

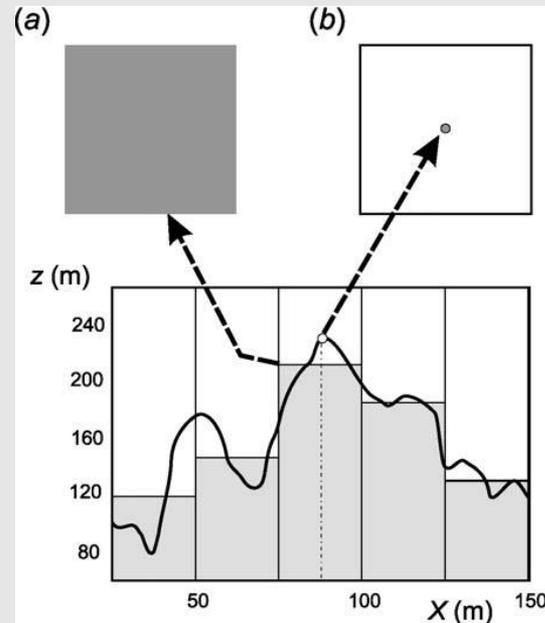
For example, if DEMs are constructed from contour lines, the surface roughness will be **under-represented**, so it is probably not very wise to extract geostatistical land-surface parameters.

Similarly, if a radar-based DEM of an area with dense vegetation has not been filtered for forest areas, it might not be a good idea to try to derive a map of ridges and streams, because the forest clear-cuts and roads may show up as artificial channels.

REMARK 1. The main objective of DEM-preprocessing is to remove artefacts, improve representation of shapes and hydrological/ecological processes.

Qual è la risoluzione ottimale (pixel size) per l'analisi geomorfometrica?

T. Hengl and I.S. Evans, 2009



The distinction between **pixel size** and **support size** in relation to the original distribution (the signal):

- (a) support size equals the resolution of an image (the averaged measurements);
- (b) support size is infinitely small (the point measurement at the center of a pixel).

The key problem when selecting a cell size for geomorphometry is that there can be significant differences between the interpolated surface elevation and the real land surface, meaning that some peaks and channels might disappear (or be displaced) when represented in a coarse raster DEM.

In general, a finer DEM resolution will also mean more accurate land-surface parameters and a higher information content (Kuo et al., 1999).

The amount of increase, however, depends on the general variability of the landscape.

A smooth, generally regular, landscape does not need a fine resolution DEM. If the DEMs are coarser than the scale of landscape processes, any results or indices derived from DEMs must be treated with caution (Pain, 2005).

If the cell size is too fine in relation to the vertical accuracy, it might introduce local artefacts and slow down the computation of land-surface parameters.

Il giusto compromesso....

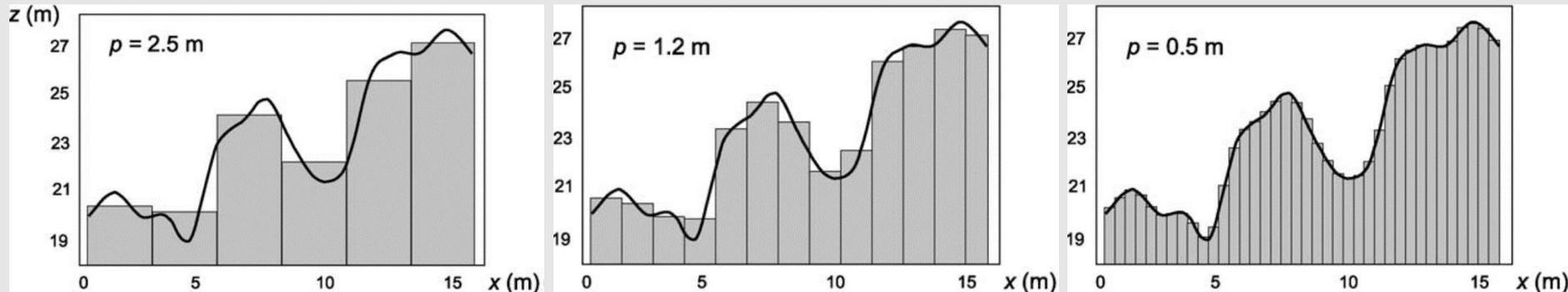
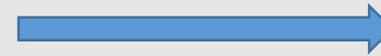
We need a **grid resolution that optimally reflects the complexity of a terrain**; one that can represent the majority of the geomorphic features (Borkowski and Meier, 1994; Kienzle, 2004; Pain, 2005; M.P. Smith et al., 2006).

Se immaginiamo di lavorare in una dimensione, lungo un profilo topografico, il profilo topografico è il segnale e la sua frequenza è determinata dalla densità dei punti di inflessione. In quest'ottica, il **teorema di Nyquist-Shannon** enuncia che **la dimensione della cella deve essere almeno pari alla metà della spaziatura media tra i punti di inflessione**.

$$\Delta s \leq \frac{l}{2 \cdot n(\delta z)}$$

Dove Δs è la dimensione della cella
 l è la lunghezza del tratto di profilo preso in considerazione
 $n(\delta z)$ è il numero dei punti di inflessione osservati

Risoluzione crescente della dimensione della cella



A schematic example showing the effect of grid resolution on the representation of topography: a cell size that is too coarse ($\Delta s = 2.5$ m) will misrepresent the topography; while a finer cell size ($\Delta s = 0.5$ m) will be more effective in representing all the peaks and channels present. From Hengl and Evans (2009).

Per capire meglio:

In the example, there are 12 inflection points and length of the segment is 16 m, suggesting a cell size of at least 0.6 m or finer.

That there is no universally suitable cell size, but only a cell size that is 'good enough' for a certain working scale.

In the 2D case, **if the DEM is based on digitised contours**, a suitable grid resolution can be estimated from the total length of the contours. Here, we do not actually have a map of inflection points, but we can approximate them using the contour map. Assuming that the contours were selected to present changes of surface function, a suitable cell size is:

$$\Delta s = \frac{A}{2 \cdot \sum l}$$

Dove A è la dimensione totale dell'area studiata e la sommatoria di l corrisponde alla lunghezza cumulativa delle curve di livello digitalizzate.

REMARK 5. A suitable cell size for a DEM can be selected to fit the complexity of terrain and/or the scale of work for the targeted application.

Gli errori nei DEM

Gli errori nei DEM e nei loro prodotti derivati possono essere raggruppati, secondo Wise (2000) in:

- errori grossolani,
- artefatti, opere umane;
- errori sistematici;
- errori casuali (*noise*).

Gli obiettivi principali delle operazioni di pre-processing dei DEM sono:

- Rimuovere gli errori grossolani e gli artefatti;
- Migliorare la rappresentazione del terreno;
- Migliorare la rappresentazione del processo oggetto di studio (idrologico, ecologico, ecc.).

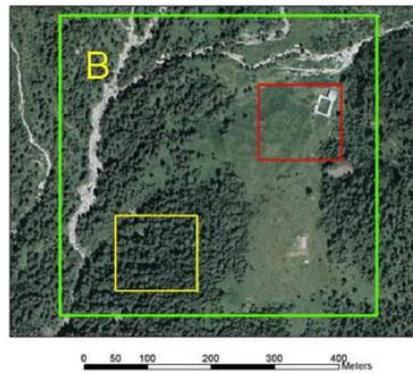
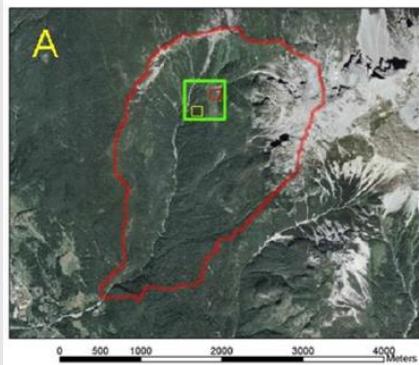
Sorgenti di errore nei LiDAR

Sorgenti di errore relative all'acquisizione dati:

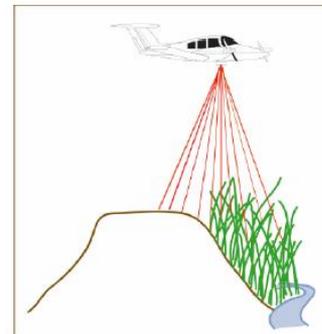
Laser scanner; DGPS (scarsa copertura satelliti, etc.); Calibrazione e sincronizzazione di laser scanner, GPS; Pianificazione del volo e procedure di calibrazione.

Sorgenti di errore relative alle procedure di post-processing:

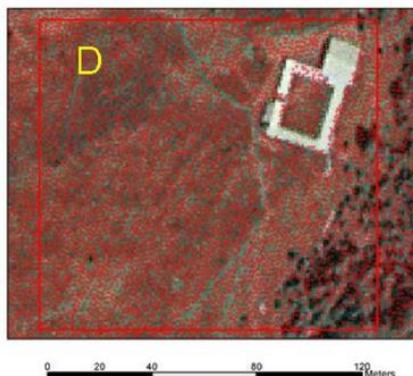
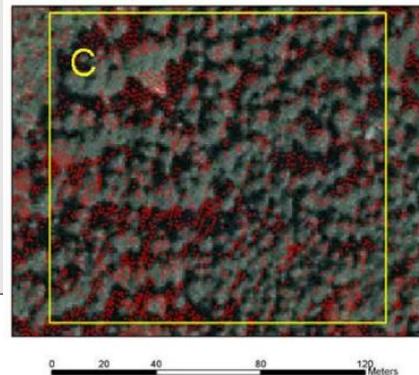
Trasformazione alle coordinate locali ed all'altezza ortometrica (datum, geoide, etc.); algoritmi di filtraggio utilizzati per rimuovere edifici e vegetazione e derivare le quote relative al suolo.



Densità spaziale dei dati "ground"



Erba alta e arbusti densi:
degrado accuratezza



LiDAR Quality Levels

A LiDAR point cloud often is described by its **Quality Level (QL), which is comprised of the nominal pulse spacing/density (NPS/NPD) and the relative and absolute vertical accuracy of the point cloud and derived digital elevation models.**

NPS and NPD refer to the point density of single swath, single instrument, first-return-only data.

