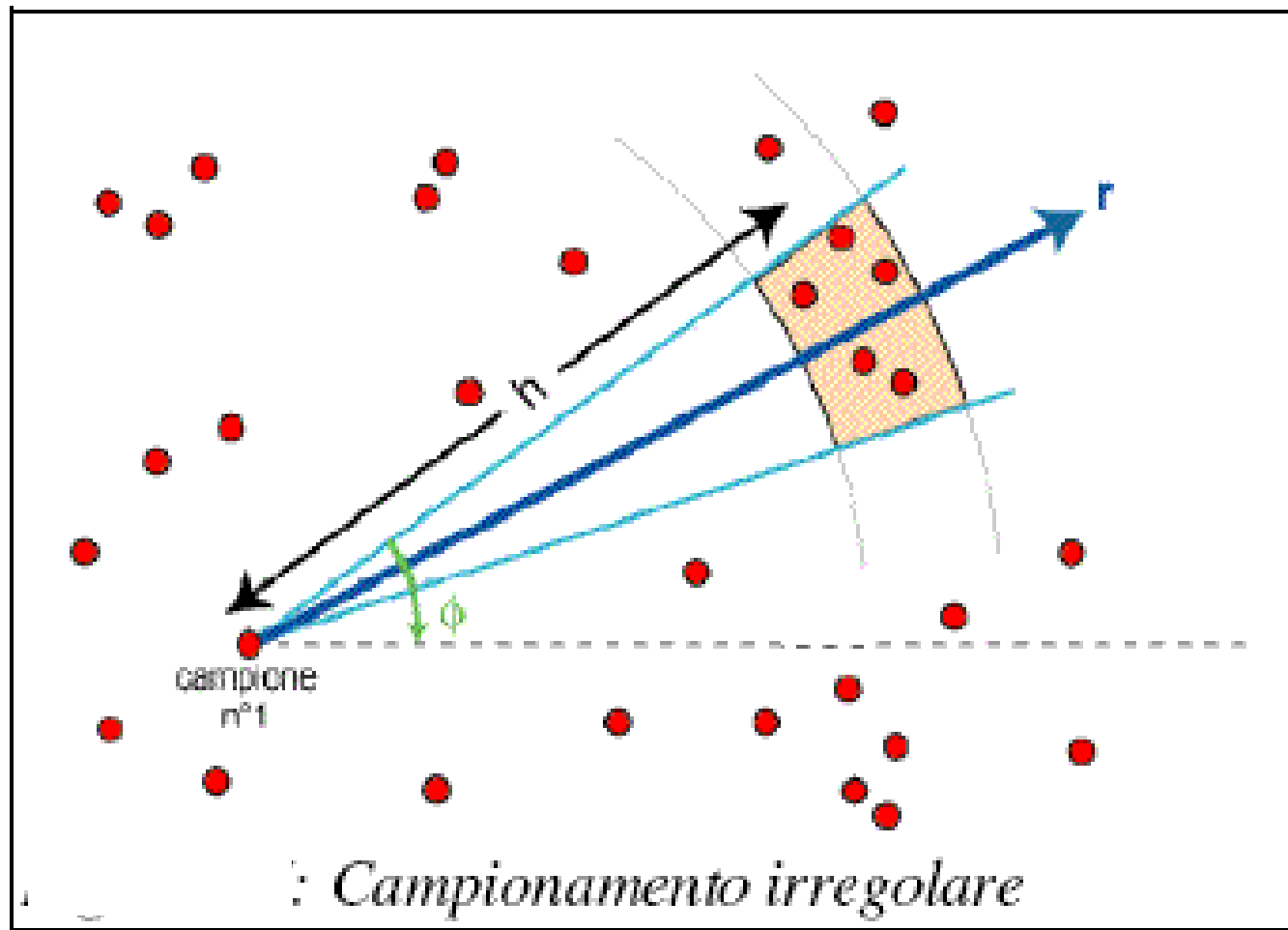
A Gaussian Model

$$\gamma(h) = C_0 + C[1 - e^{-(\frac{h}{a})^2}] \quad (4)$$

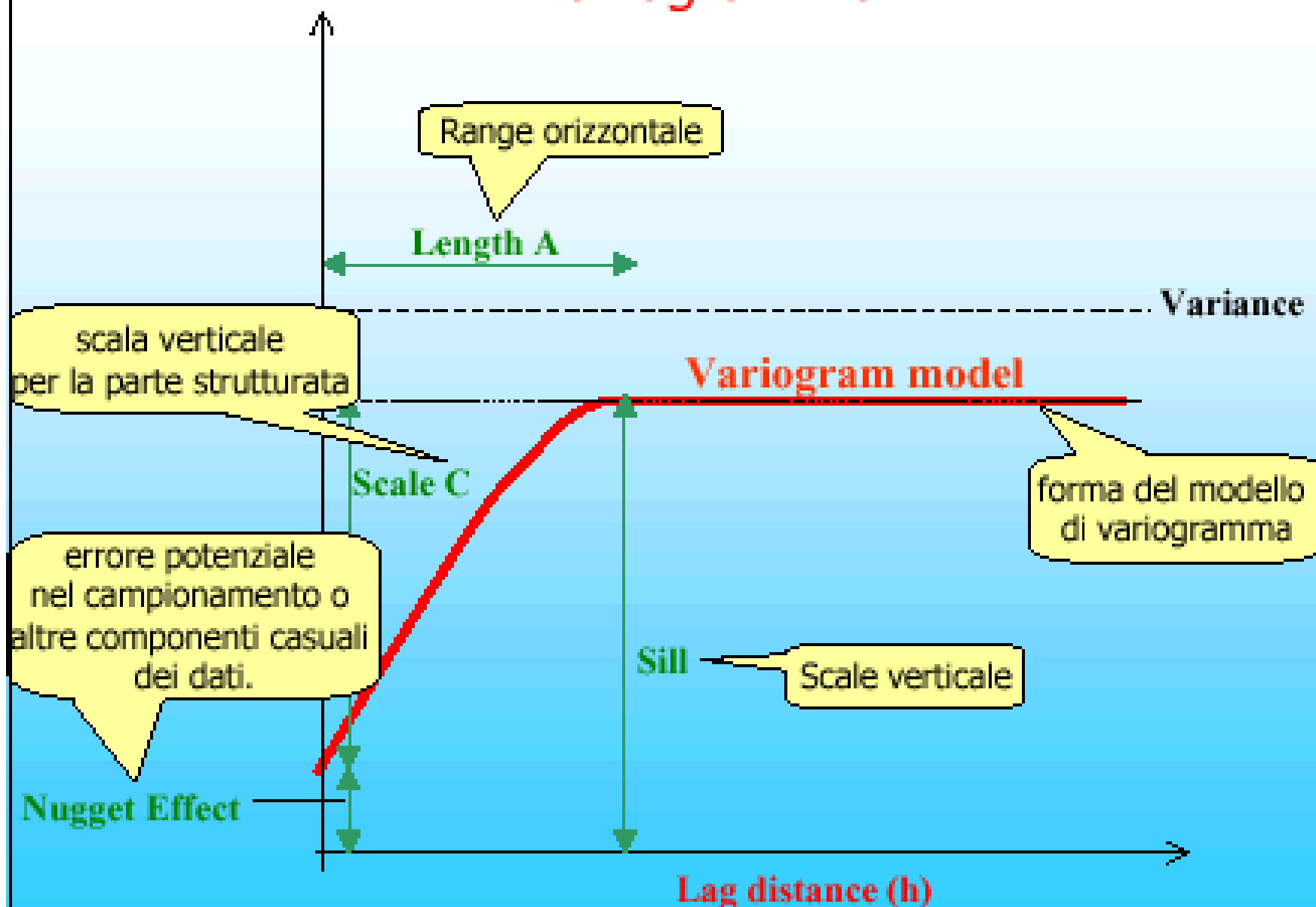
Costruzione del variogramma

- La procedura da seguire è questa:
- Si parte facendo la differenza tra i valori testa-coda di ciascun vettore $m(x)$. Le differenze saranno uguali a $m(x)$, ovvero al numero di coppie di campioni per questo lag
- Il risultato di ogni differenza si eleva al quadrato
- Si sommano tutti i quadrati
- Si divide questa somma per $2m(h)$, come da formula 6.1
- Si ripete la stessa procedura da 1 a 4 per il secondo lag (il doppio del primo)
- Si ripete la stessa procedura da 1 a 4 per il terzo lag (tre volte il primo)
- Si ripete la procedura fino all'ultima distanza di lag che si vuole.

