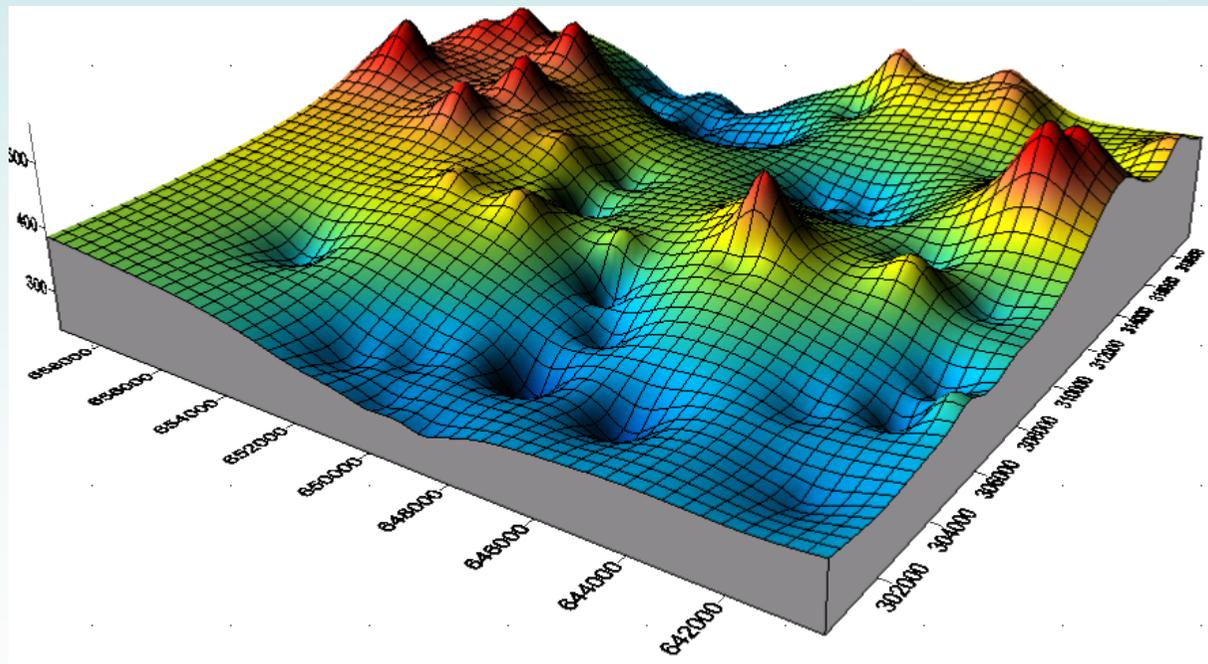


Geostatistica e interpolazione spaziale

Valerio Noti



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

Geostatistica e interpolazione spaziale

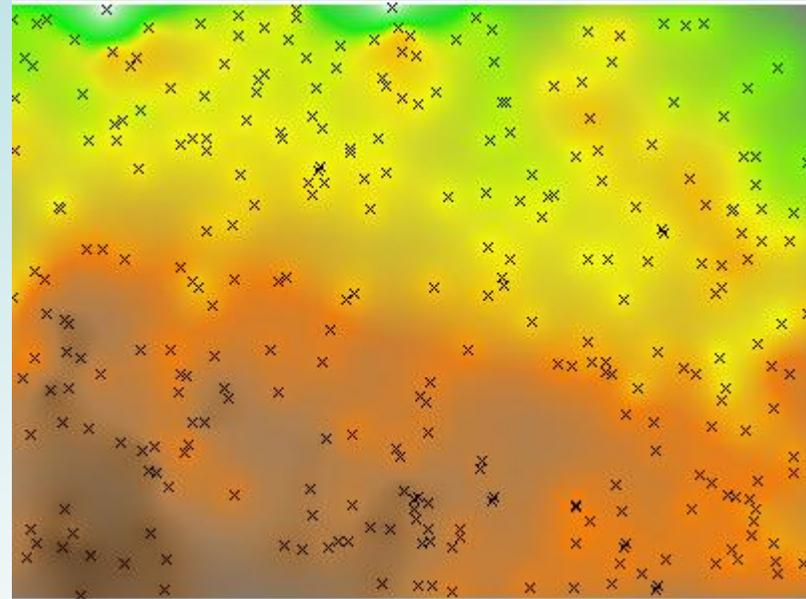
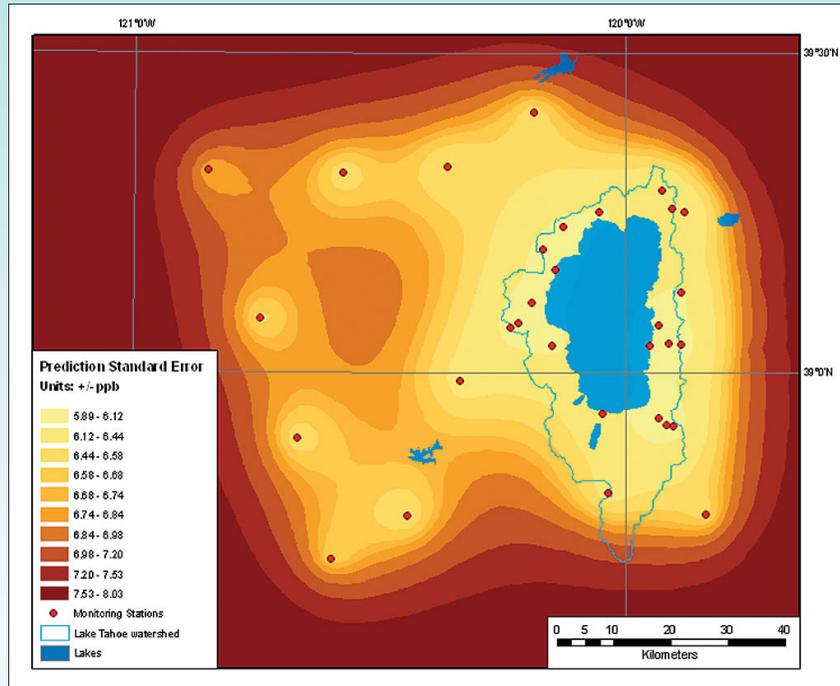
- Concetti di base e cenni storici
- Autocorrelazione spaziale
- Interpolatori deterministici
- Flusso di lavoro geostatistico
- Semivariogramma
- Kriging (cenni)
- Software geostatistici
- Tematizzazione cartografica dei dati quantitativi
- Modelli digitali del terreno: metodi geostatistici o deterministici?

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

Tipiche domande a cui la geostatistica tenta di rispondere:

- Come varia una variabile nello spazio?
- Che cosa controlla la sua variazione nello spazio?
- Quali e quanti campioni sono necessari per descrivere la sua variabilità spaziale?
- Qual è il valore della variabile in una specifica posizione (predizione)?
- Qual è l'incertezza di questa stima?

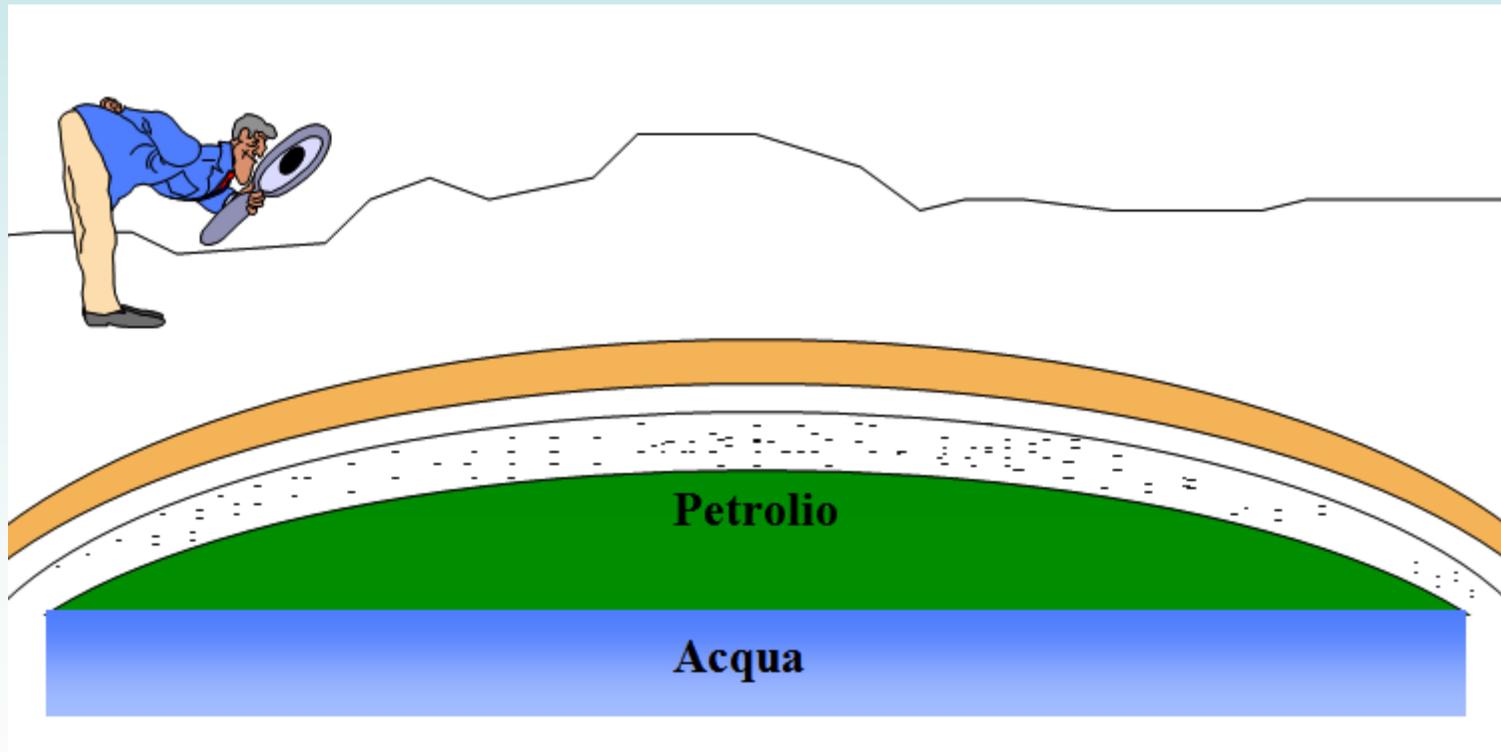
Perchè usare la geostatistica?



E' soprattutto uno strumento di pianificazione e supporto decisionale

Per molti decenni è stata applicata unicamente all'esplorazione mineraria e petrolifera come strumento di predizione.

Scopo: la stima migliore di un valore sconosciuto in una specifica posizione all'interno di un deposito minerario (best estimation of the unknown value at some location)



Tom Sheeran

Un po' di storia (anni 50-60)

D. G. Krige (1919): ingegnere minerario sudafricano. Osservò che poteva migliorare le stime di concentrazione dei minerali tenendo in considerazione quelle dei blocchi adiacenti. Notò quindi una **autocorrelazione** e la sua tecnica divenne presto una **consuetudine nella ricerca dell'oro**.

G. Matheron (1930-2000), matematico dell'Ecole des Mines di Parigi: la sua tesi di dottorato (*Les variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature, 1965*) formulò la trattazione teorica delle osservazioni di Krige e divenne il testo fondamentale della Geostatistica (**Teoria delle Variabili Regionalizzate**).

La tecnica di stima *krigeage* (in francese e tradotta in *kriging* in inglese) è stata denominata da Matheron in onore di Krige.

Il verbo *to krige* viene spesso utilizzato come sinonimo di interpolazione.

Sono dovuti passare diversi decenni per vedere la geostatistica applicata in altri settori (es. ambientale, sociologico, ecc..)

Cause:

- La complicata rappresentazione matematica e teorica
- La mancanza di integrazione con i software di informazione geografica
- L'elevato costo dei SW geostatistici fino alla metà degli anni 90

Le tecniche di stima sono ora utilizzate in tutte le situazioni in cui una variabile continua viene rappresentata con dati campionati in particolari posizioni nello spazio e nel tempo; in particolare quando un valore campionato, misurato, si presume sia influenzato dalla sua posizione e dalle sue relazioni con i valori vicini.

Principali campi di applicazione della geostatistica

- Scienze della Terra (idrogeologia, geologia applicata)
- Inquinamento e protezione ambientale
- Epidemiologia
- Agricoltura / scienze del suolo
- Ecologia, biologia, scienze ambientali in genere
- Ingegneria Civile
- Ingegneria petrolifera
- Ricerca mineraria
- Meteorologia e climatologia
- Modelli digitali del terreno
- Parametri socio-economici
- Distribuzione di specie animali e vegetali
-

Environmental features/topics	Common variables of interest to decision making	SRV	TV	VV	SSD	RSD
Mineral exploration: oil, gas, mineral resources	mineral occurrence and concentrations of minerals; reserves of oil and natural gas; magnetic anomalies;	*	-	★	*	*
Freshwater resources and water quality	O ₂ , ammonium and phosphorus concentrations in water; concentration of herbicides; trends in concentrations of pollutants; temperature change;	*	*	*	*	-
Socio-economic parameters	population density; population growth; GDP per km ² ; life expectancy rates; human development index; noise intensity;	*	*	-	★	★
Land degradation: erosion, landslides, surface runoff	soil loss; erosion risk; quantities of runoff; dissolution rates of various chemicals; landslide susceptibility;	*	*	-	-	★
Natural hazards: fires, floods, earthquakes, oil spills	burnt areas; fire frequency; water level; earthquake hazard; financial losses; human casualties; wildlife casualties;	★	★	-	*	★
Human-induced radioactive contamination	gamma dose rates; concentrations of isotopes; PCB levels found in human blood; cancer rates;	*	★	-	*	★
Soil fertility and productivity	organic matter, nitrogen, phosphorus and potassium in soil; biomass production; (grain) yields; number of cattle per ha; leaf area index;	★	*	*	*	*
Soil pollution	concentrations of heavy metals especially: arsenic, cadmium, chromium, copper, mercury, nickel, lead and hexachlorobenzene; soil acidity;	★	*	-	★	-
Meteorological conditions	temperature; rainfall; albedo; cloud fraction; snow cover; radiation fluxes; net radiation; evapotranspiration;	*	★	*	*	★
Climatic conditions and changes	mean, minimum and maximum temperature; monthly rainfall; wind speed and direction; number of clear days; total incoming radiation; trends of changes of climatic variables;	-	★	*	*	*
Global atmospheric conditions	aerosol size; cirrus reflectance; carbon monoxide; total ozone; UV exposure;	*	★	★	-	★
Air quality in urban areas	NO _x , SO ₂ concentrations; emission of greenhouse gasses; emission of primary and secondary particles; ozone concentrations; Air Quality Index;	★	★	★	★	-
Global and local sea conditions	chlorophyll concentrations; biomass; sea surface temperature; emissions to sea;	*	★	*	*	*

Le variabili ambientali

Table 1.1: Some common environmental variables of interest to decision making and their properties: SRV — short-range variability; TV — temporal variability; VV — vertical variability; SSD — standard sampling density; RSD — remote-sensing detectability. ★ — high, * — medium, - — low or non-existent. Approximated by the author.