The aggregate nominal pulse spacing/density (ANPS / ANPD) describes the net overall pulse spacing and density including overlapping swaths and multiple passes of the LiDAR system.

Table 1 and **Table 2** (from the USGS LiDAR Base Specification v1.3) describe LiDAR Quality Levels with regard to ANPS / ANPD, relative accuracy, and absolute accuracy.

Table 1
Aggregate Nominal Pulse Spacing and Density

Quality Level	Aggregate Nominal Pulse Spacing (m)	Aggregate Nominal Pulse Density (pls/m ²)
QL0	≤ 0.35	≥ 8.0
QL1	≤ 0.35	≥ 8.0
QL2	≤ 0.71	≥ 2.0
QL3	≤ 0.35	≥ 0.5

QL quality level pls/m² pulses per square meter m meter ≤ less than or equal to ≥ greater than or equal to

Table 2
Relative Vertical Accuracy for LiDAR Swath Data

Quality Level	Smooth Surface Repeatability, RMSD, (m)	Swath Overlap Difference, RMSD, (m)
QL0	≤ 0.03	≤ 0.04
QL1	≤ 0.03	≤ 0.08
QL2	≤ 0.08	≤ 0.08
QL3	≤ 0.12	≤ 0.16

QL quality level pls/m² pulses per square meter m meter ≤ less than or equal to ≥ greater than or equal to

Table 3 and **Table 4** describe land cover classes used for reporting absolute accuracy of the LiDAR point cloud.

Table 3
Land Cover Classes

Class	Land Cover Class or Description	Previous Reporting Group	Current Reporting Group
1	Clear or open, bare earth, low grass; for example: sand, rock, dirt, plowed fields, lawns, golf courses	FVA	NVA
2	Urban areas; for example: tall, dense man-made structures	SVA	NVA
3	Tall grass, tall weeds and crops: for example, hay, corn, wheat fields	SVA	VVA
4	Brush lands and short trees; for example: chaparrals, mesquite	SVA	VVA
5	Forested areas, fully covered by trees; for example: hardwoods, conifers, mixed forests	SVA	VVA
6	Sawgrass	n/a	n/a
7	Mangrove and swamps	n/a	n/a

FVA Fundamental Vertical Accuracy **SVA** Supplemental Vertical accuracy **NVA** Nonvegetated Vertical Accuracy **VVA** vegetated vertical accuracy **n/a** not applicable or not available

Table 4
Absolute Vertical Accuracy for LiDAR Data and Digital Elevation Models

Quality Level	RMSE (Non Vegetated) (m)	NVA at the 95% Confidence Level (m)	VVA at the 95th Percentile (m)
QL0	≤ 0.050	≤ 0.098	≤ 0.15
QL1	≤ 0.100	≤ 0.196	≤ 0.30
QL2	≤ 0.100	≤ 0.196	≤ 0.30
QL3	≤ 0.200	≤ 0.392	≤ 0.60

QL quality level **RMSE** root mean square error in the z direction **NVA** non-vegetated vertical accuracy **VVA** vegetated vertical accuracy **m** meter ≤ less than or equal to

Table 5 Describes the minimum classification of the LiDAR point cloud for the USGS 3D Elevation Program. **Table 6** Describes the requirements for the minimum cell size for the digital elevation models that are derived from the bare-earth classified point cloud.

Table 5
Minimum LiDAR Data Classification Scheme

Code	Description
1	Processed, but unclassified
2	Bare earth
7	Low noise
9	Water
17	Bridge deck
18	High noise
20	Ignored Ground (typically breakline proximity)
21	Snow (if present and identifiable)
22	Temporal Exclusion (typically non-favored data in intertidal zones)

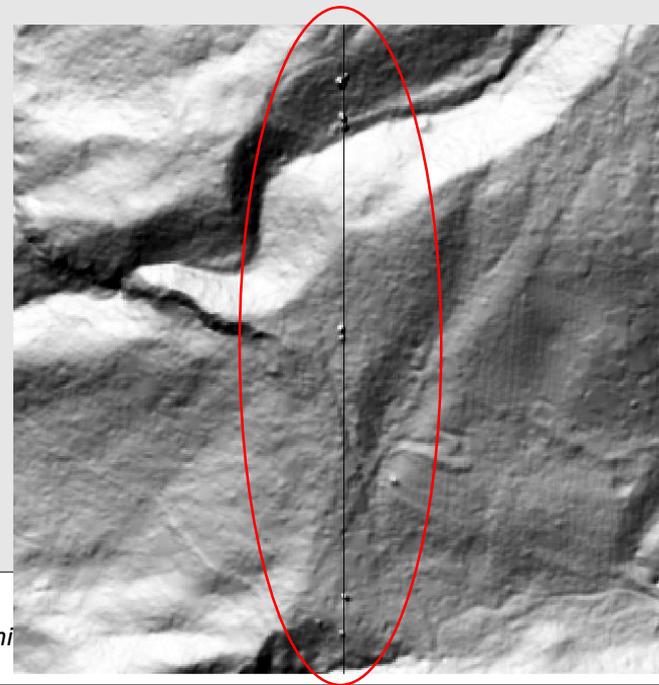
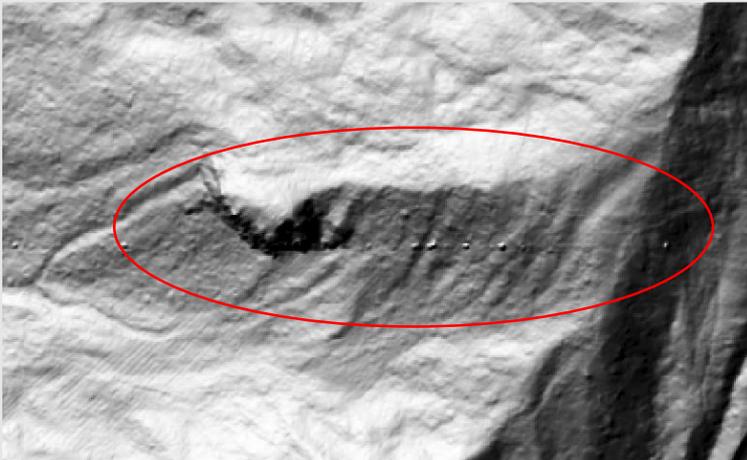
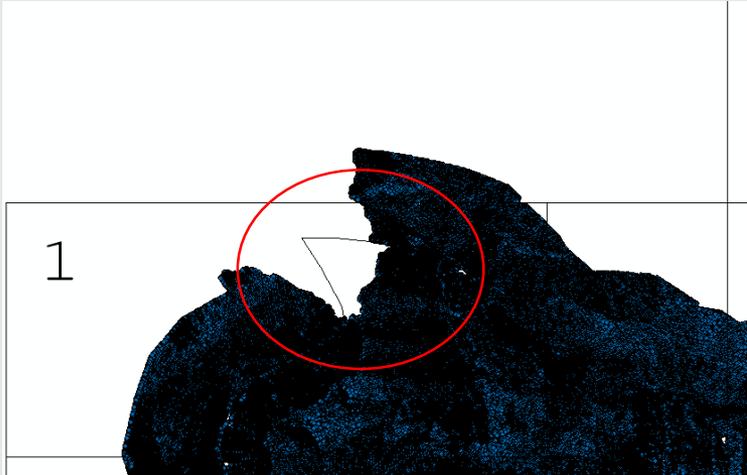
Table 6
Minimum Digital Elevation Model Cell Size

Quality Level	Minimum Cell Size (m)	Minimum Cell Size (ft)
QL0	0.5	1.0
QL1	0.5	1.0
QL2	1	2
QL3	2	5

QL quality level **m** meter **ft** feet

Prima di utilizzare un DEM, faccio un'analisi esplorativa di tipo QUALITATIVO

- mancata copertura del confine dell' area rilevata
- lunghe fasce di punti con quota erroneamente rialzata rispetto all'intorno con orientamento Nord-Sud e Est-Ovest (errori di sovrapposizione e/o ritaglio)
- mancato filtraggio di alberi isolati



Cavalli, 2017

Chiara Calligaris, Ph.D. – D.M.G. Uni

....ancora analisi qualitativa.....

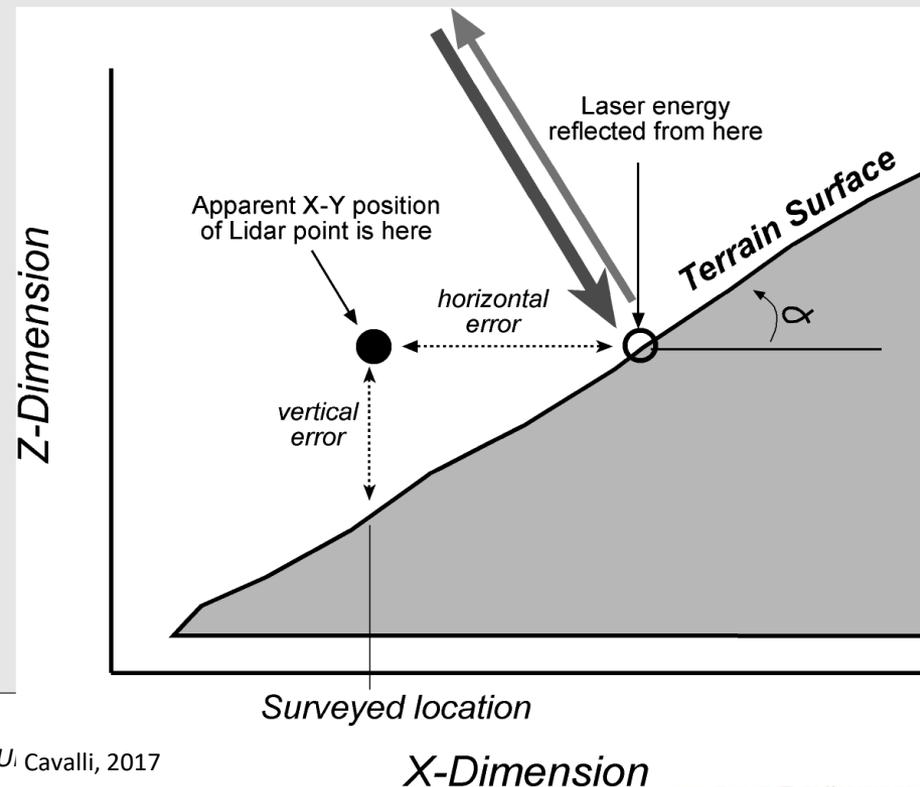
Per verificare la qualità del prodotto che sto utilizzando, devo avere dei *reference points* che possono essere messi in relazione con le coordinate del mio DEM. Devo pertanto verificare:

- L'accuratezza verticale (z)
- L'accuratezza orizzontale (x, y)

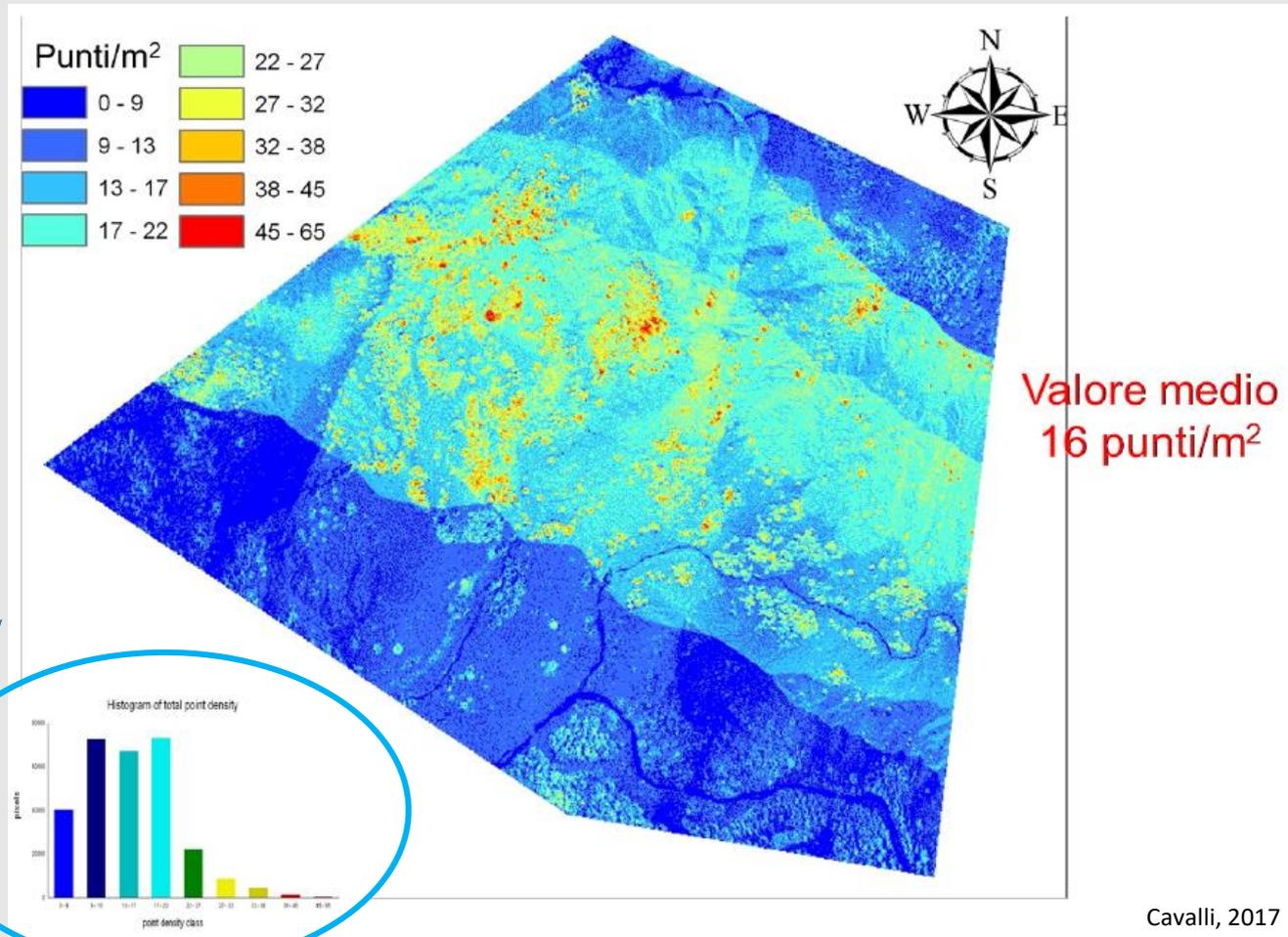
Valori da letteratura mi indicano che per avere un prodotto qualitativamente valido devono sussistere le condizioni: 15-20 cm (accuratezza verticale) 0.5-2 m (accuratezza orizzontale).

Errore verticale = $\tan \alpha$ x errore orizzontale

ad esempio, se ho una pendenza di 10° ed un errore orizzontale di 100 cm, l'errore apportato in **z** può raggiungere ± 18 cm



....ancora analisi qualitativa: DENSITA' SPAZIALE



E' importante effettuare un'analisi di densità sul dataset vettoriale prima di effettuare l'interpolazione del dato per farsi una prima idea della qualità dei dati di partenza e per avere un'indicazione per la scelta della risoluzione del DTM.

.... analisi QUANTITATIVA

L'accuratezza verticale (z)

Root Mean Square Error (RMSE)/Errore quadratico medio

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [z(s_i) - z_{REF}(s_i)]^2}{n}}$$

n number of spots, $z(s_i)$ is the elevation at the spot location and $z_{REF}(s_i)$ is the reference elevation at same location.

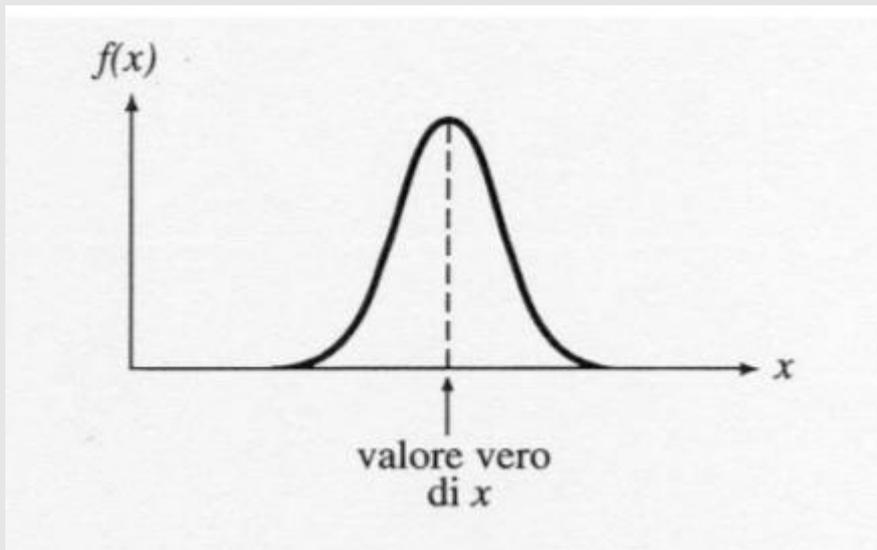
Ciò significa che l'ERRORE (RMSE) è dato dalla differenza tra i valori di quota del dataset ed i valori di quota di una sorgente indipendente campionata a più alta accuratezza per punti identici.

L'accuratezza orizzontale (x, y)

- Difficile da stimare
- Funzione dell'altezza di volo
- Errore x, y \approx altezza volo/1000

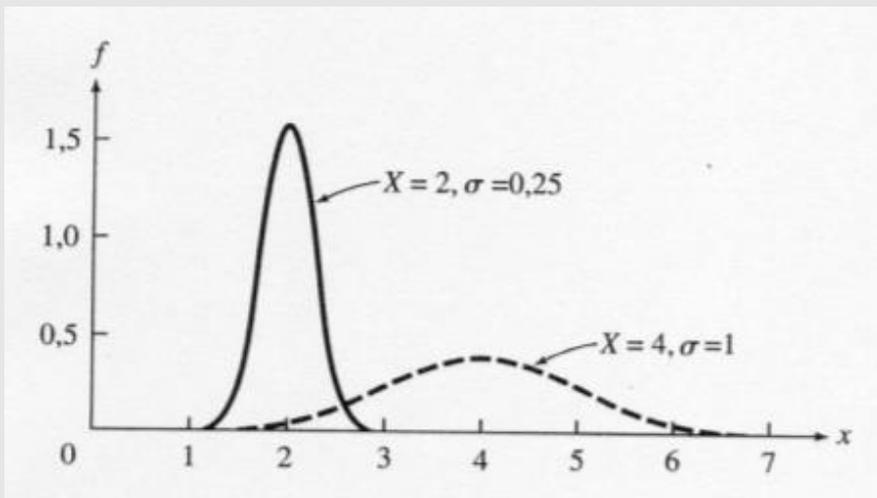
REMARK 3. *For geomorphometry, the main concern is the accuracy of the outputs (parameters and objects) and not the absolute accuracy of measured elevations.*

La distribuzione di frequenza dell'errore



La **funzione di Gauss** rappresenta una curva simmetrica e centrata sul valore sul valore $x = X$ dove la funzione assume il valore 1.

Se, invece, le misure hanno errori sistematici apprezzabili, allora non ci si deve aspettare che la distribuzione limite sia centrata sul valore "vero".



Il parametro σ dipende dalla precisione dell'apparato di misura.