- Purtroppo però non sempre si ha una maglia regolare di campionature, bisogna allora aggiustare la tecnica per il calcolo del variogramma. In questo caso infatti potrebbe accadere che in una certa direzione r non cada nessun punto campionato, perché disposti in modo irregolare.
- Per ovviare a ciò allora si considera una direzione r , individuata tramite un angolo a partire da una direzione di riferimento, con una tolleranza angolare Δ . Una tolleranza Δh deve anche essere data sulla distanza. Così facendo tutte le coppie aventi distanza compresa tra $h - \Delta h$ e $h + \Delta h$ e allineate secondo la direzione compresa tra $-\Delta\Phi$ e $+\Delta\Phi$ contribuiscono al calcolo del variogramma.



Il Variogramma



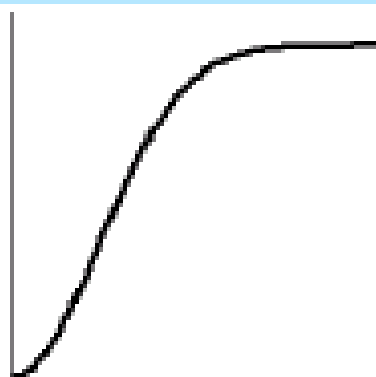
- Nugget: descrive il livello di variabilità casuale;
- Sill: valore massimo della semivarianza quando si ha stazionarietà (esso approssima per eccesso la varianza campionaria);
- Range: rappresenta la distanza massima entro la quale si manifesta correlazione tra semivarianza e *lag*;

Semi-variogramma

- Si adottano modelli matematici che di adattano meglio al variogramma dei nostri dati mathematical function
 - es. lineare, sferico, gaussiano, esponenziale, ecc

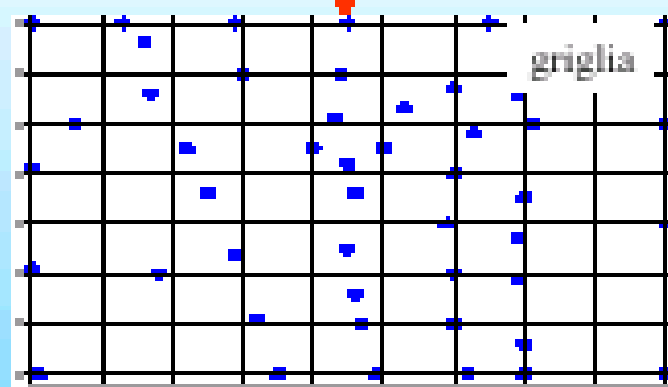
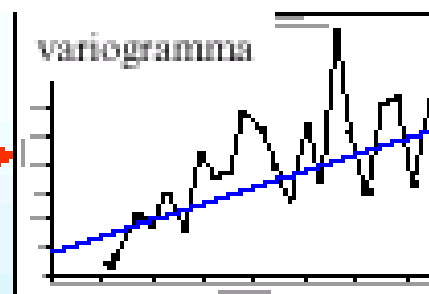
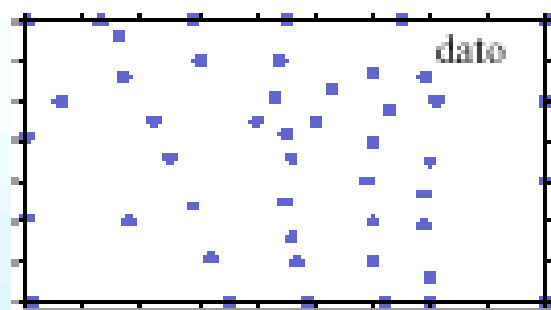


Exponential Model
Cressie (1991, p. 61)



Gaussian Model
Fornari (1996, p. 30)