GIS e Geostatistica

**Geostatistica e GIS hanno più o meno la stessa età
ma radici profondamente diverse**

Il **GIS** è nato come strumento per automatizzare la creazione di modelli esatti e deterministici del mondo in un contesto cartografico dominante, con particolare attenzione alla precisione geometrica del dato.

La **geostatistica** ha come scopo quello di fare predizioni in condizioni di incertezza e informazione limitata.

L'iniziale diffidenza **della comunità GIS verso la geostatistica** è dovuta a molti fattori:

- L'elevato costo dei SW geostatistici fino alla metà degli anni 90
- L'orientamento del settore verso l'esplorazione mineraria
- La complessa trattazione matematica che caratterizza la geostatistica
- La lingua del testo fondamentale di Matheron del 1965 (francese)

GIS e Geostatistica

A partire dagli anni 90 alcuni fattori hanno favorito la loro integrazione:

- **La crescente disponibilità di dati telerilevati e di strumentazione per il monitoraggio ambientale.**
- Software meno costosi o open source con interfacce intuitive
- Espansione dei campi di applicazione
- Utilizzo in ambito aziendale e, in generale, extra-accademico

La visione strettamente deterministica dei GIS è risultata limitata in un mondo caratterizzato da dati sempre più numerosi ma incompleti ed incerti.

Remote Sensing e Geostatistica

Le applicazioni geostatistiche utilizzano frequentemente dati telerilevati (es. topografici, idrogeologici, vegetazionali, antropici).

E' altrettanto frequente **l'utilizzo di tecniche geostatistiche nell'analisi di immagini telerilevate:**

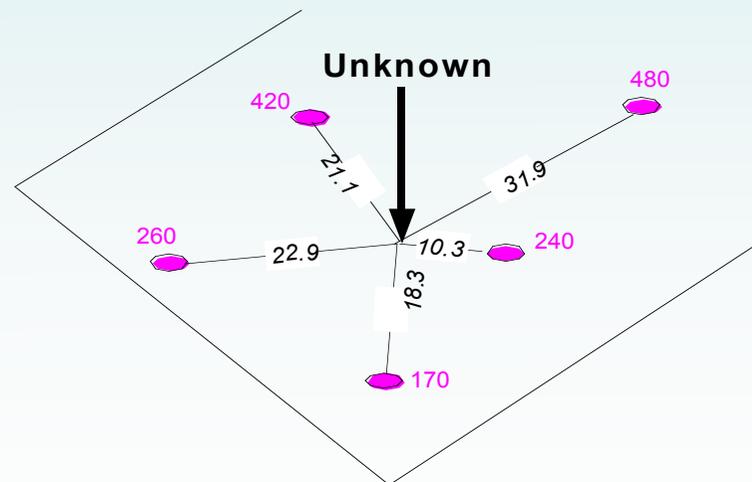
- Analisi di una variazione spaziale telerilevata
- Miglioramento dell'accuratezza delle interpretazioni (es. classificazione uso del suolo)
- Filtraggio immagini e valutazione rumore
- Miglioramento nella generazione di DEM ad alta e bassa risoluzione.
- Generazione DEM da dati LIDAR
- Ottimizzazione dei campionamenti
- Simulazioni di dataset multiscala e multitemporali

Le proprietà ambientali sono il risultato di azioni e interazioni di diversi processi e fattori; variano da luogo a luogo con grande complessità, in modi non lineari, a scale spaziali diverse (dai micrometri alle centinaia di chilometri).

L'ambiente è un dominio continuo nelle tre dimensioni e possiamo tentare di effettuare misure solamente in un numero finito di siti

La cosa migliore che possiamo fare è stimare (**predire**) in senso spaziale cercando di minimizzare gli errori

Questa è la principale ragione della geostatistica

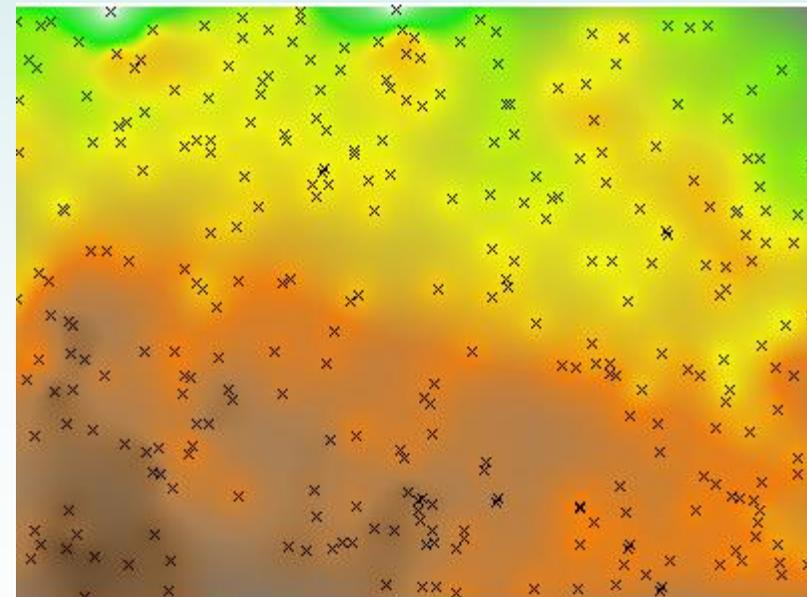
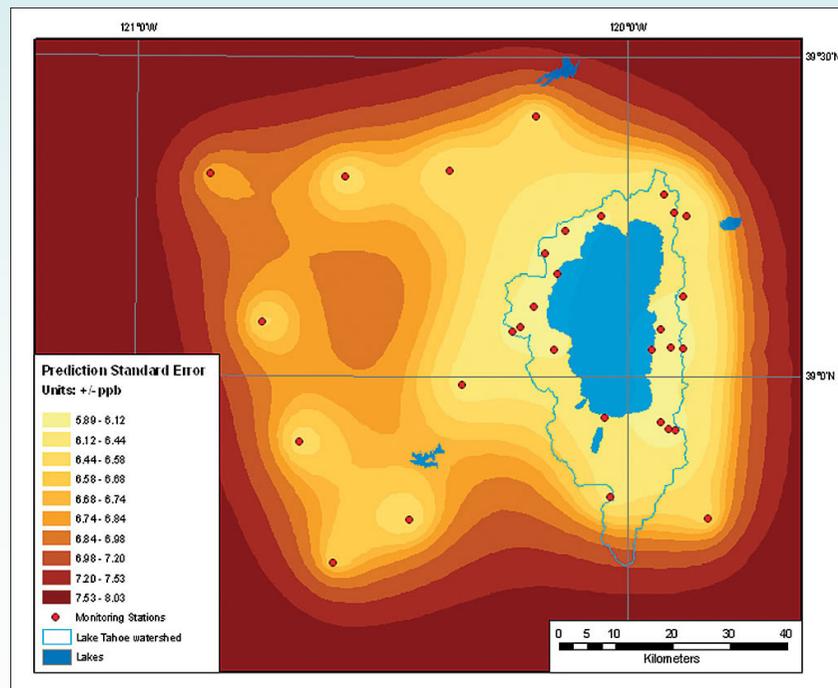


I campioni sono misure di un certo fenomeno di solito molto più piccoli dei volumi indagati. La distanza tra due campioni successivi è di vari ordini di grandezza superiore rispetto al diametro dei supporti stessi.

La Geostatistica è quella scienza che consente di quantificare il continuo spaziale tramite un dataset discontinuo estratto dallo stesso dominio.

Ci permette di far questo riducendo al minimo gli errori

unbiased estimates: in media la geostatistica è corretta nelle sue stime



\odot^5

A
•

\odot^3

\odot^1

\odot^2

\odot^4

Quali fattori di peso assegniamo?

Quanti campioni consideriamo?

Se ci fosse un campione 6 al doppio della distanza dal 5 lo dovremmo considerare?

Quanto è affidabile la stima che otteniamo?

⊙³

⊙⁵

A
•

⊙²

⊙¹

⊙⁴



Autocorrelazione spaziale

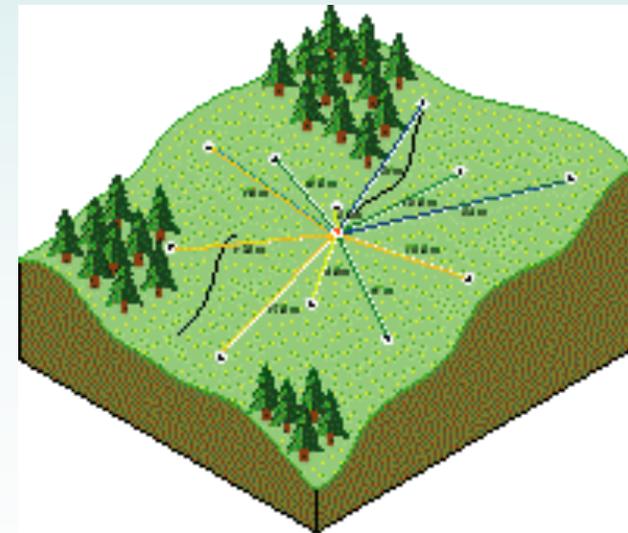
**"Everything is related to everything else,
but near things are more related than distant things"
(Tobler's First Law of Geography)**

I valori delle proprietà ambientali ad una certa scala sono positivamente relazionati (autocorrelati). **Punti vicini tra di loro tendono a valori simili, mentre punti lontani differiscono, in media, in misura maggiore.**

La geostatistica misura quanto oggetti spazialmente vicini siano simili.

La posizione del campione di cui si conosce la misura è importante quanto il valore misurato.

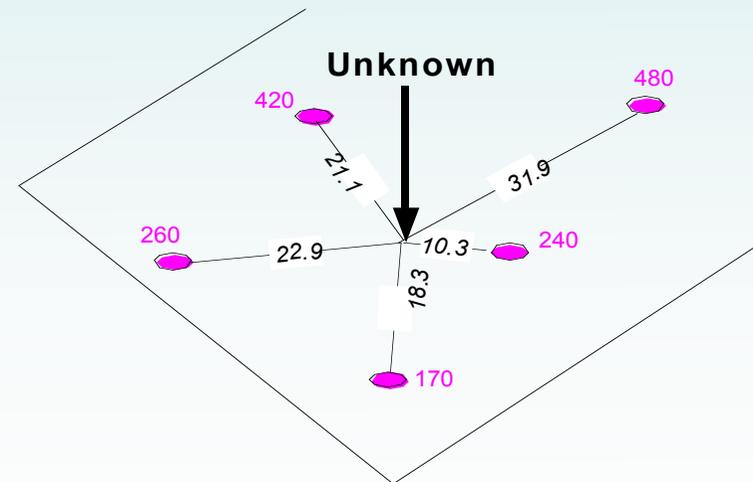
Introducendo l'autocorrelazione alcuni campioni saranno più importanti di altri e influenzeranno la predizione del valore di un punto sconosciuto.



Geostatistica

Obiettivi principali:

- Valutare la **variabilità spaziale** dei dati (presenza o meno di autocorrelazione)
- Applicare tecniche di **interpolazione spaziale** (predizione o stima locale di una variabile)



Interpolazione spaziale → predizione

La stima sul valore assunto da una variabile in una posizione in cui la misurazione non è stata effettuata

Attraverso tecniche di interpolazione di dati misurati puntuali si giunge ad una superficie statistica continua (CARTA DELLE PREDIZIONI) di solito in formato matriciale con risoluzione scelta dall'utente

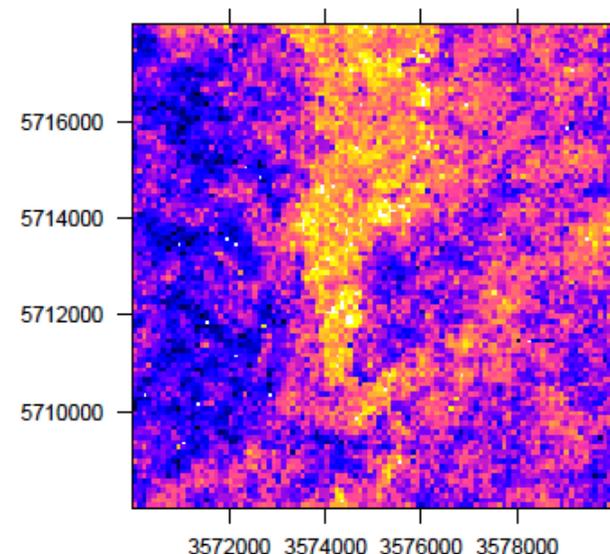


Fig. 1.2: If we were able to sample a soil variable over the whole area of interest, we would probably get an image such as this. This image was, in fact, produced using the geostatistical simulations with a regression-kriging model (see further §4.18).

Gridding

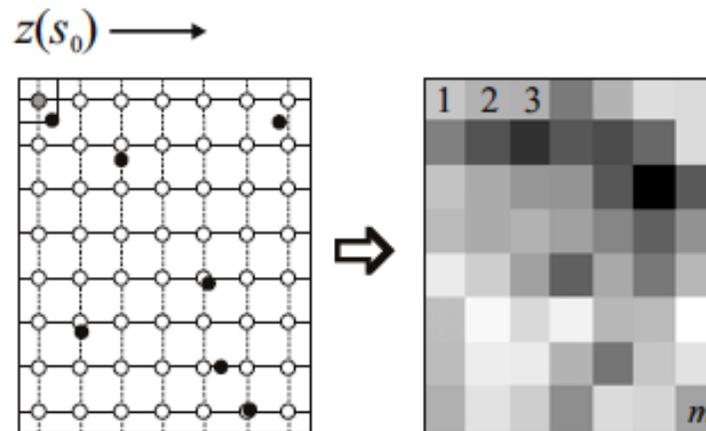


Fig. 1.6: Spatial prediction implies application of a prediction algorithm to an array of grid nodes (*point á point* spatial prediction). The results are then displayed using a raster map.