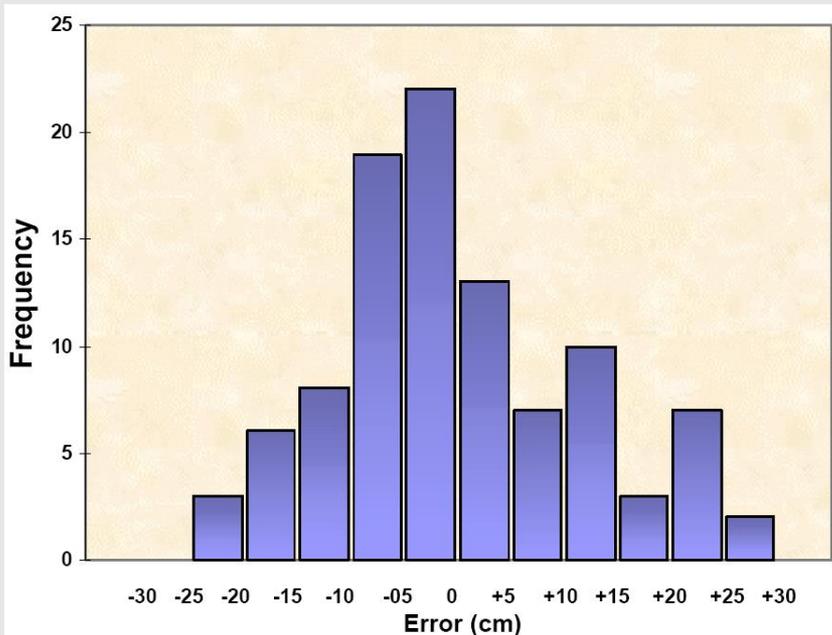
A piccoli valori di σ corrispondono curve strette, ossia distribuzioni fortemente addensate intorno al valore $x = X$, che a loro volta corrispondono ad una misura molto precisa, mentre per valori grandi di σ la distribuzione si allarga, il che corrisponde ad un apparato di misura e/o ad una procedura di misura poco precisi.

Errori casuali vs. errori sistematici

Gli **errori casuali** variano in modo imprevedibile da una misura all'altra e influenzano il risultato qualche volta per eccesso, qualche altra volta per difetto.

Gli **errori sistematici** avvengono sempre nello stesso senso: o sempre per eccesso, o sempre per difetto.

La distribuzione di frequenza dell'errore

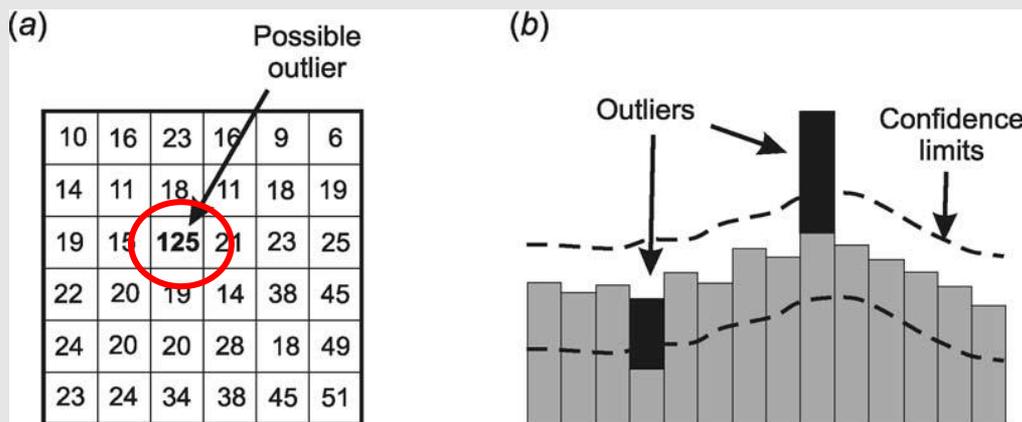


Creare un istogramma di tutti gli errori rappresenta la forma più comune di analisi dell'errore;

- L' istogramma evidenzia la forma della distribuzione (normale o no) e l'eventuale presenza di errori sistematici.

Importante calcolare anche:

- La tendenza centrale (media, mediana, etc.) -> se $\neq 0$ errore sistematico
- Variabilità generale (ad es. deviazione standard) -> outliers
- Asimmetria (generalmente se $> \pm 0.5$ la distribuzione non è normale)



Schematic example showing the expected local outliers and the result of filtering. The black color in (b) indicates change in the original values derived using the filter.

**Passiamo da una nuvola di punti ad una
superficie:**

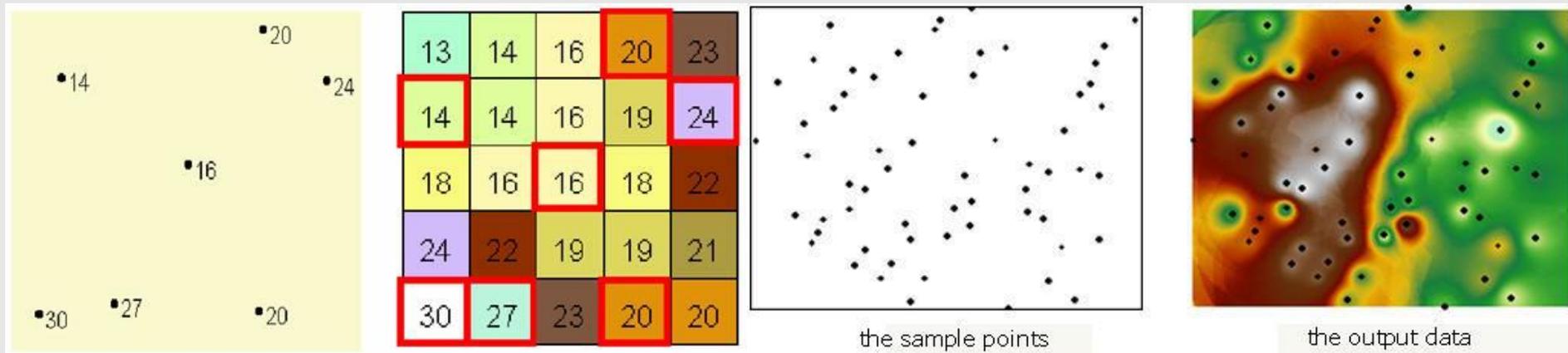
Gli interpolatori



L'INTERPOLAZIONE

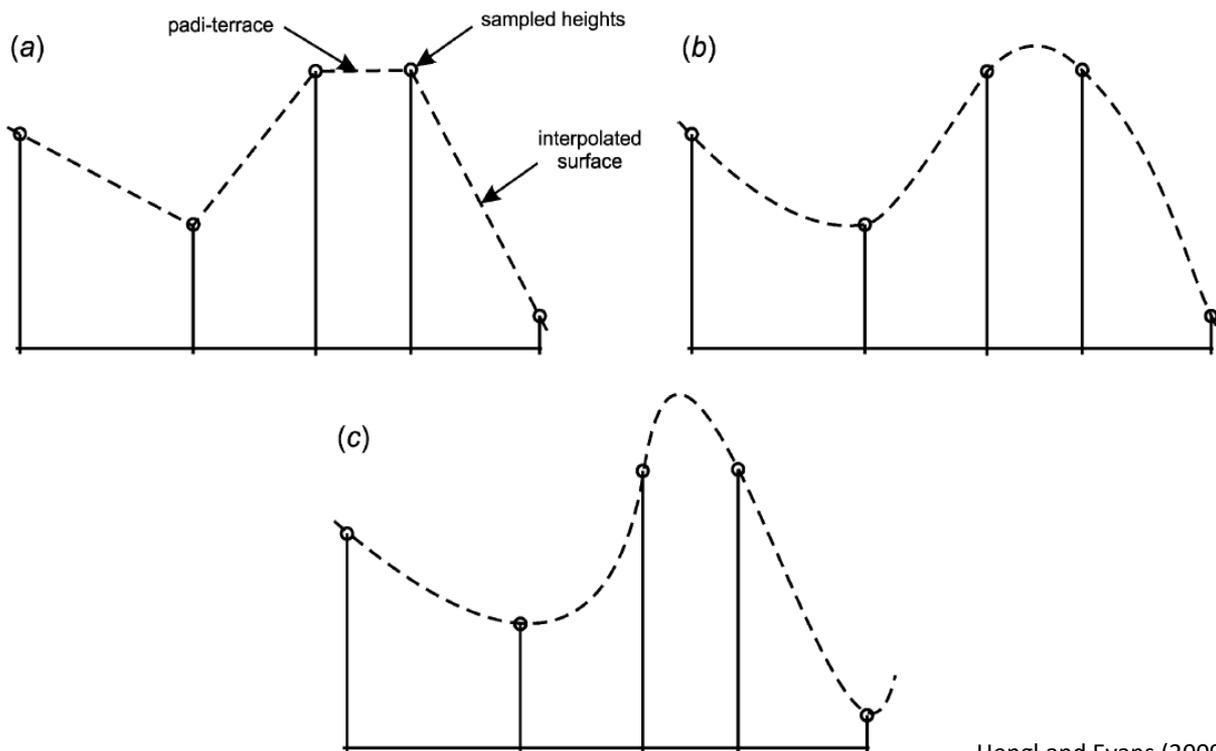
Qualora i punti quotati (punti LiDAR o curve di livello) siano uniformemente distribuiti nello spazio, il successo nella creazione di un DEM dipende fortemente dal **metodo di interpolazione prescelto**.

L'interpolazione (**gridding**) converte i punti vettoriali quotati in DEM continui di tipo raster, stimando i valori sconosciuti di ciascuna cella sulla base dei valori noti dei punti vicini.



ALGORITMI DI INTERPOLAZIONE

Gli algoritmi sono molto numerosi e sono stati implementati in diversi software GIS (Geographic Information Systems). Esistono anche software specifici che permettono un maggior controllo dei parametri: il software **Surfer** è uno di questi.