Dal Variogramma alla ricostruzione spaziale del dato



Modello
raster

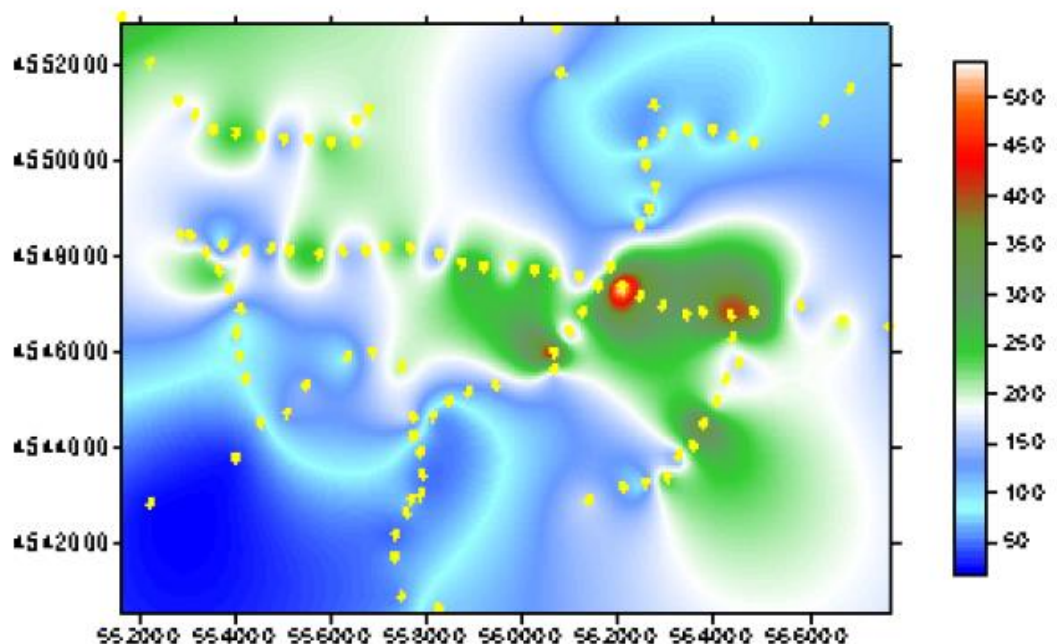


8
7. interpolazione

90,09	89,26	88,33	8	
78,89	77,57	76,24	7.	interpolazione
65,42	64,09	62,78	61,49	60,22
53,36	52,37	51,47	50,64	49,91
49,28	50,03	50,94	51,96	53,02
59,06	59,74	60,23	60,44	60,31
54,01	52,63	51,26	49,93	48,89
46,40	47,18	48,12	49,11	50,06
51,14	50,43	49,57	48,58	47,50
39,77	38,36	36,92	35,49	34,06

Il Kriging risulta particolarmente adatto e mostra i suoi vantaggi rispetto ad altri stimatori in situazioni in cui la **geometria di campionamento** si presenta **irregolare**. Da quanto esposto si capisce inoltre che la struttura di continuità è fondamentale nella definizione dei pesi per la stima.

Mappa di suscettività magnetica ottenuta attraverso una procedura di kriging spaziale



Kriging

Non esiste un solo metodo di Kriging, ma più versioni (Ordinary Kriging, Universal Kriging, ...).

In definitiva, la modalità operativa è la seguente: definito lo scatterogramma, la quota calcolata in ciascun nodo viene adattata in maniera tale da far sì che lo scatterogramma dei punti calcolati ricalchi quello dei punti misurati (o meglio sia simile a quello ottenuto con un modello matematico interpretativo dello scatterogramma, come quello esponenziale, gaussiano, ecc.).

Come si verifica l'attendibilità dei risultati ottenuti con il Kriging?

Come per i DTM ottenuti con metodi deterministi, l'attendibilità dei risultati del Kriging può essere valutata vedendo quali residui il modello produce (nella direzione delle z ovvero delle profondità) su punti noti non utilizzati per il processo di interpolazione.

In alternativa, si può utilizzare la Cross Validation.

Cross Validation

- Si toglie un punto dalla serie di partenza e si calcola la quota nel punto tolto, utilizzando le quote degli altri punti e l'algoritmo prescelto (ad esempio ordinary kriging, universal kriging, ecc.)
- Si calcolano tante quote quanti sono i punti, cioè si toglie un punto alla volta e si calcola ivi la quota;
- Si controlla la differenza tra il singolo punto misurato e il valore ivi predetto (calcolato);
- Si considerano i valori statistici di tali residui (es: RMS) e si decide, in base a tali valori, se il grid è buono o no;
- Se i residui sono considerati accettabili, il grid viene realizzato;
- Attenzione: il grid finale è diverso da quelli intermedi utilizzati solo per vedere la differenza tra misurato e predetto: quello finale è calcolato su tutti i punti, quelli intermedi hanno ciascuno un punto in meno (ad ognuno ne manca uno).