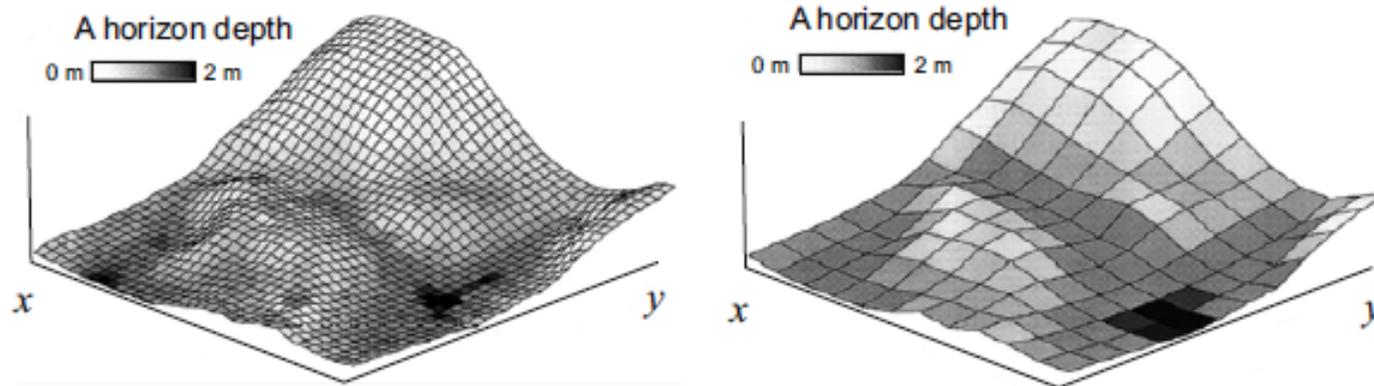


Fig. 1.4: Influence of the support (grid cell) size: predictions of the same variable at coarse grid will often show much less contrast. Example from Thompson et al. (2001).

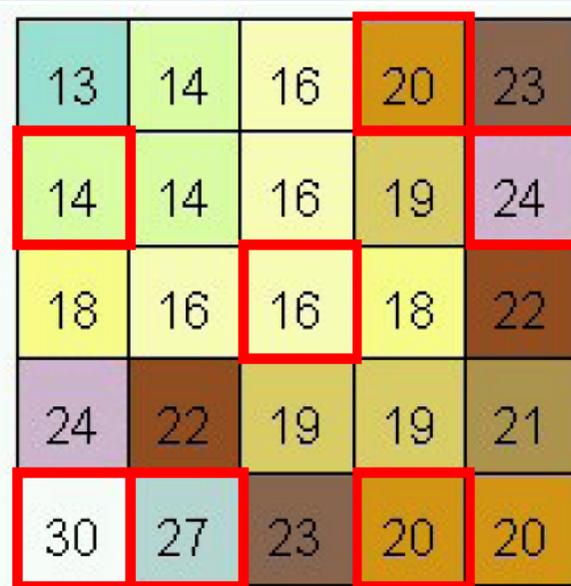
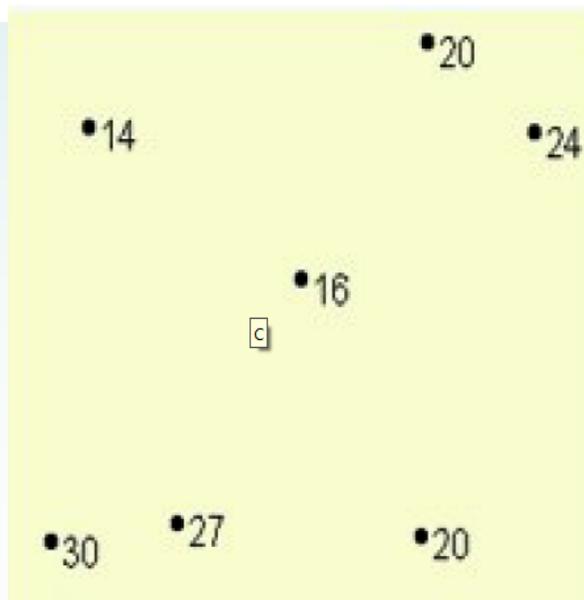
Gridding

Conversione di un campione di dati puntuali in una copertura completa dell'area di studio.

Quasi tutti i metodi di predizione (interpolazione) possono essere considerati medie pesate dei dati:

$$Z(x_0) = \sum_{i=1}^n \lambda_i Z(x_i)$$

x_0 = punto da predire
 $Z(x_0)$ = valore da stimare
 $Z(x_i)$ = valori misurati della variabile Z
 λ_i = pesi assegnati ai valori misurati



Gridding

La qualità del processo dipende:

Qualità dei dati di input

Densità

Distribuzione

Variabilità spaziale

Interpolatore applicato

Inserimento di breaklines

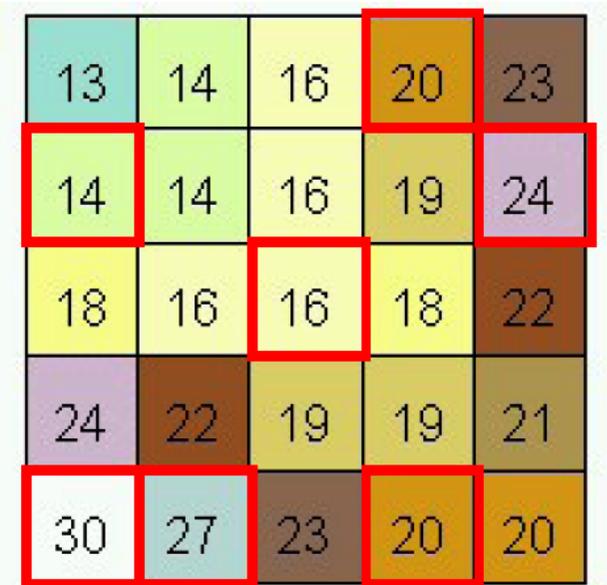
Scelta dei pesi

Definizione del numero dei punti oggetto di interpolazione

Forma e dimensioni dell'area di vicinato (neighbourhood)

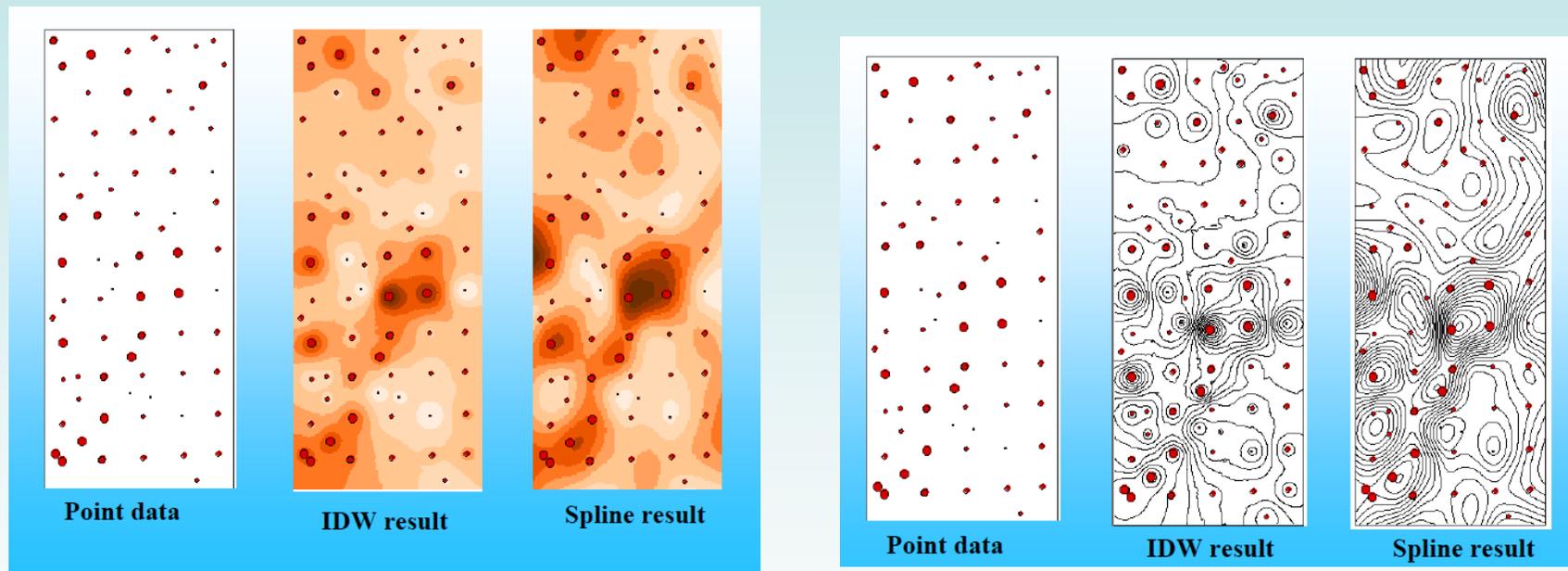
Valutazione della qualità

Disponibilità di dataset per cross-validation



Interpolazione spaziale

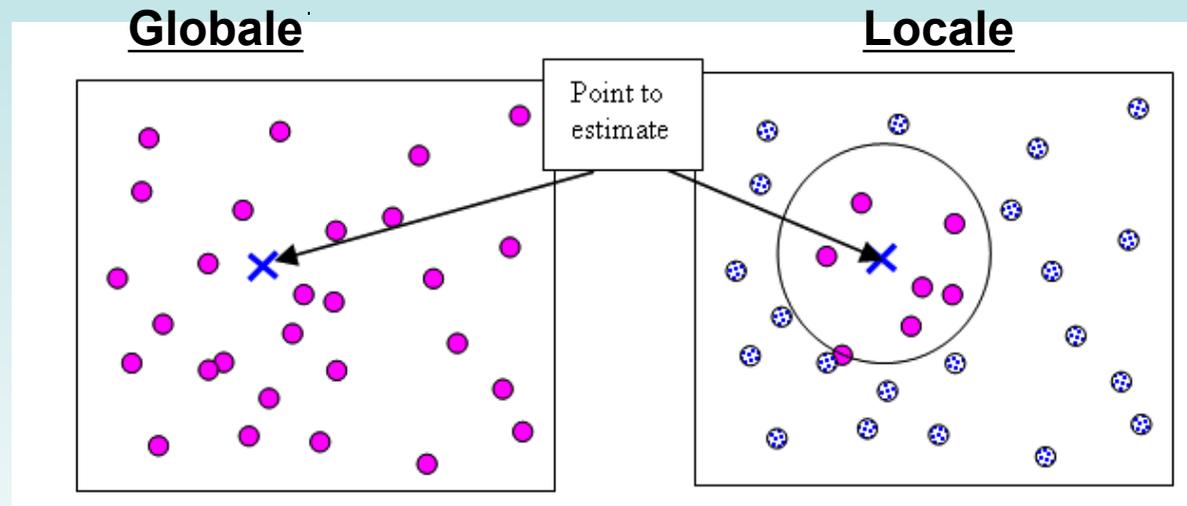
I risultati ottenuti da diversi algoritmi di interpolazione sono spesso enormemente diversi l'uno dall'altro.



Quali sono i processi e le caratteristiche implementate dagli algoritmi di interpolazione?

Interpolazione spaziale

Estensione della funzione



Vengono utilizzati tutti i dati misurati del dominio per la stima di un valore in un dato punto.

Una singola funzione matematica rappresenta l'intera regione.

La variazione di un singolo dato di input influenza l'intera superficie di output.

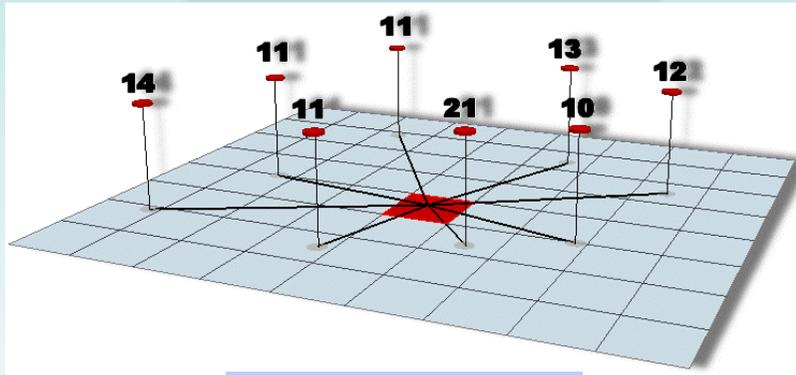
Vengono utilizzati alcuni dati misurati nelle vicinanze del punto da stimare (es. gli n punti più vicini oppure all'interno di un certo raggio)

Viene applicato l'algoritmo ripetutamente per ogni piccola porzione dell'intera regione. **Una variazione di un dato di input influenza l'output solamente all'interno della finestra di indagine locale.**

Interpolazione spaziale

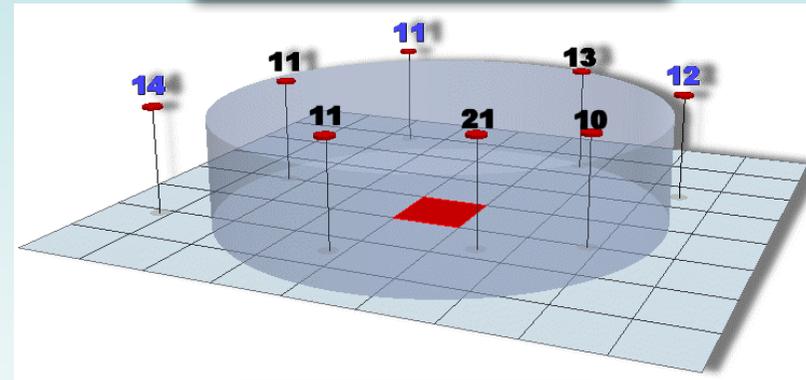
Raggio di ricerca

Variable



Samples = 8
Radius = ?

Fixed

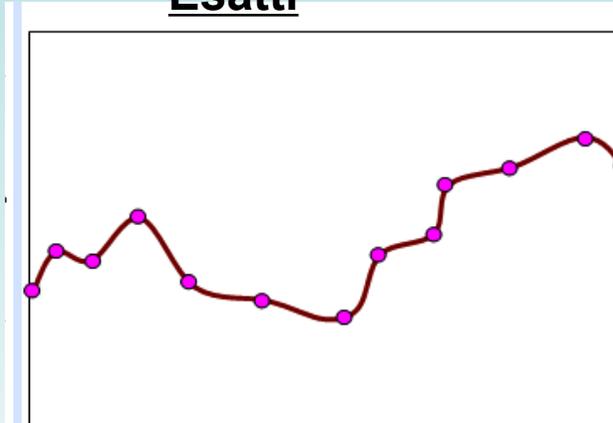


Radius =
1000
Samples = ?

Interpolazione spaziale

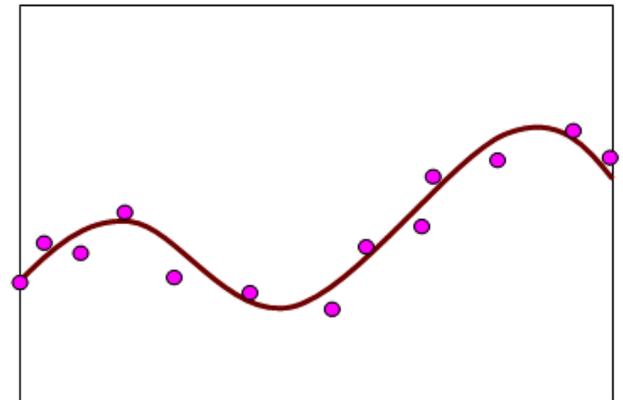
Approssimazione

Esatti



La superficie interpolata
passa attraverso le misure

Approssimati



La superficie interpolata NON
passa attraverso le misure.