A schematic example, showing the effect of choice of interpolation algorithm on the quality of output:

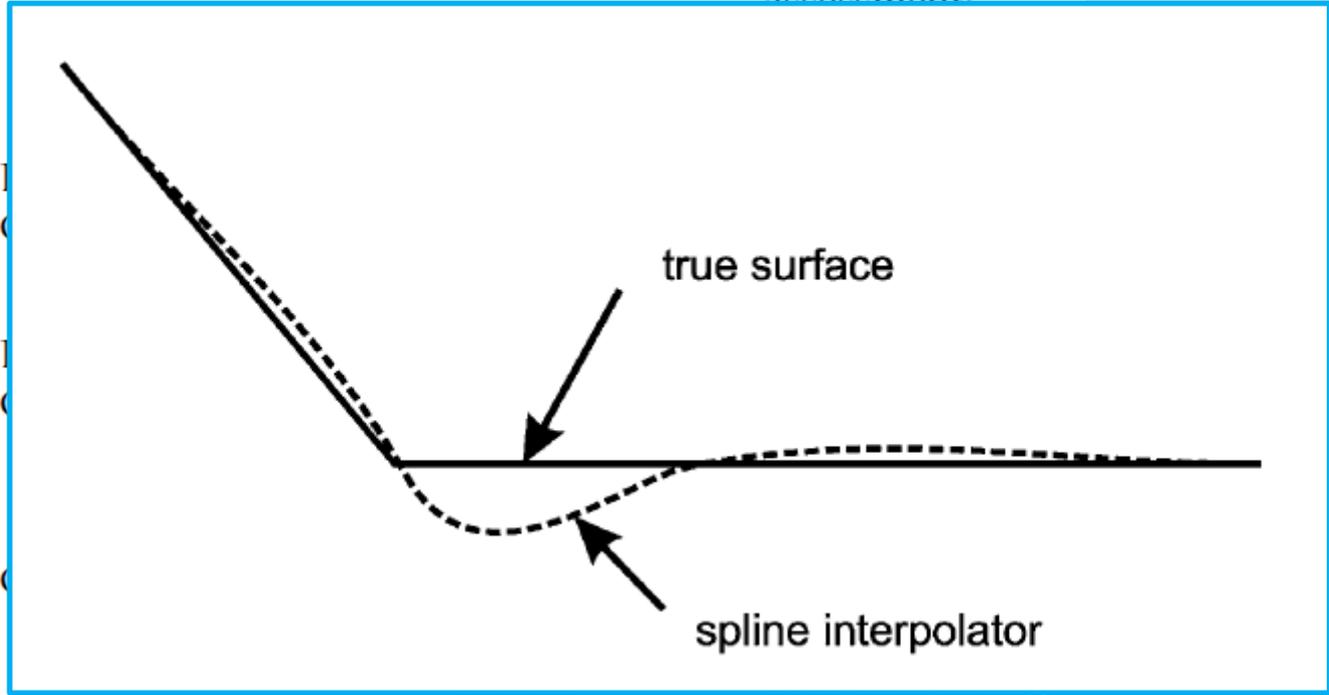
- (a) **pure linear interpolation** can cause obvious artefacts such as multiple terraces and flat summits;
- (b) **splines** are often very successful interpolators, because they can represent the shapes correctly;
- (c) **if unrealistic parameters are used with splines**, the final result can be even poorer than if a simple linear interpolator had been used.

Hengl and Evans (2009)

Method	Smoothing effect	Local/ Global	Deterministic/ Stochastic	Requirements/ Inputs	Possible problems
Linear interpolation	Low	Local	Deterministic	None	No error assessment; cut-offs and similar artefacts
Inverse distance interpolation	Low	Local/ Global	Deterministic	Weighting function, search radius	No error assessment; over-smoothing
Ordinary kriging	Medium	Local/ Global	Stochastic	Variogram model, search radius	Over-smoothing; statistical assumptions
Moving surface	High	Global	Deterministic/ Stochastic	Polynomial order, search radius	Possible over-fitting; over-smoothing
Splines	High	Local	Deterministic	Smoothness factor, search radius	Overshooting; over-smoothing
ANUDEM	High	Local/ Global	Deterministic	Smoothness factor, search radius, streams	Over-smoothing

Hengl and Evans (2009)

Method	Smoothing effect	Local/ Global	Deterministic/ Stochastic	Requirements/ Inputs	Possible problems
Linear interpolation	Low	Local	Deterministic	None	No error assessment
Inverse distance interpolation	Low	Local	Deterministic	None	No error assessment
Ordinary kriging	Medium	Local	Deterministic	None	No error assessment
Moving surface	High	Local	Deterministic	None	No error assessment
Splines	High	Local	Deterministic	Smoothness factor, search radius	Overshooting; over-smoothing
ANUDEM	High	Local/ Global	Deterministic	Smoothness factor, search radius, streams	Over-smoothing



L'interpolazione può essere effettuata globalmente o localmente

METODI GLOBALI

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

- trend (tendenza)
- serie di Fourier

METODI LOCALI

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

- spline
- IDW (inverse distance weight)
- kriging (collocazione)

I metodi globali non possono modellare bene gli andamenti locali e sono quindi usati per descrivere variazioni ad ampio raggio.

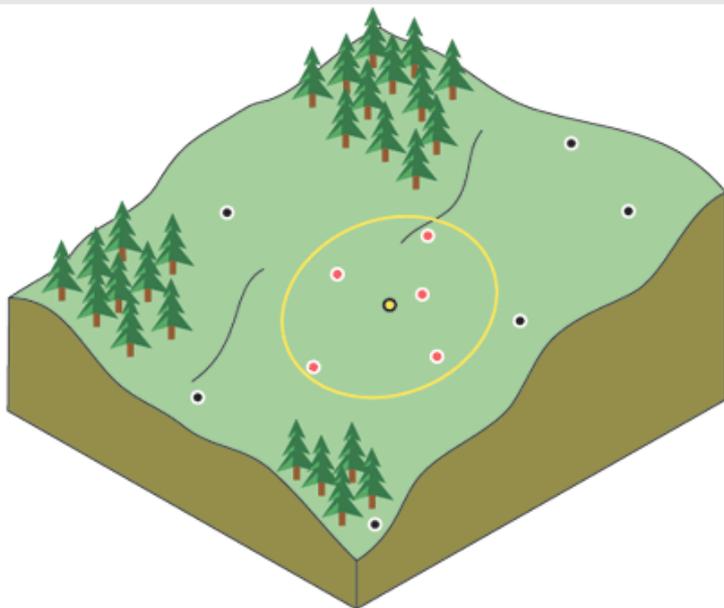
IDW ovvero INVERSE DISTANCE WEIGHTED

$$\hat{Z}(\text{yellow}) = \sum_{i=1}^5 \lambda_i (d_{\text{yellow}, \text{red}_i}) \times Z(\text{red}_i)$$

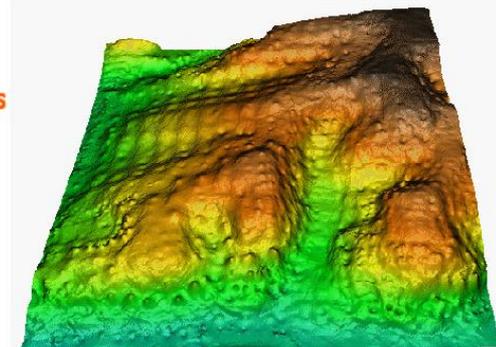
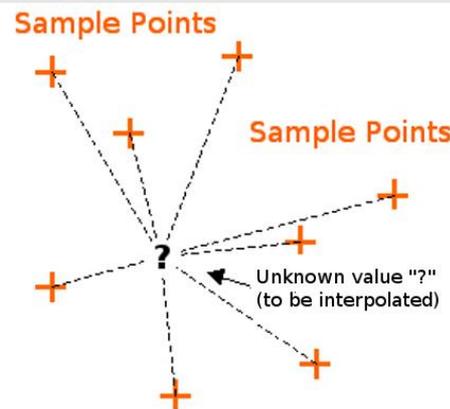
$$\lambda_i = \frac{d_i^{-p}}{\sum_{j=1}^5 d_j^{-p}}$$

DISTANZA INVERSA PONDERATA: è un metodo di interpolazione che, in funzione dei dati osservati, applica per l'identificazione del nuovo punto, una **media pesata in base alla distanza dal punto di osservazione (in giallo bordato nero)**.

In genere viene utilizzato un raggio di ricerca (cerchio giallo) che permette di limitare il numero di punti utilizzati.



Dove p è un numero reale positivo chiamato *power parameter* (tipicamente, $p=2$) e d è la distanza dal punto quotato in input al punto di interpolazione.



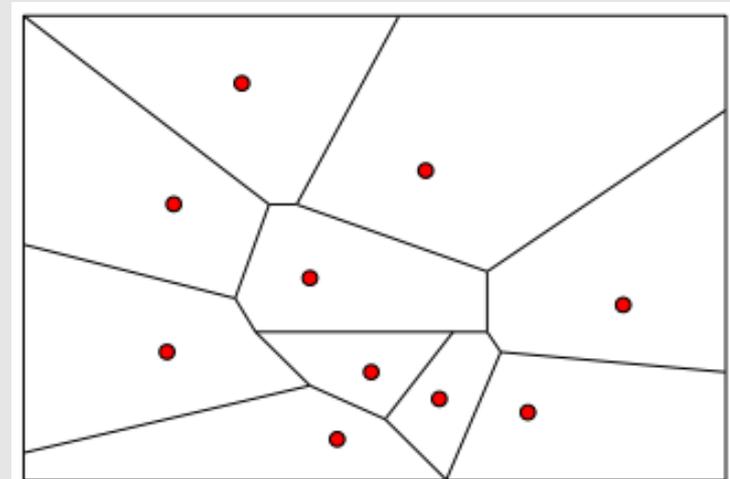
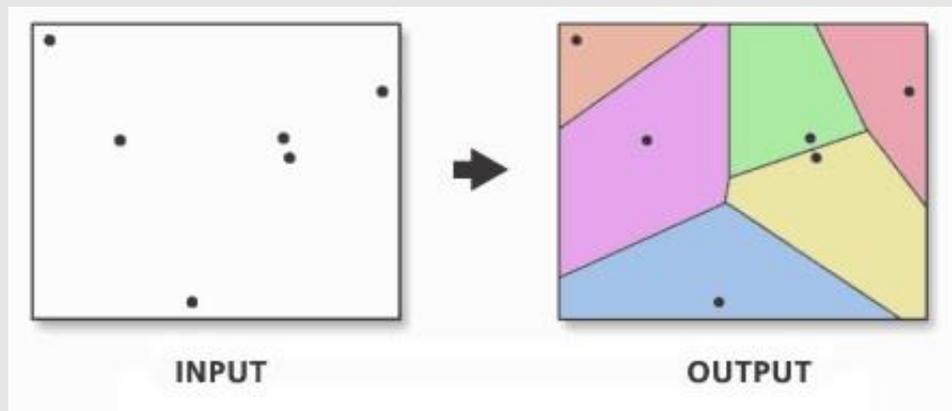
Poligoni di Thiessen (celle di Voronoi)

L'interpolatore suddivide il piano con dei poligoni aventi la proprietà che tutti i punti interni sono più vicini al punto osservato interno che a tutti gli altri punti osservati.