Sono utilizzati quando si prevede
margine di incertezza sui dati
campionati.

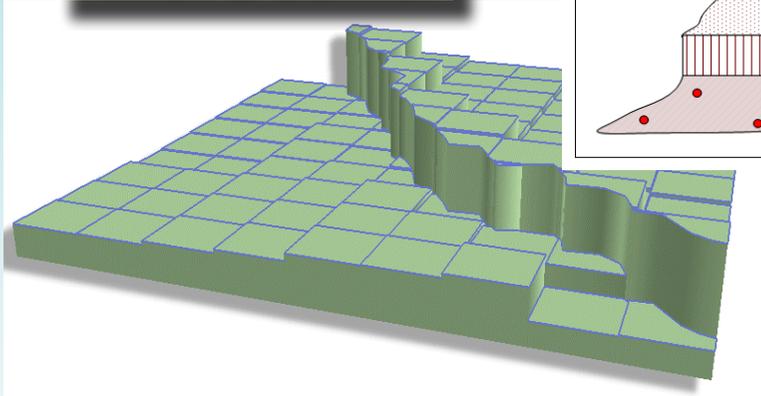
Le superfici risultanti sono
smussate

Interpolazione spaziale

Presenza di barriere

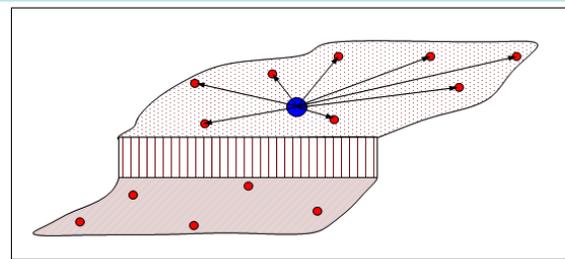
Bruschi

With a barrier



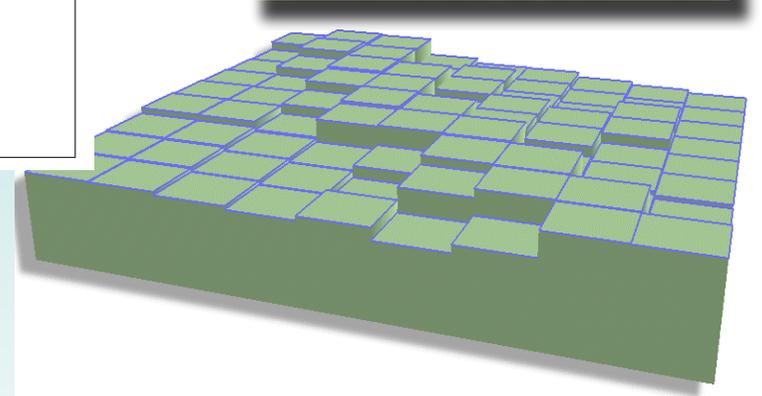
Il processo interpolativo prevede l'inserimento di barriere.

Producono superfici “a gradini” appropriati per dati con alta variabilità locale e discontinuità



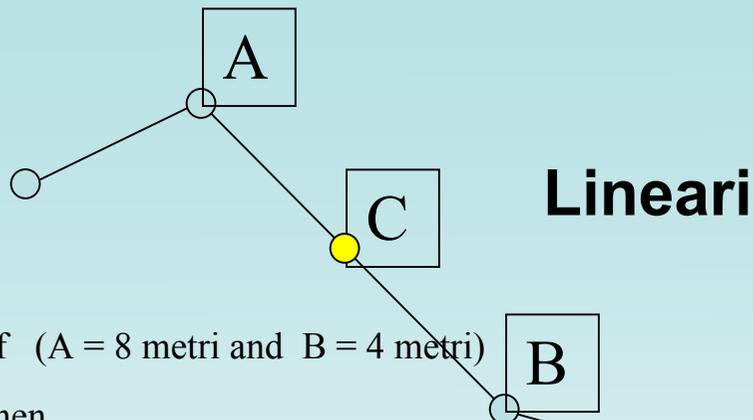
Graduali

Without a barrier



Producono superfici smussate, appropriati per dati con bassa variabilità locale.

Interpolazione spaziale



If (A = 8 metri and B = 4 metri)
then
 $C = (8 + 4) / 2 = 6$ metri



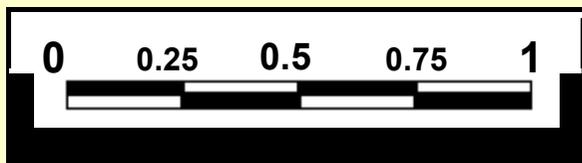
Porta spesso a risultati più
realistici ma la stima dei valori
interpolati è più complessa.

Es. Tecniche pesate

Known rainfall values (inches)



Interpolated rainfall values



1 Mile

Metodi di interpolazione spaziale

DETERMINISTICI

Mappa delle predizioni

- Non utilizza le **proprietà statistiche** dei punti misurati.
- Ogni valore predetto è il risultato di una funzione deterministica: **le variabili possono assumere in un determinato istante uno e un solo valore.**
- Sono utilizzati se sussistono sufficienti conoscenze sulla superficie da modellare.
- Non effettuano stime di errore sul risultato ottenuto

GEOSTATISTICI

Mappa delle predizioni

Mappa degli errori

- Utilizzano le **proprietà statistiche** dei punti misurati.
- Incorporano il **concetto di casualità**
- Sono molto **più complicati da implementare.**
- La superficie interpolata è concettualizzata come **una delle superfici possibili** che potrebbero essere osservate, tutte coerenti con i dati misurati.
- Queste tecniche producono anche superfici di errore o incertezza, dando indicazioni sulla capacità predittiva del metodo applicato.

A Comparison of the Geostatistical Analyst methods

Method	Deterministic/ Stochastic	Output Surface Types	Computing Time/ Modeling Time ¹	Exact Interpolator	Advantages	Disadvantages	Assumptions ²
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

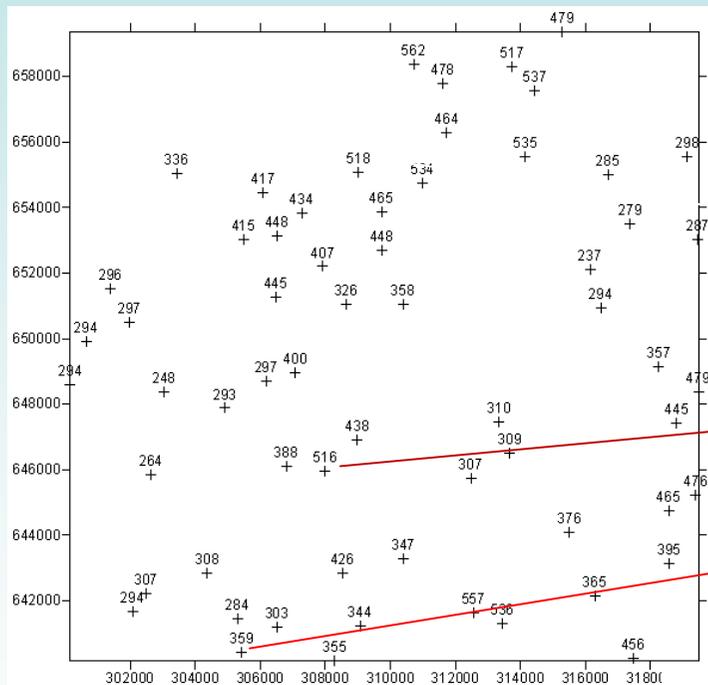
1. Computing time is computer-processing time to create a surface. Modeling time includes user-processing time to make decisions on model parameters and search neighborhoods.

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

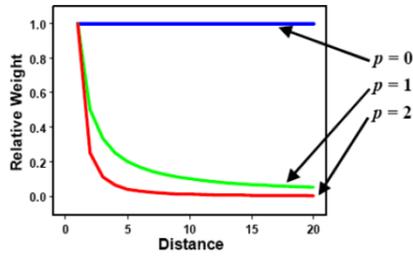
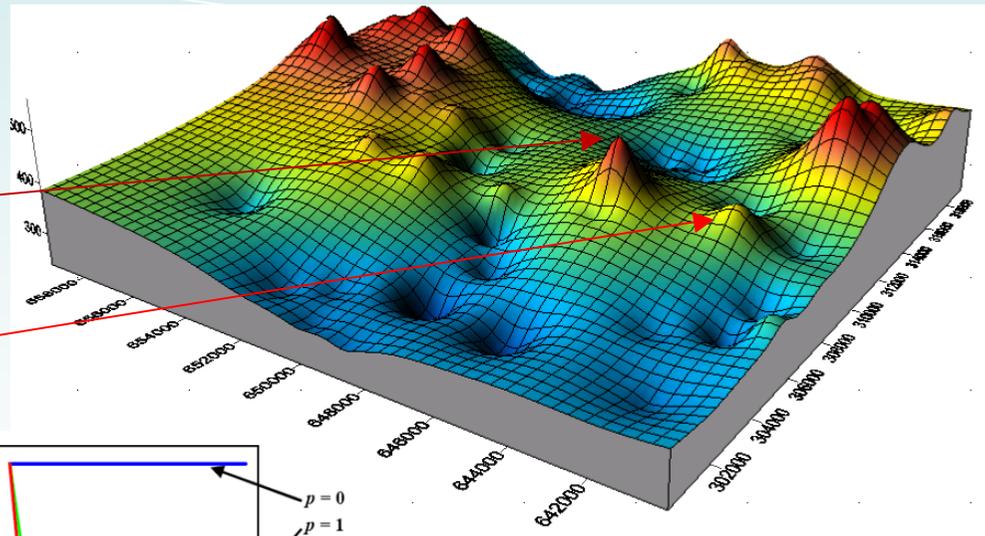
Metodi deterministici: Inverse Distance Weighted (IDW)

Il peso (influenza) di un dato di input è inversamente proporzionale alla distanza dal valore da predire

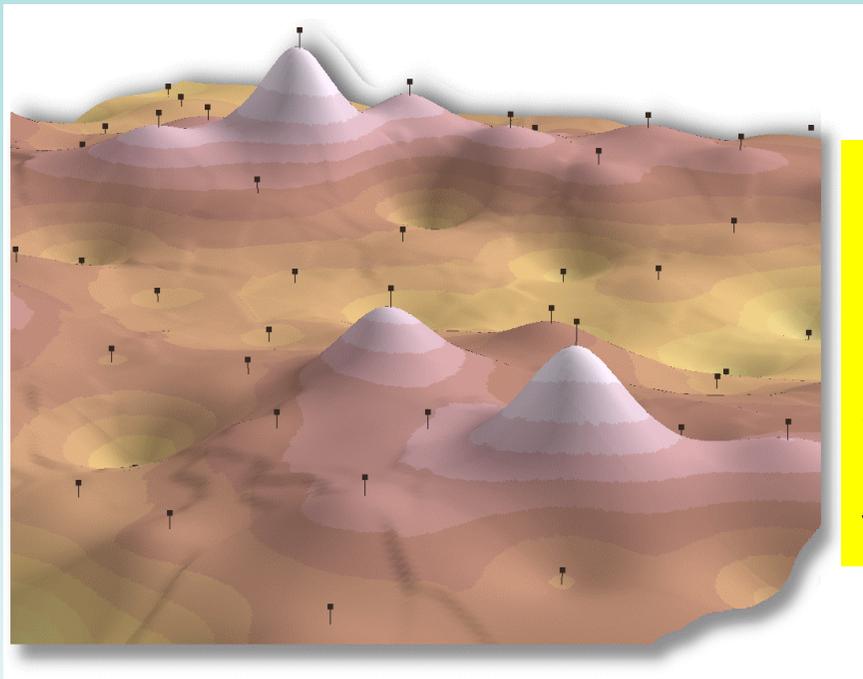
Locale (si può utilizzare un numero di punti predefinito di input oppure un raggio di ricerca);
Esatto (la superficie passa attraverso i punti misurati).



$$\hat{z}(x_0) = \frac{\sum_{i=1}^n z(x_i) \cdot d_{ij}^{-r}}{\sum_{i=1}^n d_{ij}^{-r}}$$



Inverse Distance Weighted (IDW)



Esempio di utilizzo: superficie di previsione di acquisti da parte dei consumatori. Negozi distanti hanno poca influenza (è più probabile che si acquisti vicino casa). Possono essere impostate barriere come ferrovie, viabilità principale, ecc..

- E' il metodo più popolare e semplice da implementare
- **Poche decisioni da prendere** e configurazioni da impostare per la sua applicazione
- Nessuna discontinuità interpolativa: i pesi non possono annullarsi, le superfici di output sono “apparentemente” realistiche, motivo che spiega il grande utilizzo di questo interpolatore.

Inverse Distance Weighted (IDW)

- **Scelta dei pesi arbitraria: difficoltà nello stabilire la potenza e il numero dei punti/raggio sorgente da interpolare**
- **Nessuna possibilità di valutazione di errore** (se non attraverso il confronto con altri dati sorgente)
- Può produrre superfici con numerosi picchi o buche **“hollow/bulls eyes”** intorno ai dati conosciuti.
- La disposizione spaziale dei campioni non è tenuta in considerazione: **punti isolati e cluster hanno la stessa importanza.**

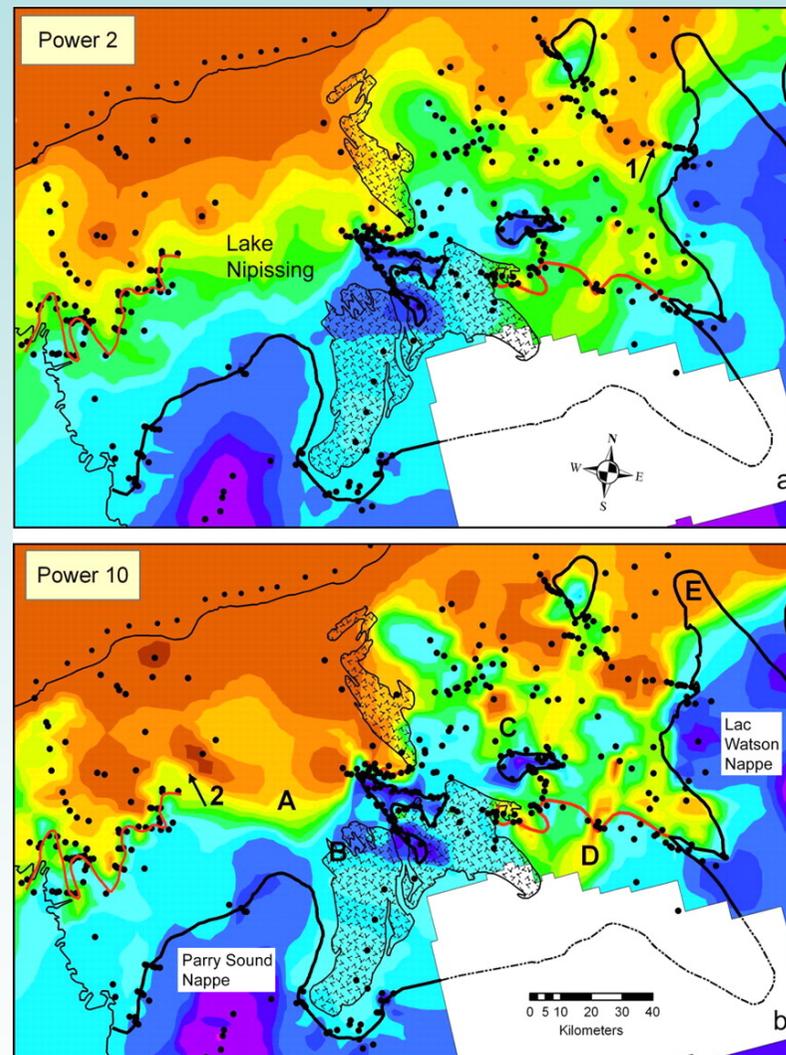


Diagramma di Voronoi

- *...oppure: tassellatura di dirichlet o decomposizione di voronoi o poligono di thiessen*

- **uno dei più vecchi e semplici metodi di interpolazione conosciuti**

Suddivide lo spazio in **poligoni adiacenti con numero variabile di lati** che racchiudono tutti i punti posti alla minima distanza dal centro del poligono.

- Ogni punto appartiene a un solo poligono ad eccezione dei punti di bordo che sono condivisi tra le poligoni adiacenti

- Definita **l'attrazione di un punto** sulla superficie (es. centro commerciale, ospedale più vicino ad una certa abitazione) come una grandezza proporzionale alla distanza dal punto, il **poligono di Voronoi** identifica **l'area contenente tutti i punti (le località geografiche) più vicini al punto generatore (es. centro commerciale)**.

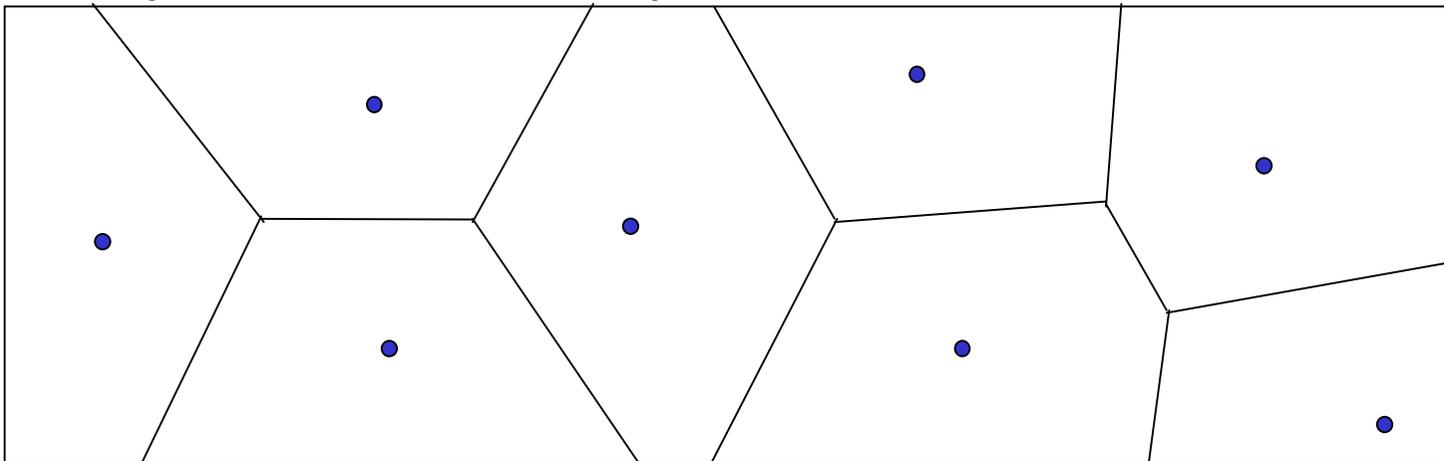
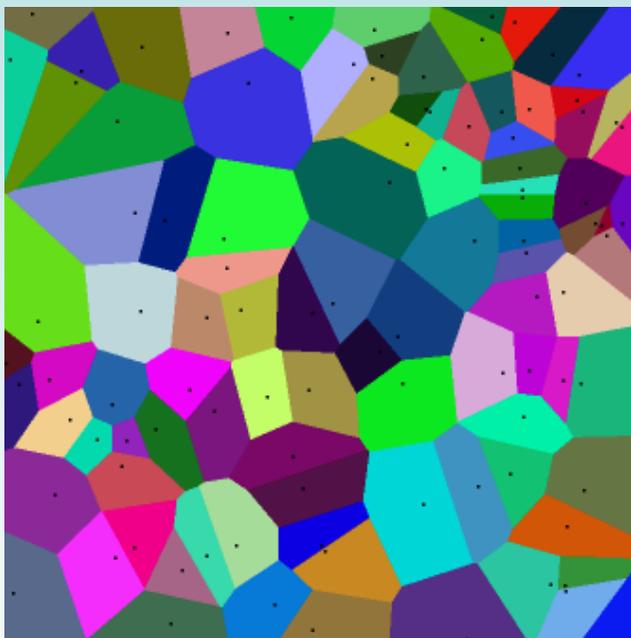


Diagramma di Voronoi



- Lo spazio viene suddiviso in poligoni o tasselli con bisettrici perpendicolari tra i punti campionati
- All'interno di ogni poligono qualunque posizione è più vicina al punto interno al poligono rispetto a qualsiasi altro punto all'esterno.
- Il valore di tutti i punti all'interno del poligono è uguale al valore del punto misurato.
- tutti i punti all'interno di ogni poligono assumono lo stesso peso mentre al di fuori del poligono viene attribuito un peso nullo.

Non c'è nessuna influenza dei valori vicini

Nessuna stima di errore

Il risultato finale è un mosaico....

TIN (Triangulated Irregular Network)

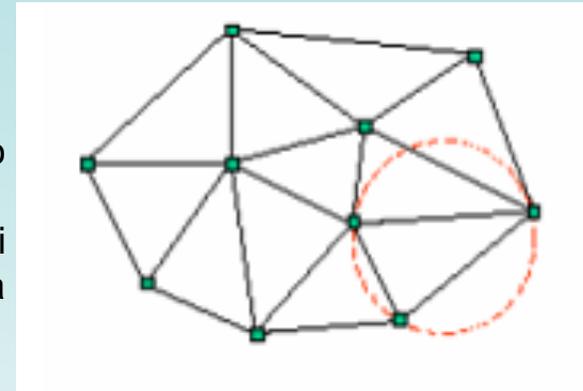
Metodologia di interpolazione vettoriale.

Una rete di triangoli adiacenti, non sovrapposti calcolati a partire da punti irregolarmente spazati

Il TIN viene costruito rispettando la **regola di Delaunay**:

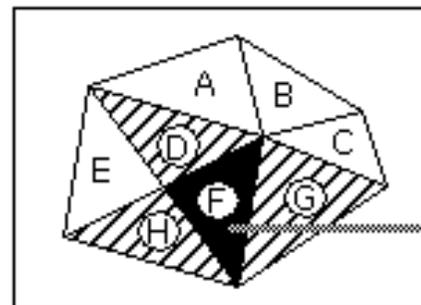
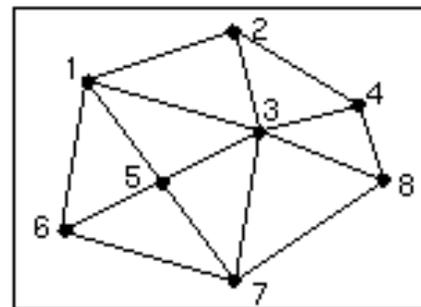
Il cerchio che passa per i 3 vertici di un triangolo non contiene alcun altro vertice della rete triangolata.

C'è un teorema che dimostra che comunque siano dati un insieme di punti su un piano esiste sempre almeno una triangolazione che verifica la regola di Delaunay.



Le faccette triangolari del TIN rappresentano visivamente e concettualmente un triangolo orientato nello spazio.

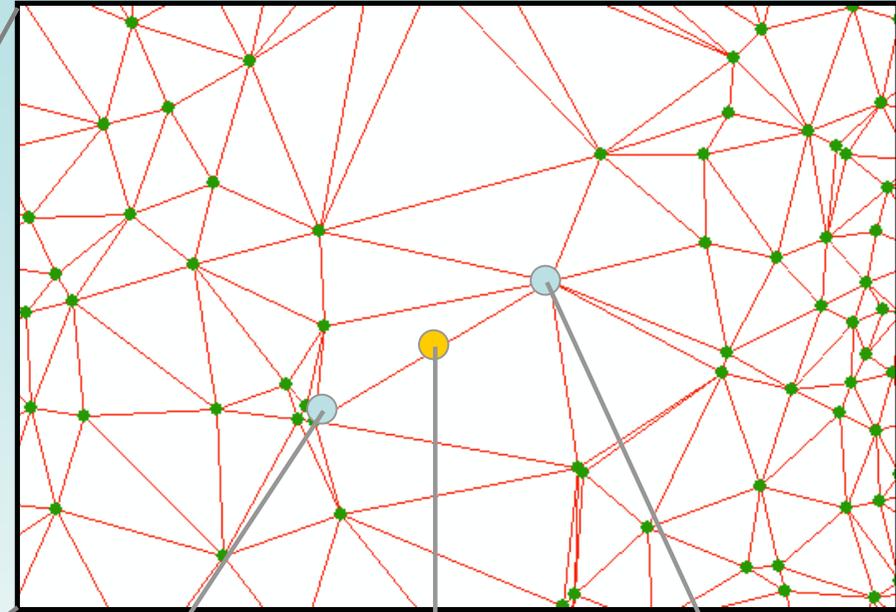
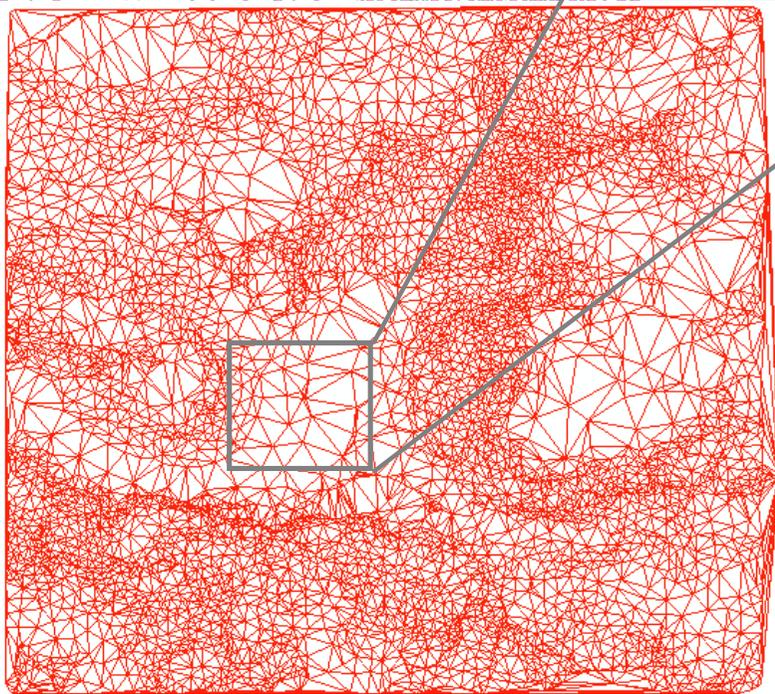
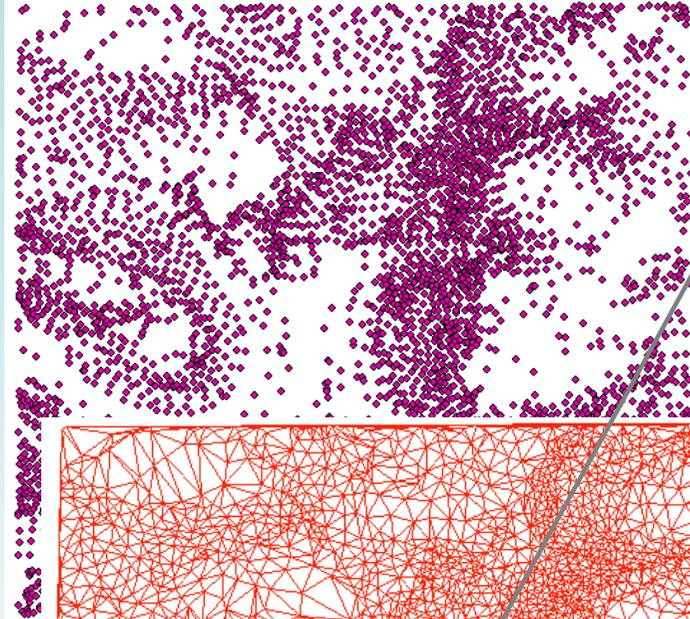
Il valore di un punto interno ad un qualsiasi triangolo viene calcolato con interpolazione lineare dei vertici del triangolo.