La procedura comincia con la connessione mediante triangolazione dei punti noti; infine, linee di connessione sono tracciate perpendicolarmente ai lati di ogni triangolo nel loro punto mediano.

Il risultato è la tracciatura di regioni a valori costanti “attorno” ai punti campionati:
tassellazione (Poligoni di Thiessen o di Voronoi, celle di Dirichlet)

- la configurazione delle regioni (poligoni) è univocamente determinata dalla posizione dei punti;
- se i punti sono distribuiti regolarmente le regioni sono rettangolari.

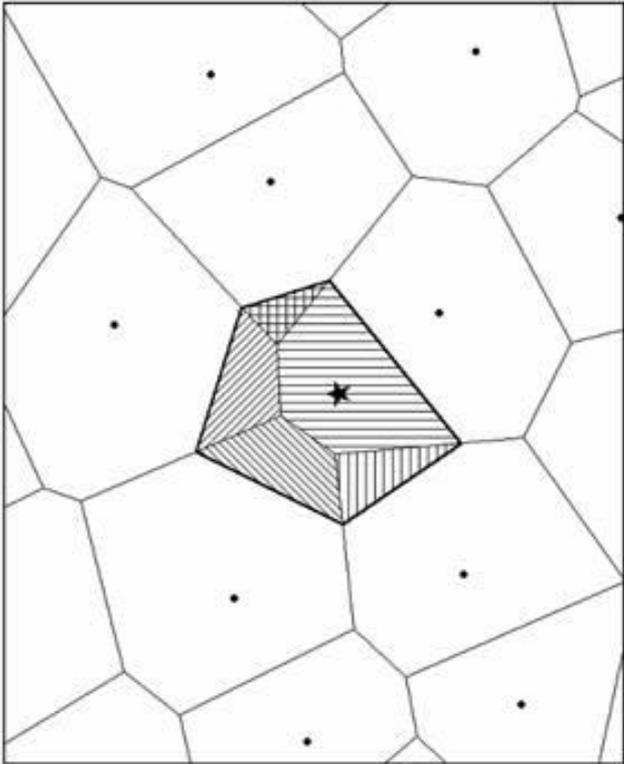


DICCA - Università degli Studi di Genova

Chiara Calligaris, Ph.D. – D.M.G. Università degli Studi di Trieste

NATURAL NEIGHBOUR

Il *natural neighbour* utilizza come pesi il rapporto tra la cella di Voronoi del punto la cui quota deve essere stimata e la porzione dei poligoni di Voronoi creati a partire dai punti di quota nota.



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

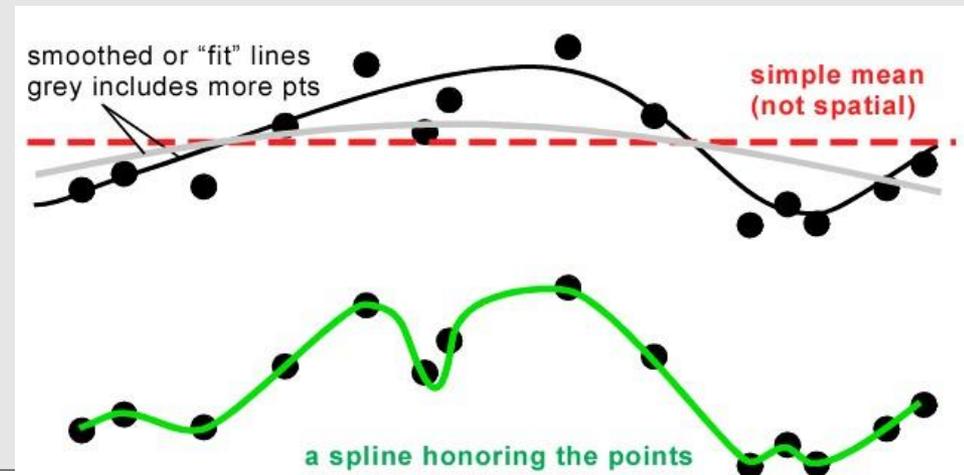
Dove $Z(x,y)$ è il valore del punto da stimare (x,y)
 w_i sono i pesi assegnati e $f(x_i, y_i)$ sono i valori noti di (x_i, y_i) .

I pesi w_i , sono determinati dalle porzioni di sovrapposizione tra poligoni di Voronoi.

Da notare la diversa assegnazione di pesi rispetto a IDW.

SPLINE

- Si tratta di un metodo di interpolazione che utilizza una funzione matematica sulla base dei valori dei punti scelti;
- Ha la caratteristica di fittare una superficie di minima curvatura ai punti noti e la superficie interpolata passa esattamente per i punti di input;
- Nella sua forma base è conosciuto come “*thin plate interpolation*”
- Il **metodo REGULARIZED** produce superfici smussate che non passano sempre per i punti noti. Alti valori del peso aumentano l'effetto ‘smussato’ della superficie;
- Il **metodo TENSION** cerca di onorare maggiormente i punti noti, producendo quindi superfici meno lisce rispetto al metodo precedente. Aumentando il peso la superficie appare più grossolanamente discretizzata;
- È possibile scegliere il numero di punti utilizzabili (locale o globale).

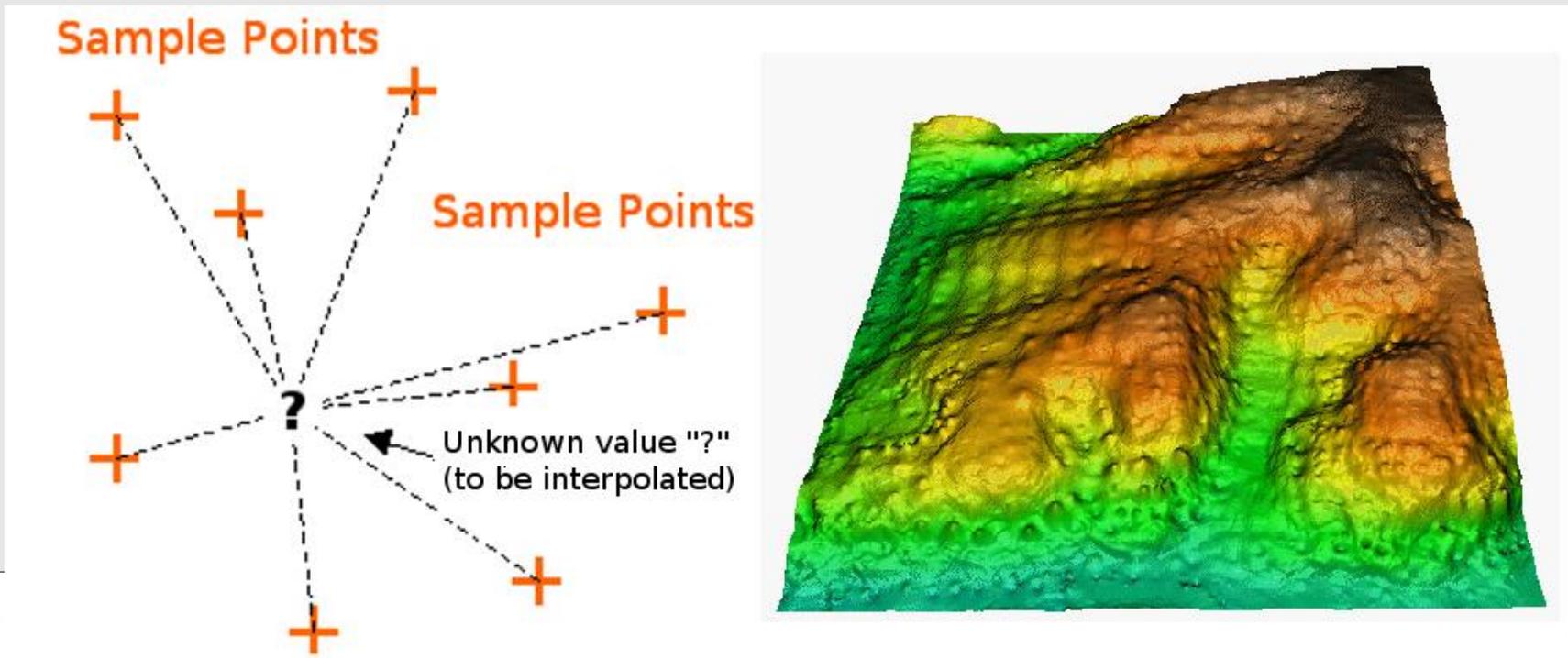


KRIGING

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 minerali.

È un metodo di interpolazione che si basa sull'autocorrelazione della grandezza, cioè l'assunto che la grandezza in oggetto vari nello spazio con continuità; detto in parole più semplici le cose più vicine sono più simili rispetto alle cose più lontane.

I valori in ogni punto rosso cambieranno, perché la variabile avrà una certa varianza locale. Comunque, la media generale, assumendo che si possa campionare ovunque, rimarrà costante e uguale alla media aritmetica dei punti rossi.