TIN Triangle List

Triangle	Node List	Neighbors
A	1, 2, 3	- B, D
B	2, 4, 3	-, C, A
C	4, 8, 3	-, G, B
D	1, 3, 5	A, F, E
E	1, 5, 6	D, H, -
F	3, 7, 8	G, H, D
G	3, 8, 7	C, -, F
H	5, 7, 6	F, -, E

TIN (Triangulated Irregular Network): la superficie di output contiene i dati di input



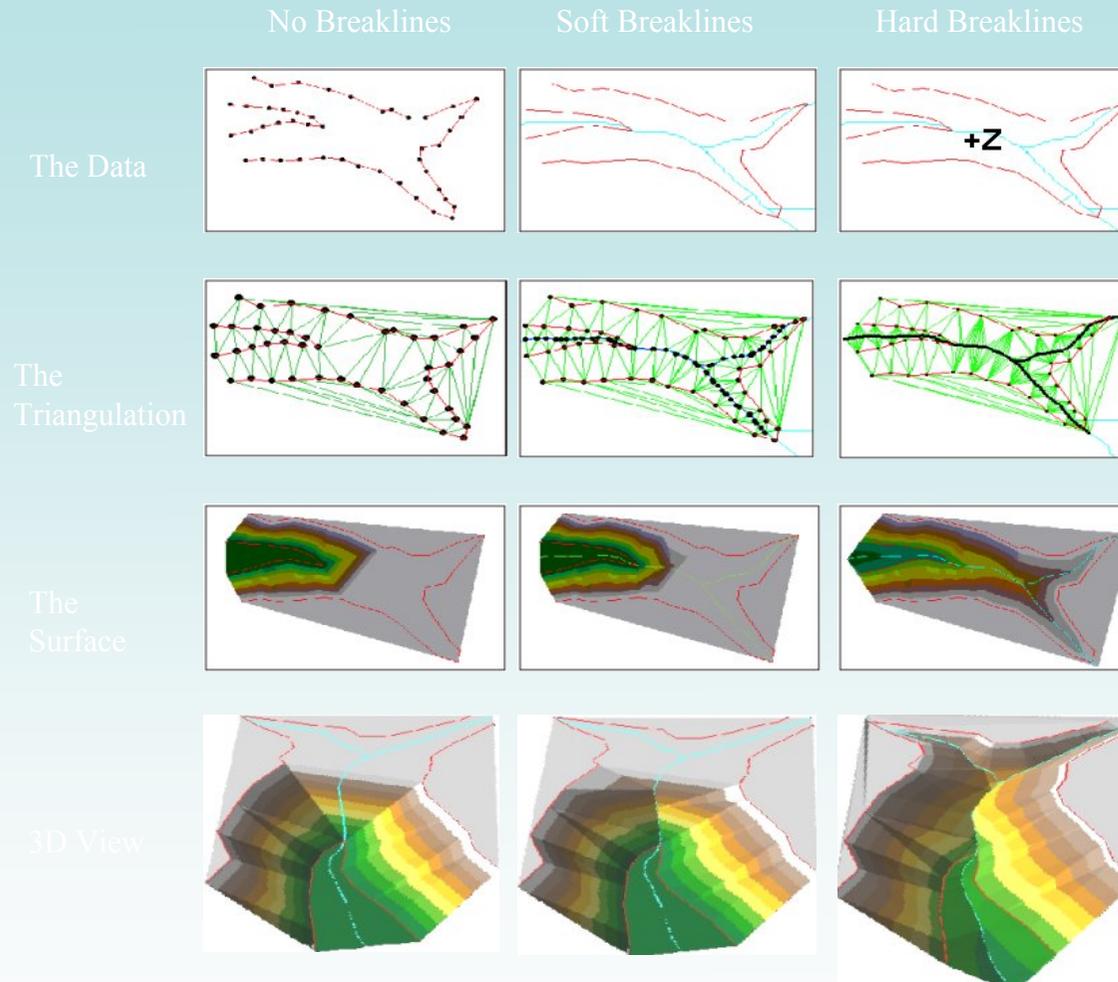
X=3970
Y=3869
Z=7746

X=4266
Y=4044
Z=7826

X=4562
Y=4219
Z=7906

Il risultato è unico e non dipende dal punto di inizio di tracciatura dei triangoli

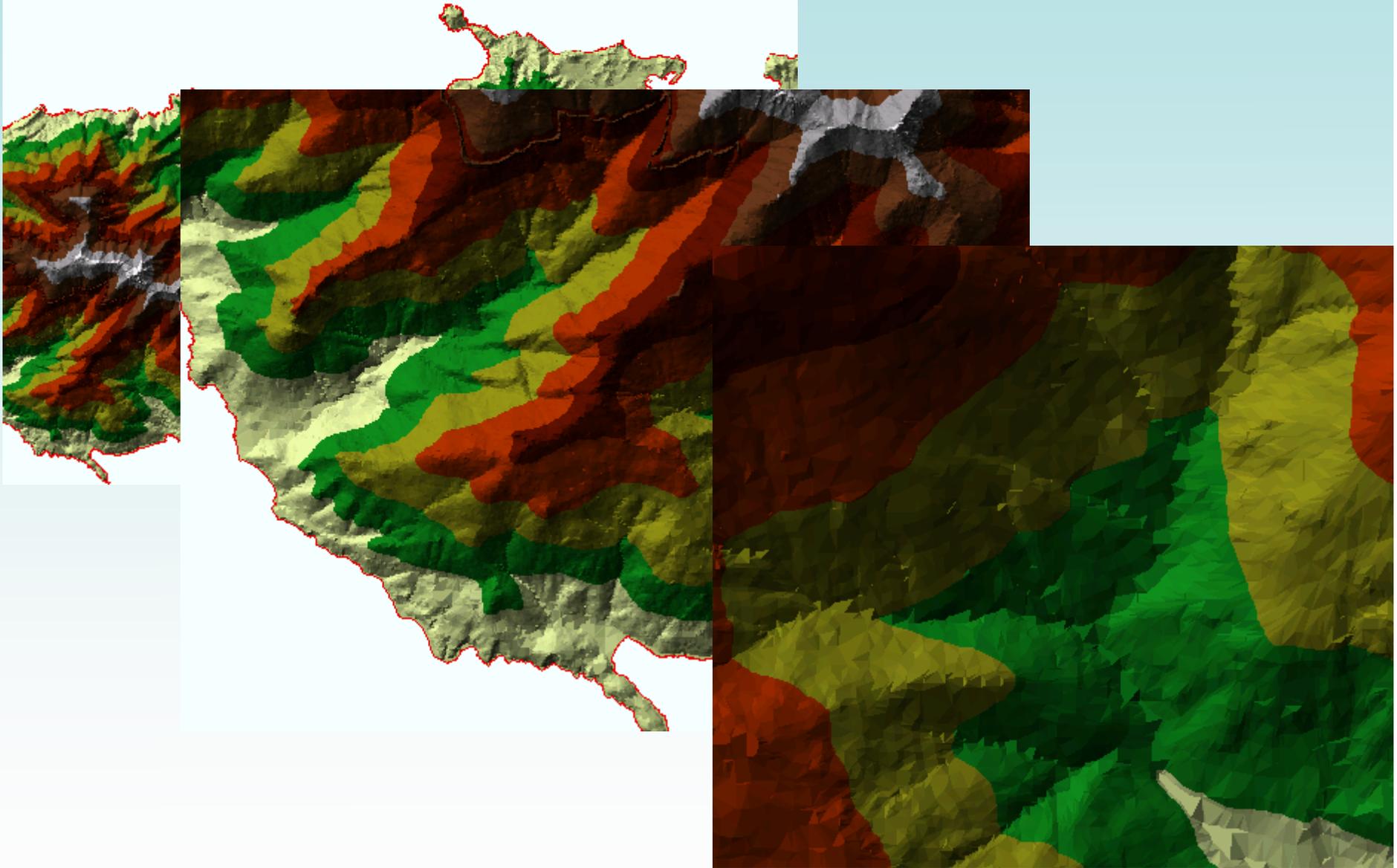
TIN (Triangulated Irregular Network): breaklines



Break lines

Linear features which define and control surface behavior in terms of smoothness and continuity.

TIN (Triangulated Irregular Network)



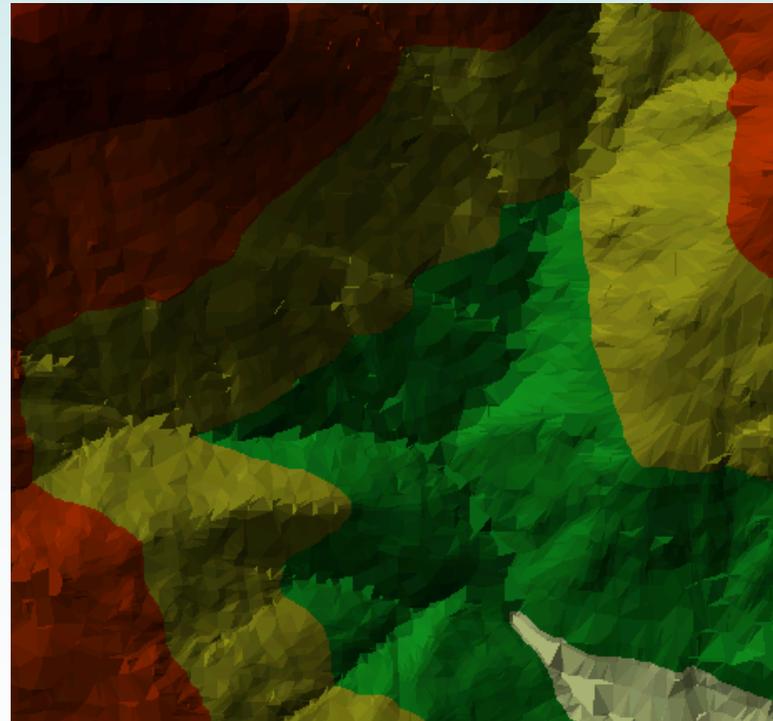
TIN (Triangulated Irregular Network)

- Molto utilizzato nella **generazione dei DEM** in contrapposizione alle metodologie raster
- **Esatto**: i valori dei dati di input sono contenuti nel modello finale come vertici di triangoli.
- La **dimensione dei triangoli** è dipendente dalla configurazione dei dati di input (es. in zona pianeggiante, numero inferiore e maggiori dimensioni dei triangoli)

Nessuna misura dell'errore di predizione

Presenza di brusche variazioni al passaggio da un triangolo all'altro

Possibile (spesso inevitabile) presenza di numerosi artefatti



Natural neighbour

E' molto simile a IDW

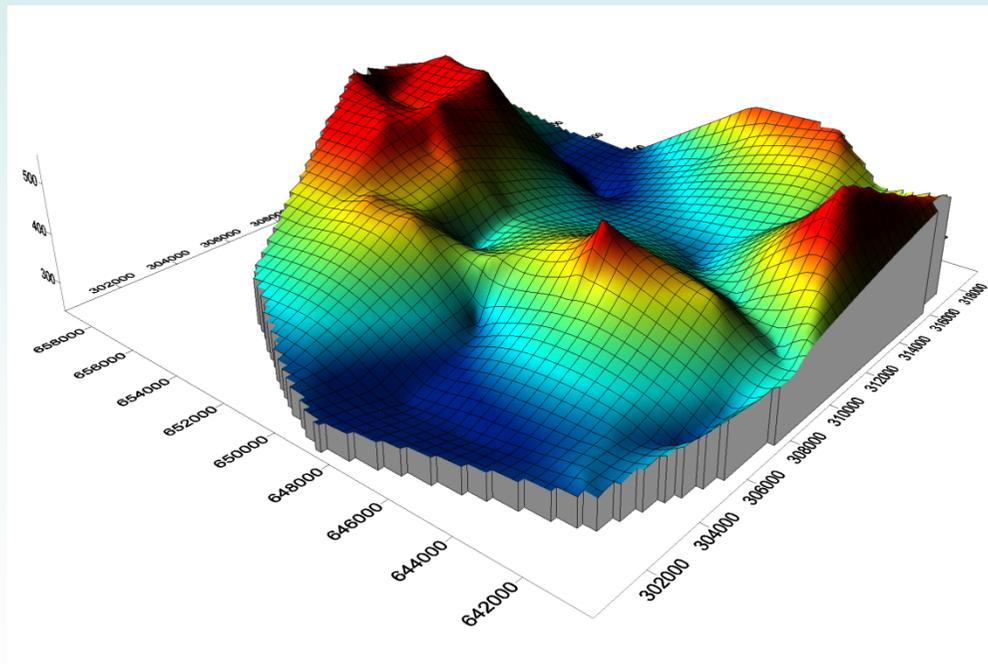
Semplice da utilizzare: non richiede la definizione del raggio di ricerca o il numero dei vicini

Uniformità all'interno delle zone

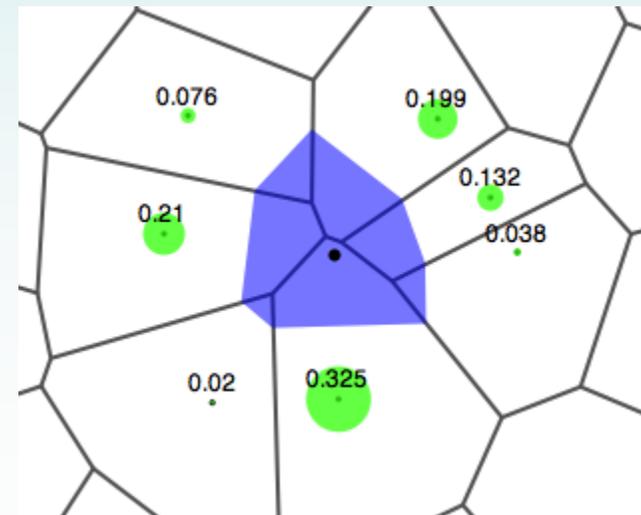
Affidabile per dataset molto ampi

Si assume che ogni punto misurato abbia una naturale "sfera di influenza" (basata sui poligoni di Voronoi)

E' di fatto una **media pesata in cui i pesi sono definiti dalla porzione di area utilizzata di ogni poligono di Voronoi.**



$$G(x, y) = \sum_{i=1}^n w_i f(x_i, y_i)$$

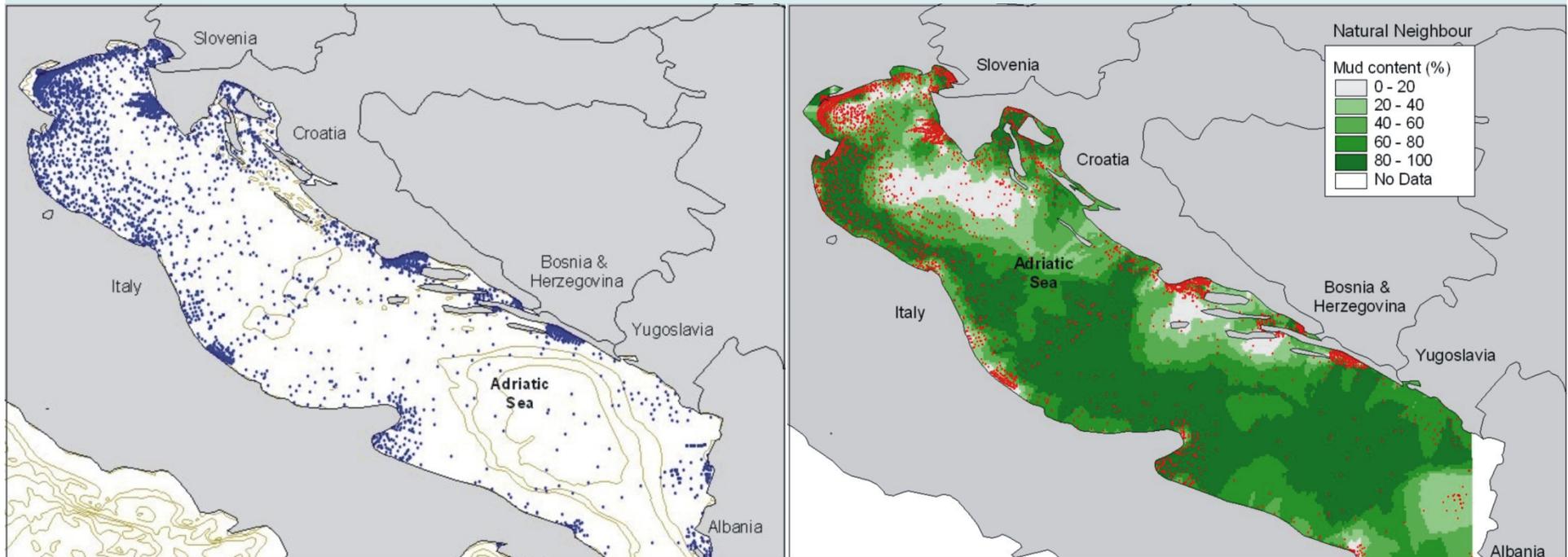


Natural neighbour

Lavora bene quando i campioni sono distribuiti in maniera irregolare

Produce numerosi artefatti (“punte”) se la densità dei dati di input è elevata.

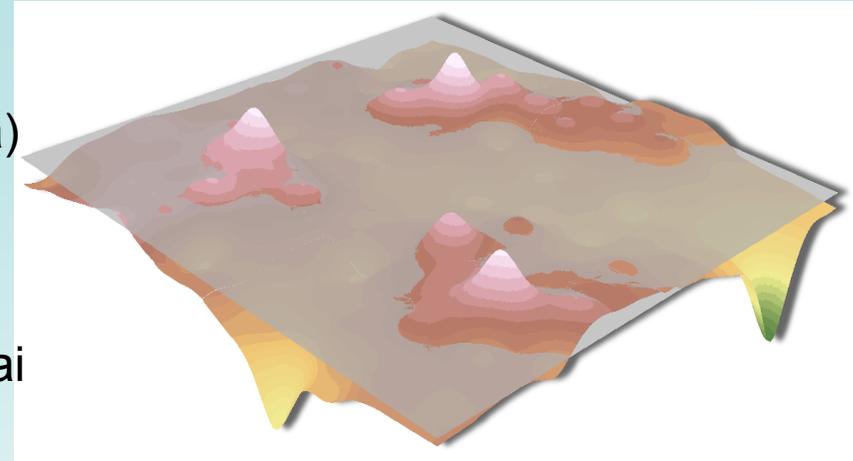
Nessuna stima dell'errore



Trend

Metodo globale, non esatto.
Generalizza una superficie (molto smussata)
ed **evidenzia una tendenza nei dati**.

Utilizza una regressione polinomiale per
adattare una superficie dei minimi quadrati ai
dati misurati. L'ordine della polinomiale è
definito dall'utente. Un ordine più elevato
genera superfici più complesse (ma non
necessariamente più accurate).



Gray plane is the trend surface

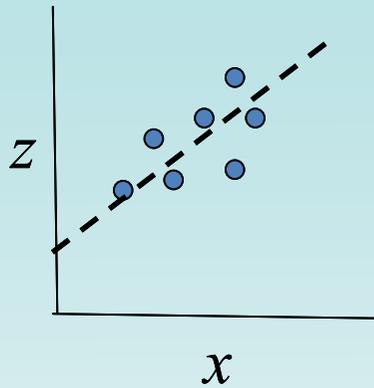
- Order 1: No curve (flat tilted surface)**
2: One curve
3: Two curves
4: Three curves, and so on



First order trend surface

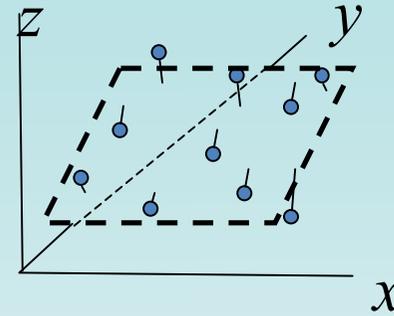
Superfici Trend di Primo Ordine

$$z = b_0 + b_1x + e$$

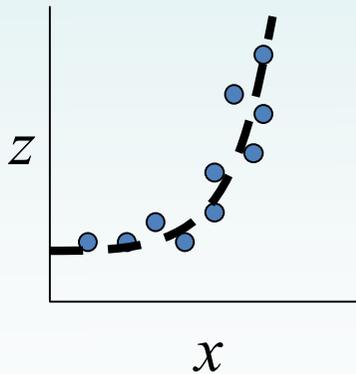


In two dimensions: z varies as a linear function of x and y

$$z = b_0 + b_1x + b_2y + e$$



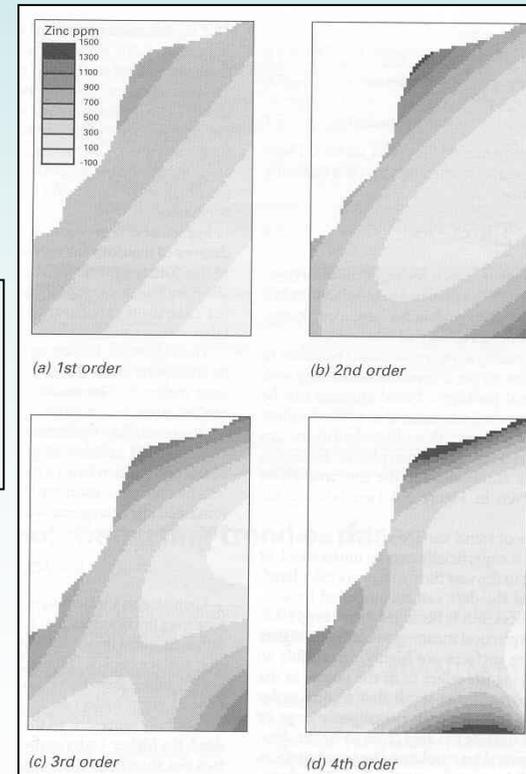
Superficie Trend di Secondo Ordine



In one dimension: z varies as a non-linear function of x

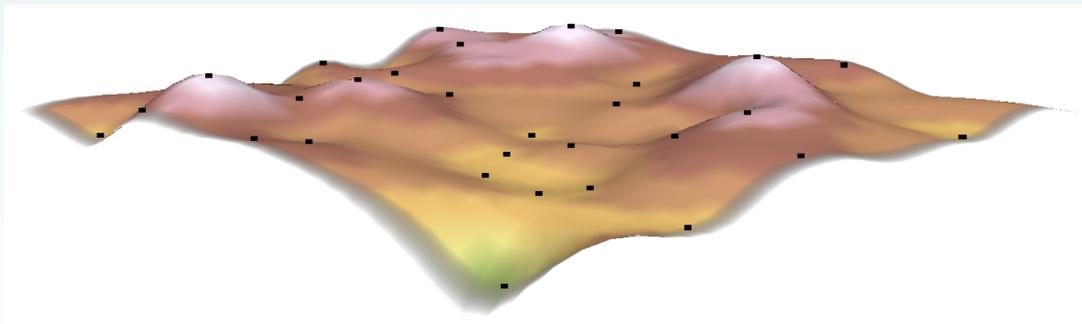
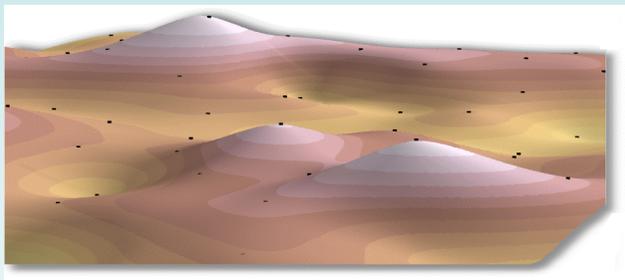
$$z = b_0 + b_1x + b_2x^2 + e$$

**Superfici Trend di
ordini superiori in
due dimensioni**



Spline (Radial Basis Function)

- Non lineare; esatto (la superficie passa dai punti di input)
- Spline esegue una stima dei valori attraverso una funzione matematica (applicata ai dati misurati nel vicinato) che **minimizza la curvatura**, dando origine ad una **superficie molto smussata** (simula uno strato di gomma che si conforma ai dati di input).
- Adatto per **superfici con variabilità debole** (es: temperature, superfici piezometriche, concentrazioni di inquinanti, ecc.)



Tipi di spline

In realtà è una famiglia di interpolatori diversi:

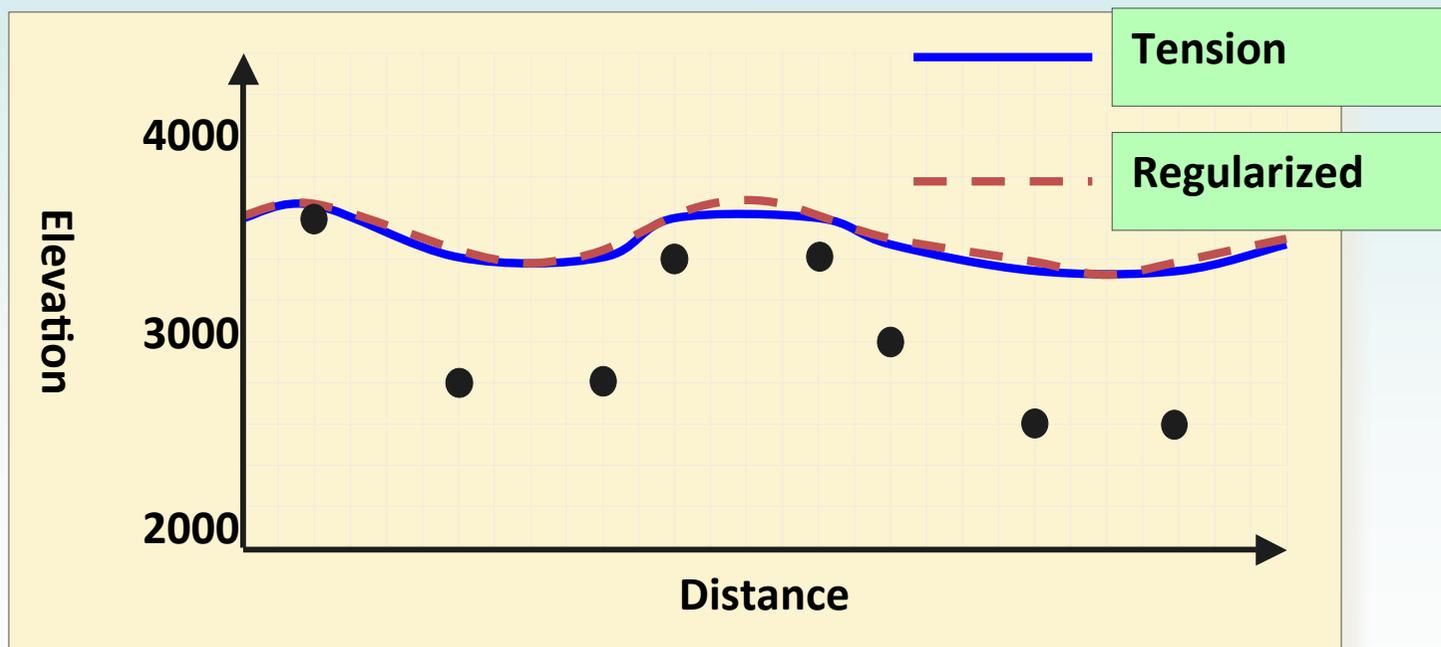
thin-plate spline, spline with tension, completely regularized spline, multiquadric function, and inverse multiquadric spline, ecc.

•Regularized

- Dossi più elevati e depressioni più profonde.
- Superficie più smussata

Tension

“Forza” l’adattamento della superficie
Dossi e depressioni meno accentuate
Superfici più grossolane



Spline (Radial Basis Function)

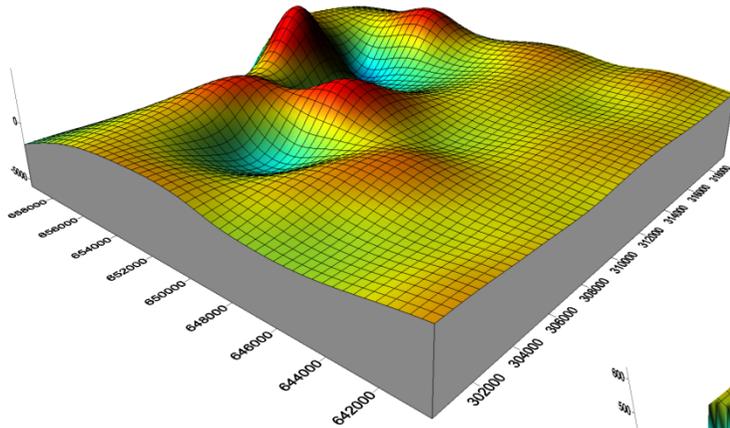
Pregi

- Visivamente **attraente e realistico**: fornisce risultati più morbidi e realistici di IDW e altri metodi deterministici.
- Produce una superficie di minima curvatura molto smussata
- Interpolatore **esatto**: la superficie passa esattamente per i punti di input
- Ampiamente utilizzato in Scienze della Terra
- Concettualmente si avvicina al kriging

Difetti

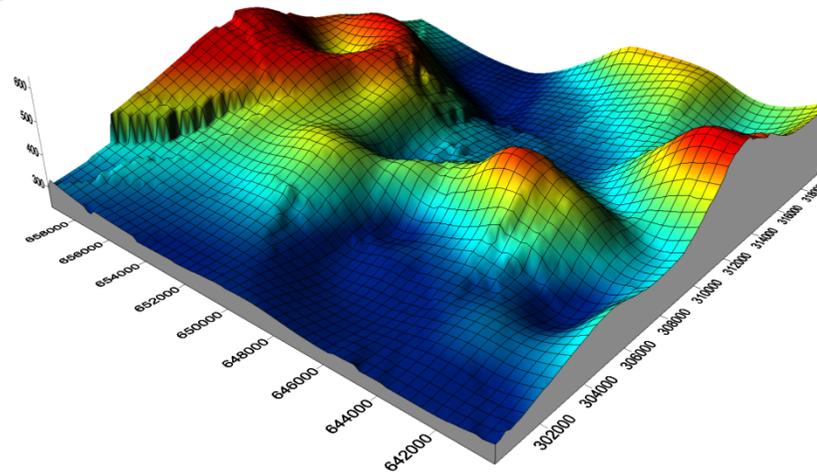
- I parametri di smussamento devono essere settati dall'utente
- Nessuna stima dell'errore
- Non lavora bene con variabili caratterizzate da bruschi cambiamenti di valore nello spazio.
- Eccessivi artefatti in aree con assenza di dati di input
- Se il numero di dati di input supera un certo limite, cresce l'influenza dei punti lontani con eccessivo smussamento della superficie di output

Thin Plate Spline



Using all 62 data points

Using 12 point neighbourhoods

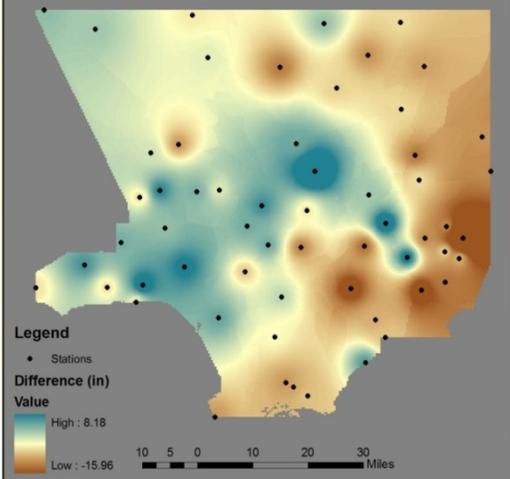




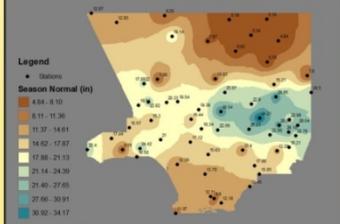
IDW Interpolation of Rainfall for Los Angeles County