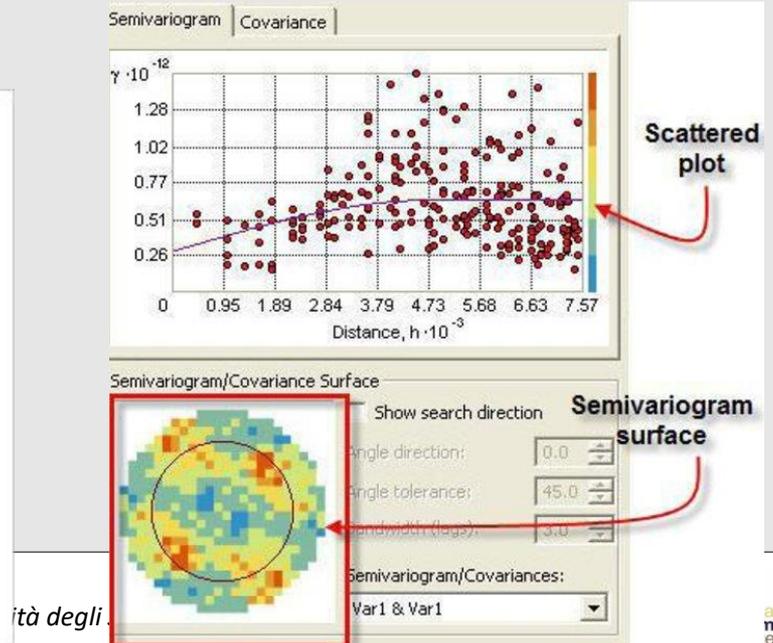
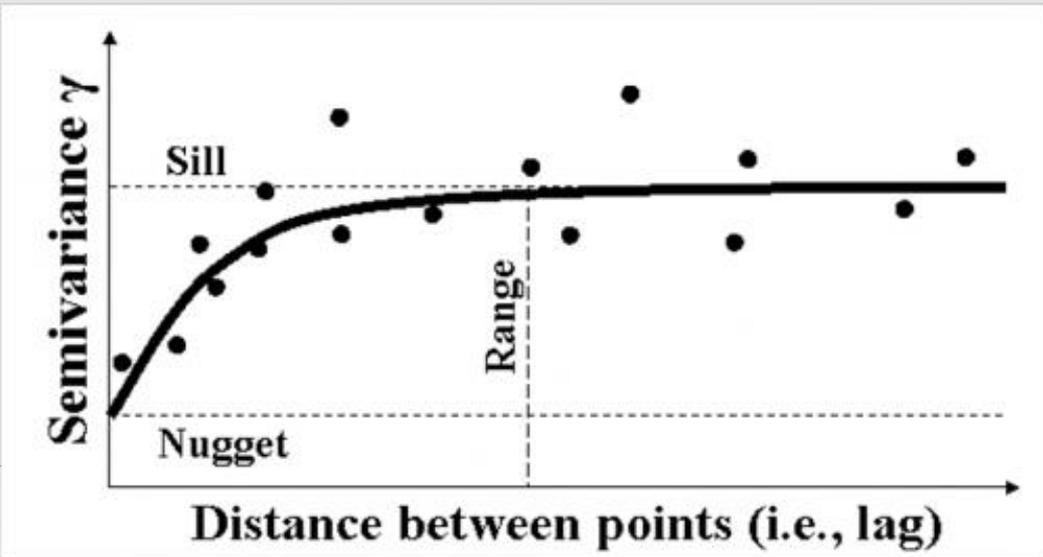
Se si assume che la variabile possa essere approssimata ad una distribuzione normale, conoscendo la media, è possibile calcolare il valore ignoto del pallino giallo aggiungendo o sottraendo la varianza. La varianza locale, la differenza tra il punto giallo e i vicini punti rossi dipenderà pertanto dalla distanza tra loro.

Il Kriging modella proprio questa differenza creando **il variogramma**, che consiste in una funzione che ci dice quant'è la varianza del pallino giallo in relazione alla sua distanza dai vicini rossi.

Dove la varianza è una funzione che fornisce una misura della variabilità dei valori assunti dalla variabile stessa.

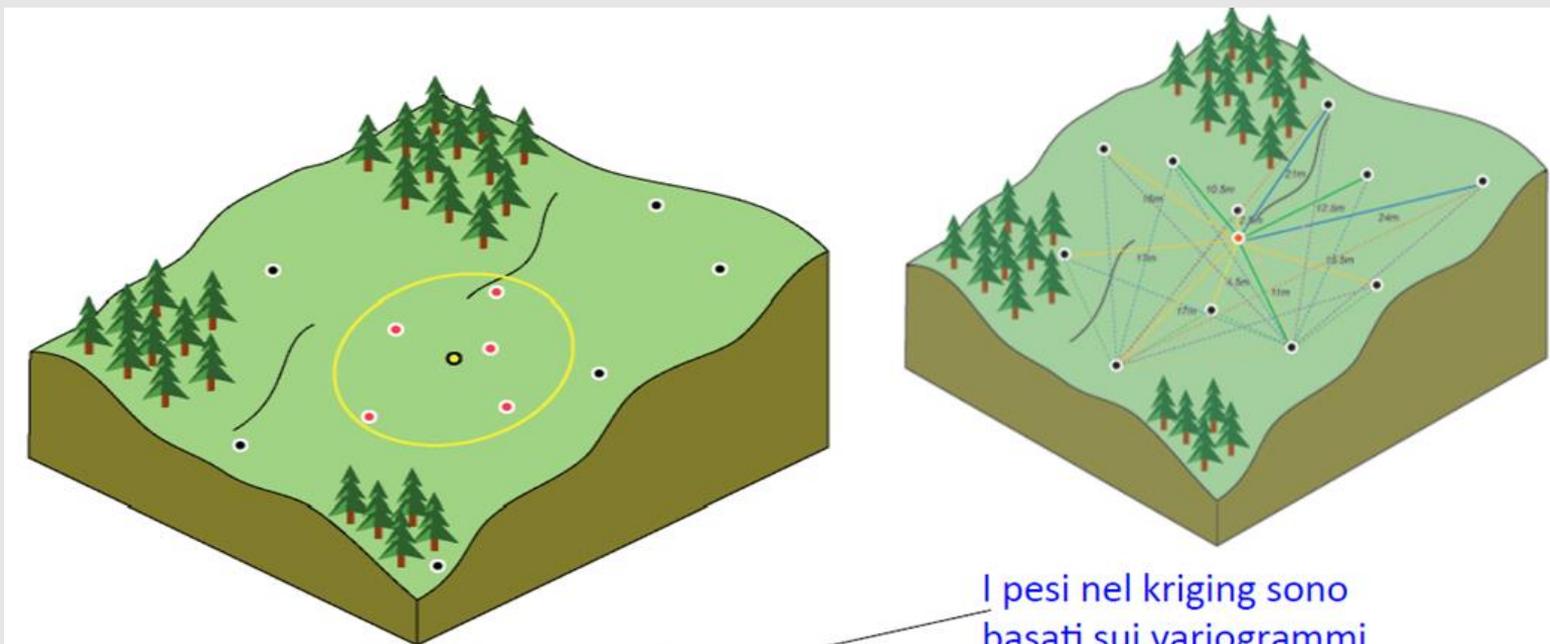
Il variogramma è semplicemente una rappresentazione sintetica (media) della nuvola di punti delle semivarianze (la semivarianza è la metà della somma delle varianze (quadrato della deviazione standard) tra ogni valore Z e ognuno dei punti alla stessa distanza) vs distanze.

La nube di punti contiene tutte le relazioni spaziali nei dati per tutte le possibili distanze tra i campioni, ma non è una funzione continua. E' difficile interpretarla e comprendere l'esistenza di correlazioni spaziali.



Questo metodo 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.

La versione standard di Kriging è chiamata **Ordinary Kriging (OK)**. Può essere vista come un Inverse Distance Weighted molto più sofisticato: i pesi anziché determinati in modo arbitrario, riflettono la struttura di autocorrelazione spaziale definita dalla funzione di variogramma.



I pesi nel kriging sono basati sui variogrammi

$$\hat{Z}(\text{yellow}) = \sum_{i=1}^5 \lambda_i(d_{\text{yellow}, \text{red}_i}) \times Z(\text{red}_i)$$

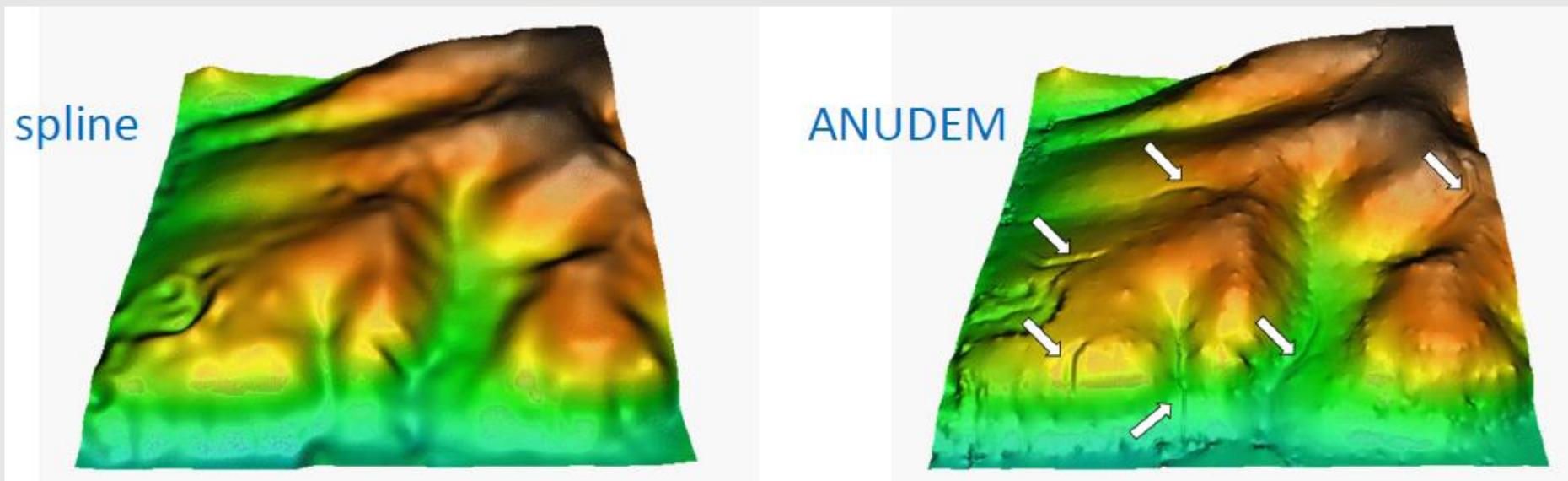
ANUDEM

Conosciuto anche come *Topo To Raster* è un metodo di interpolazione sviluppato da Michael Hutchinson (1988) specificamente sviluppato per la creazione di DEM idrologicamente corretti;

Uno dei pochi algoritmi in grado di gestire intelligentemente curve di livello e altre utili informazioni lineari (e.g. corsi d'acqua, creste);

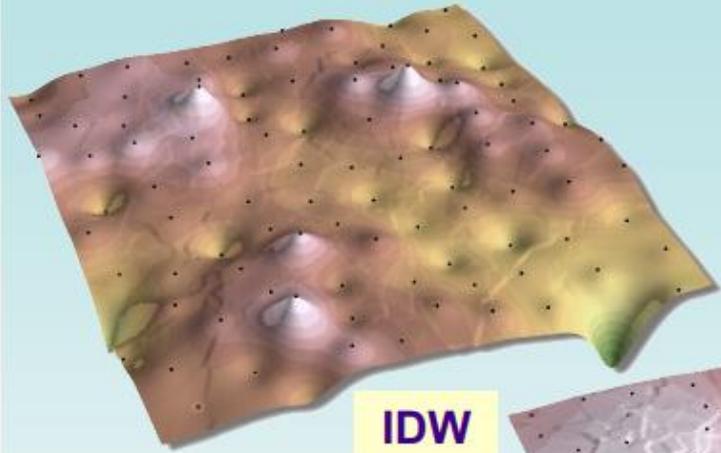
È una tecnica di interpolazione iterativa alle differenze finite;

È essenzialmente un *thin plate spline* con modifiche che garantiscono la rappresentazione di bruschi cambiamenti di superficie.

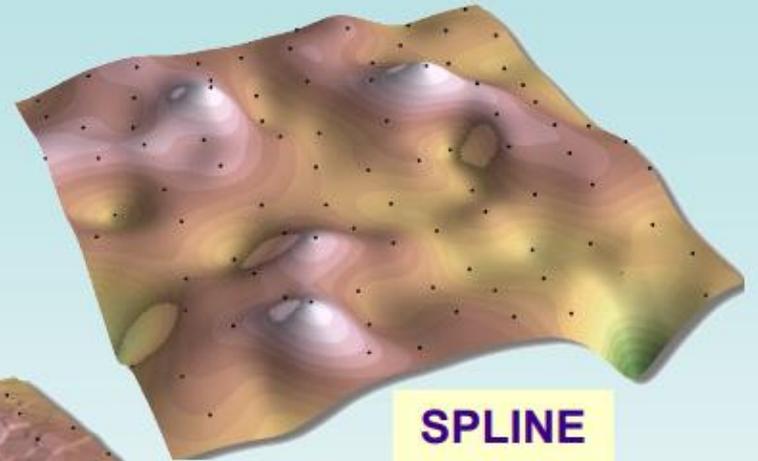


Cavalli and Tarolli, 2017

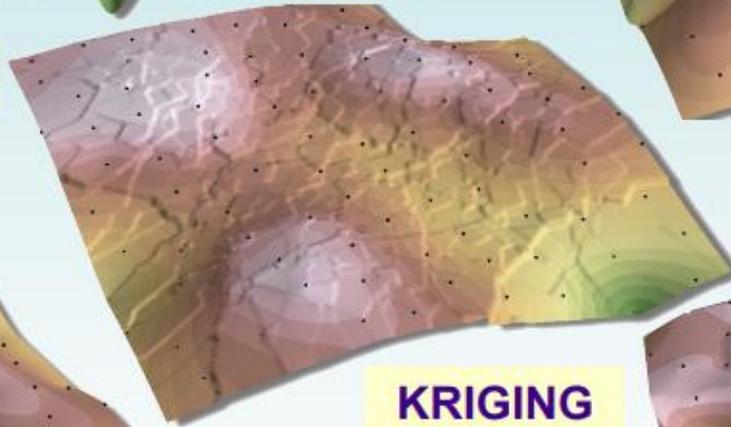
RISULTATI A CONFRONTO



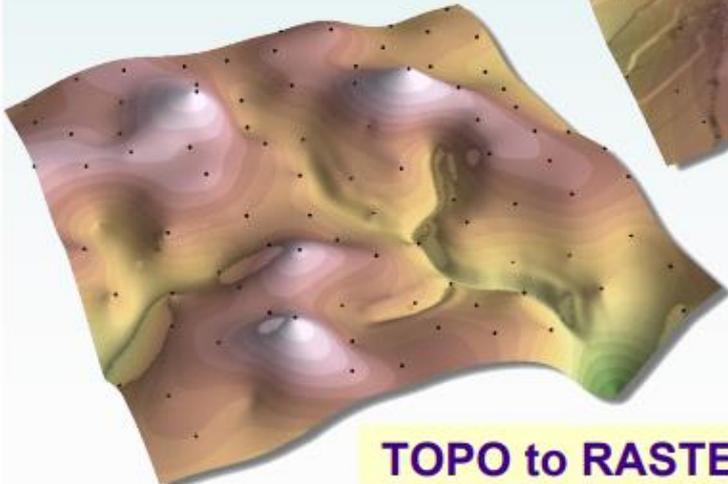
IDW



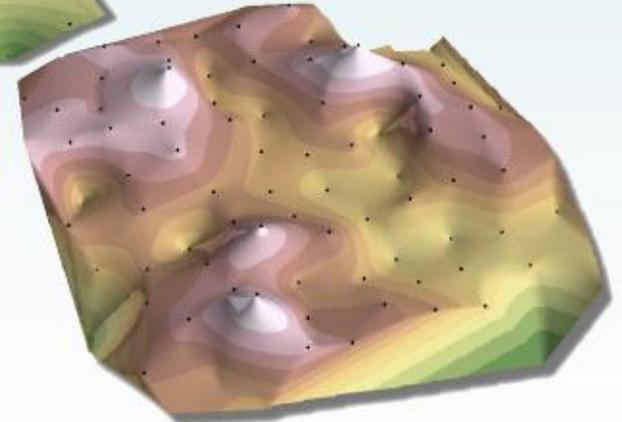
SPLINE



KRIGING



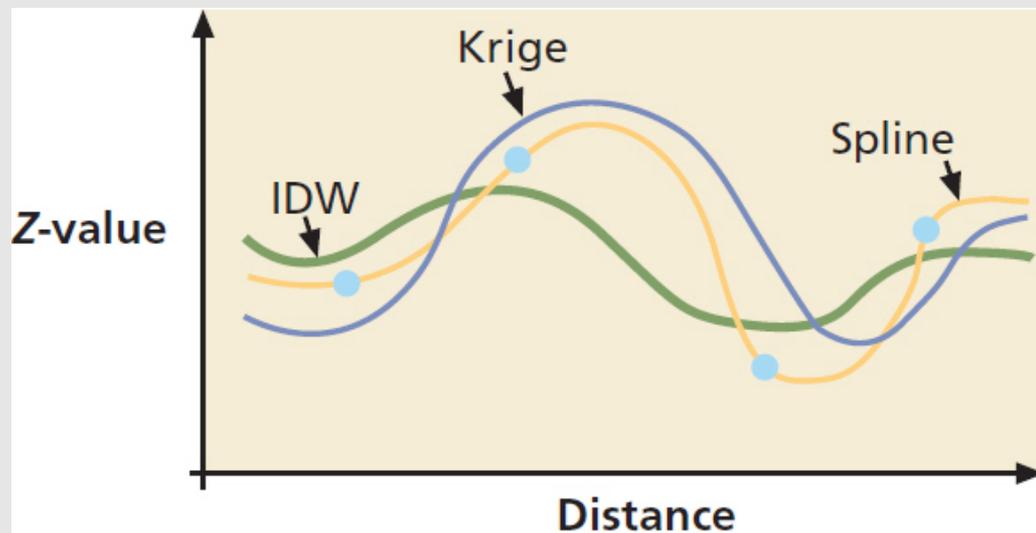
TOPO to RASTER



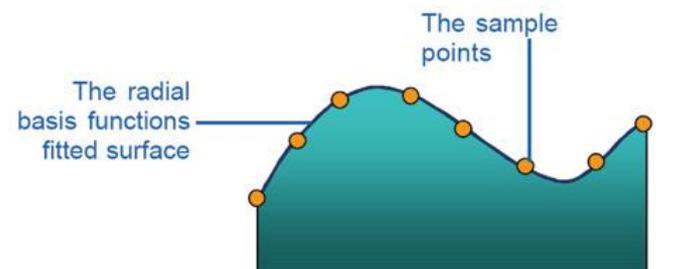
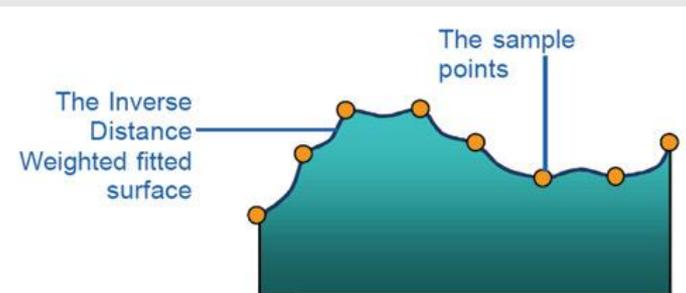
NATURAL NEIGHBOR

www.esri.com

METODI A CONFRONTO



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.



Johnston et al., 2003, p. 126