By Vivian Lee
Geography 168
March 02, 2011

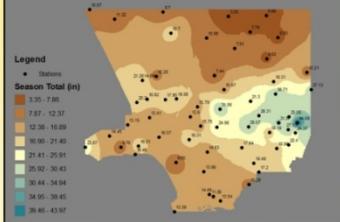
Rainfall Difference of Normal Season and Total Current Season



Rainfall for Normal Season



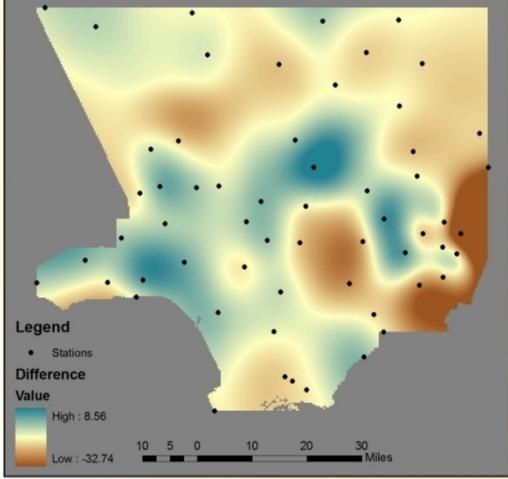
Total Rainfall for Current Season



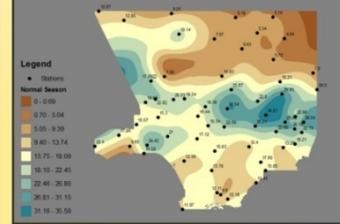
Spline Interpolation of Rainfall for Los Angeles County

By Vivian Lee
Geography 168
March 02, 2011

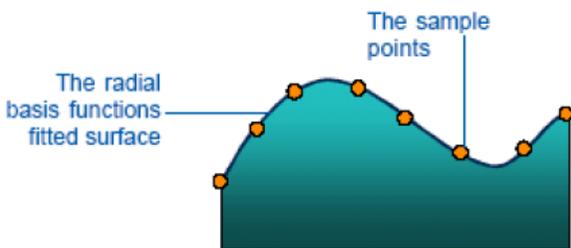
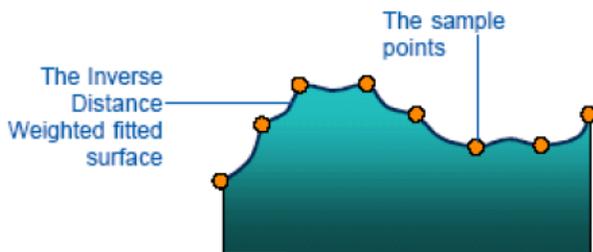
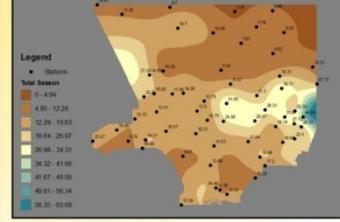
Rainfall Difference of Normal Season and Total Current Season



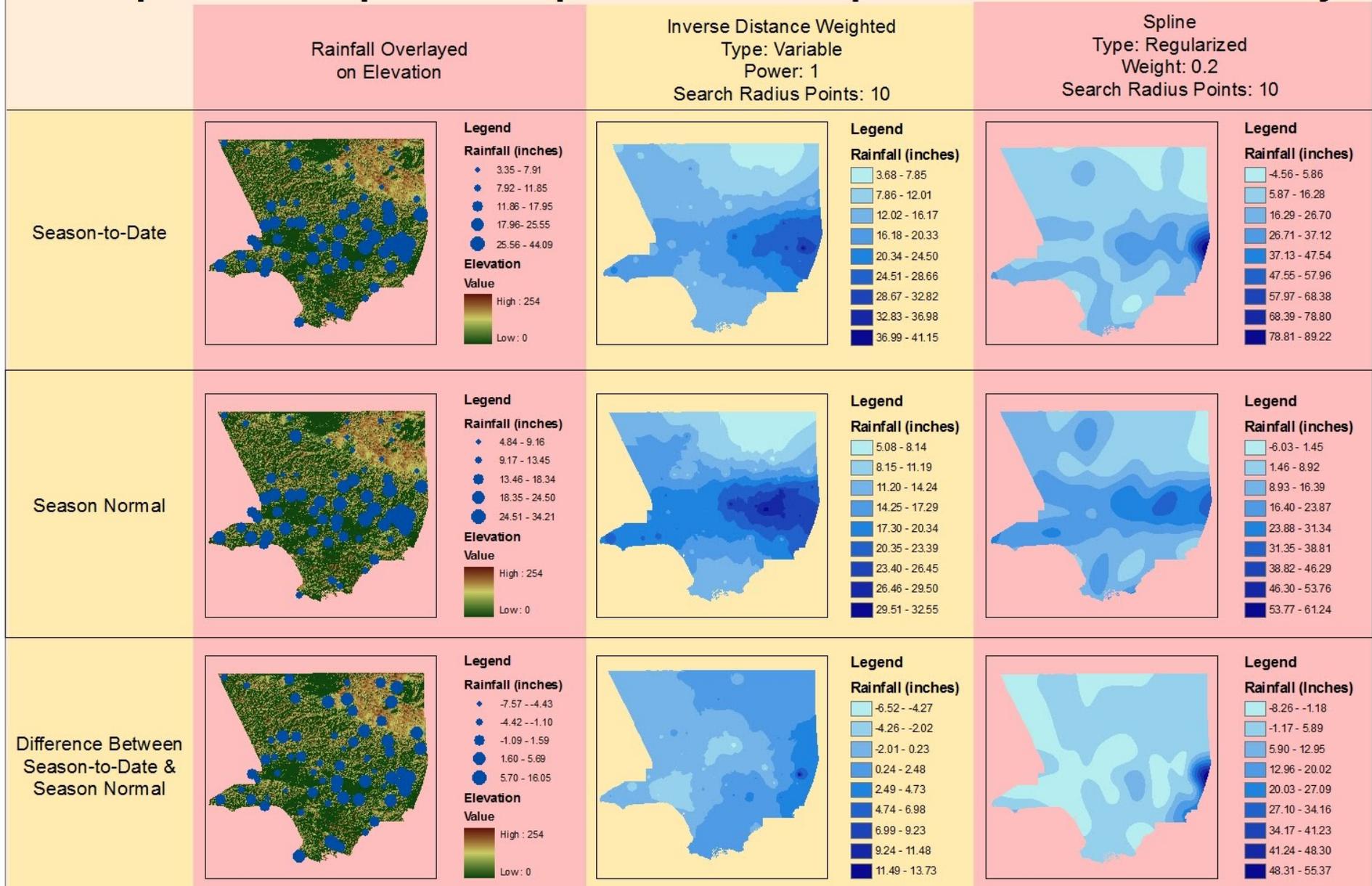
Rainfall for Normal Season



Total Rainfall for Current Season



Comparison of Spatial Interpolation Techniques: Rainfall in LA County



Michael Paul
 March 2, 2011
 Geographic Coordinate System: GCS_WGS_1984
 Datum: D_WGS_1984

20 10 0 20 40 60 Miles
 Note: All maps use the same scale.



Altri metodi deterministici

Global Polynomial

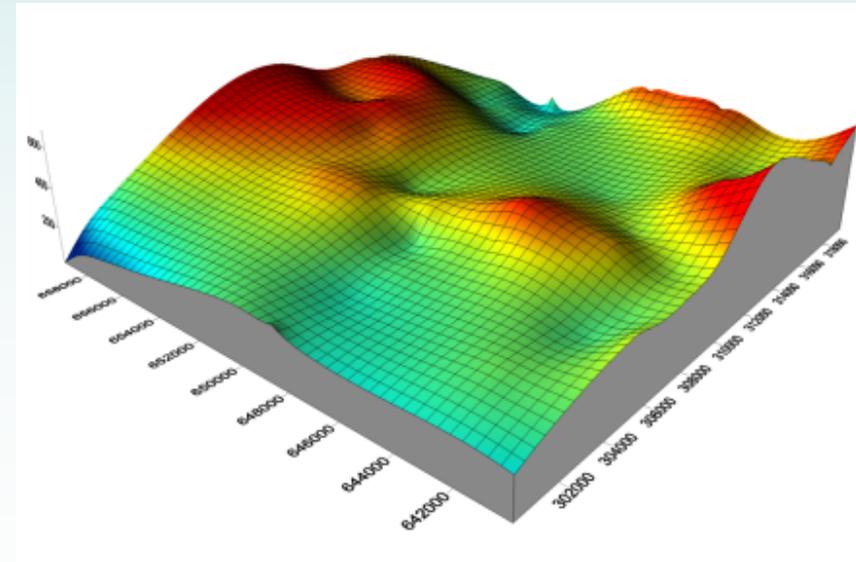
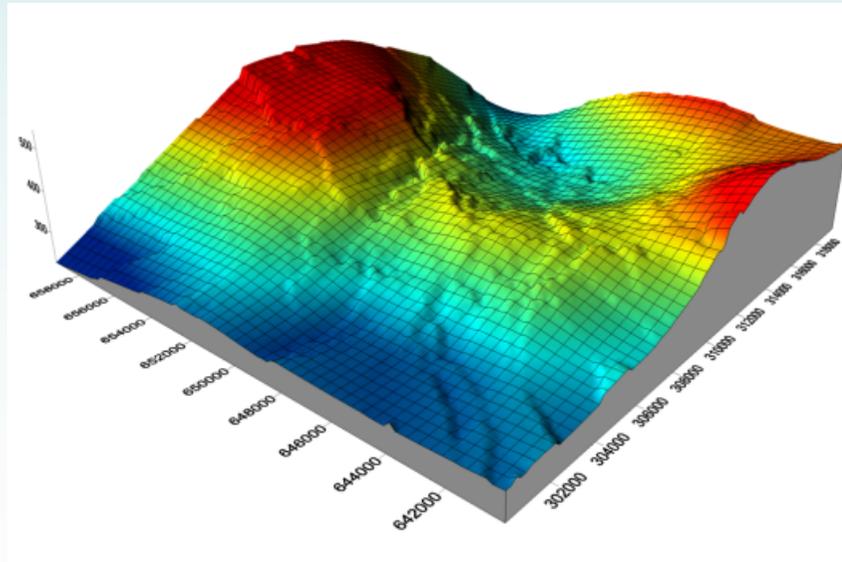
Local Polynomial

Minimum curvature

Triangulation with non-linear interpolation

Moving Average

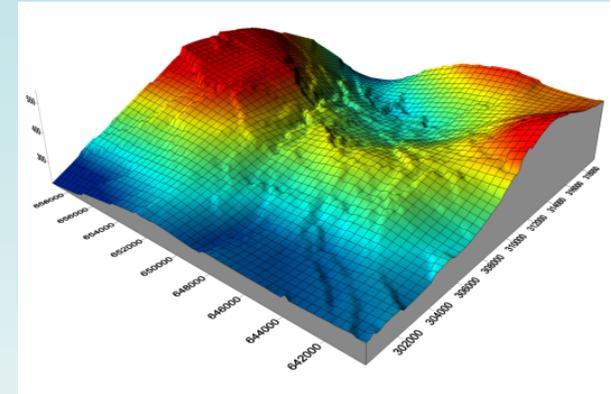
Topogrid (Digital Elevation Models) →



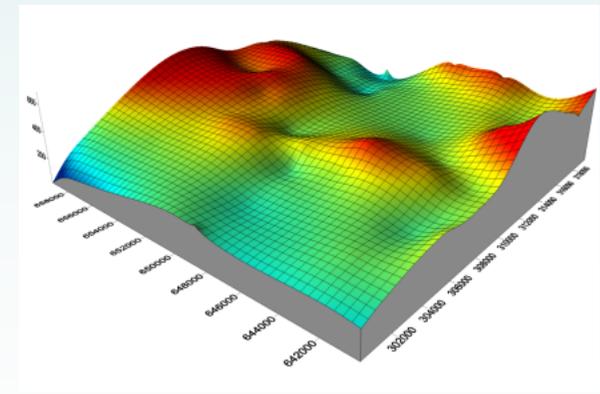
Altri metodi deterministici

Global Polynomial (GP) is a quick deterministic interpolator that is smooth (inexact). There are very few decisions to make regarding model parameters. It is best used for surfaces that change slowly and gradually. However, there is no assessment of prediction errors and it may be too smooth. Locations at the edge of the data can have a large effect on the surface. There are no assumptions required of the data.

- **Local Polynomial (LP)** fits a local polynomial (e.g. quadratic) to each grid point, using window (e.g. circle) of given size or number of points. It is a moderately quick deterministic interpolator that is smooth (inexact). It is more flexible than the global polynomial method, but there are more parameter decisions. There is no assessment of prediction errors. The method provides prediction surfaces that are comparable to Kriging with measurement errors. Local polynomial methods do not allow you to investigate the autocorrelation of the data, making it less flexible and more automatic than Kriging. There are no assumptions required of the data.



- **Minimum curvature**, seeks smoothed 'elastic membrane' fit to surface; fit to residuals after linear regression of original dataset (linear component then added back)
- **Topogrid** →



Diagnostica

Cross Validation:

Valutazione dell'accuratezza di una superficie statistica:
processo iterativo

1. Rimuovere un campione misurato
2. Interpolare nuovamente
3. Confrontare il campione rimosso con la superficie ottenuta
4. L'interpolatore ha predetto il campione mancante?
Errore = valore predetto – valore conosciuto.
5. Inserire nuovamente il campione e ripetere l'operazione con un altro.

Validation Subset:

Si rimuove una parte dei dati (test dataset) e si usa il resto (training dataset) per ricalcolare il modello.

Validation Points:

Si utilizzano punti di validazione indipendenti dal dataset utilizzato per il processo di interpolazione.

Metodi deterministici

In linea generale i metodi deterministici non tengono conto della variabilità nel dominio spaziale del parametro da interpolare e la loro implementazione risulta arbitraria.

Inoltre i risultati sono estremamente dipendenti dalla configurazione spaziale dei dati di input e dagli schemi di campionamento.

L'ambiente è il risultato di azioni e interazioni di diversi processi e fattori; che interagiscono spesso in maniera casuale.

L'approccio deterministico in molti casi non è adeguato per interpretare queste interazioni.

A Comparison of the Geostatistical Analyst methods

Method	Deterministic/ Stochastic	Output Surface Types	Computing Time/ Modeling Time ¹	Exact Interpolator	Advantages	Disadvantages	Assumptions ²
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 19th Advanced Summer School in Regional Science
- UB Geology GLY560: GIS
- 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.