esercitazione 4

interpolazione spaziale stocastica (geostatistica)
pericolosità sismica a Belluno, 2006

per saperne di più su
- pericolosità sismica nella provincia di Belluno
- interpolazione spaziale con kriging

in che modo
- studio del variogramma
- kriging ordinario
- kriging universale

facendo attenzione a
- scelta dei parametri per l'algoritmo

bibliografia
- de Smith MJ, Goodchild MF, Longley PA, Geospatial Analysis. A comprehensive Guide to Principles, Techniques and Software Tools, 3rd ed., Splint, 2009 http://www.spatialanalysisonline.com/

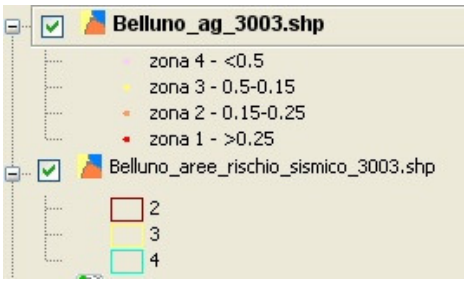
dati a disposizione
- tabellari: Belluno_ag_3003.xls
- vettoriali: Belluno_ag_3003.shp, Belluno_aree_rischio_sismico_3003.shp
- sistema di riferimento: epsg 3003 Monte Mario / Italy Zone 1

software
- excel
- jmp – analisi statistica – trial 30 giorni
- gvsig – gis – free open source

piccoli dettagli utili
- link alle informazioni sulla pericolosità sismica in Italia http://zonesismiche.mi.ingv.it/

parte 1
analisi esplorativa dei dati

aprite gvSIG



definite una nuova vista con epsg 3003 e aggiungete i layer

- Belluno_ag_3003.shp

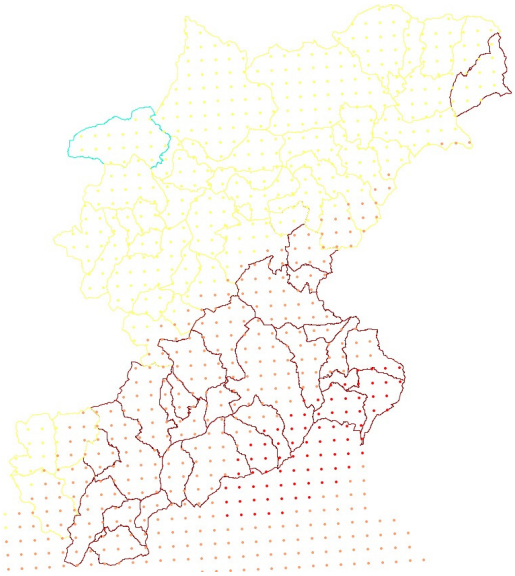
  dati puntuali di pericolosità sismica al 2006
  pubblicati dall'INGV Istituto Nazionale di Geofisica e Vulcanologia

  espressi in ag = frazione di accelerazione di gravità
  con probabilità di eccedenza del 10% in 50 anni
  riferita a suoli rigidi

  è una misura dell'accelerazione attesa
  a cui è sottoposto il suolo
  calcolata su una griglia regolare di punti

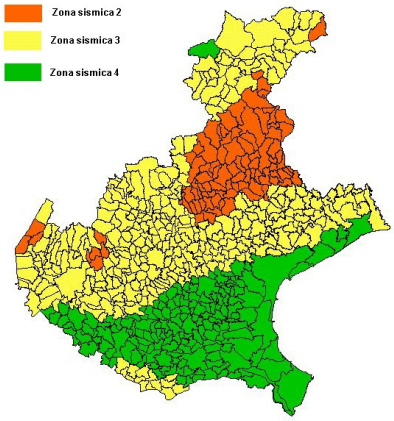
  classi standard:

< 0.05	zona 4 pericolosità minima
0.05 - 0.15	zona 3
0.15 - 0.25	zona 2
> 0.25001	zona 1 pericolosità massima



- Belluno_aree_rischio_sismico_3003.shp

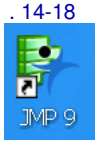
  classificazione sismica dei comuni
  ordinanza del Presidente del Consiglio dei Ministri n.3274/03



obiettivo:
stimare la superficie di ag nella provincia di Belluno per il 2006
confrontarla con la classificazione del 2003

c'è qualche comune per cui andrebbe rivista la classificazione sismica?

aprite Belluno_ag_3003.xls con

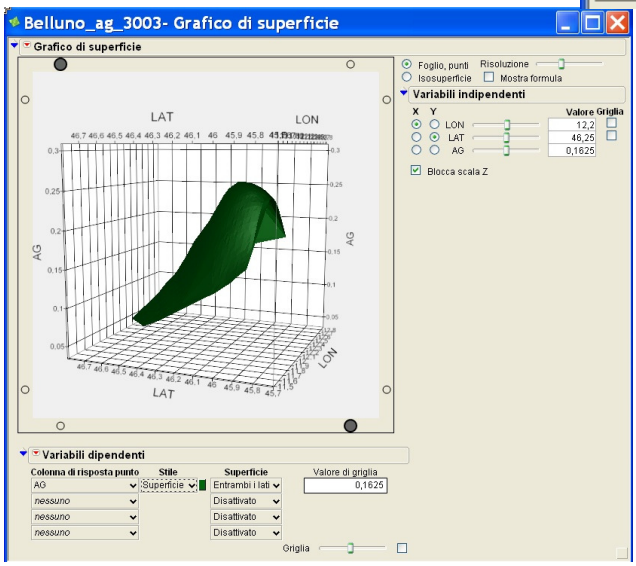


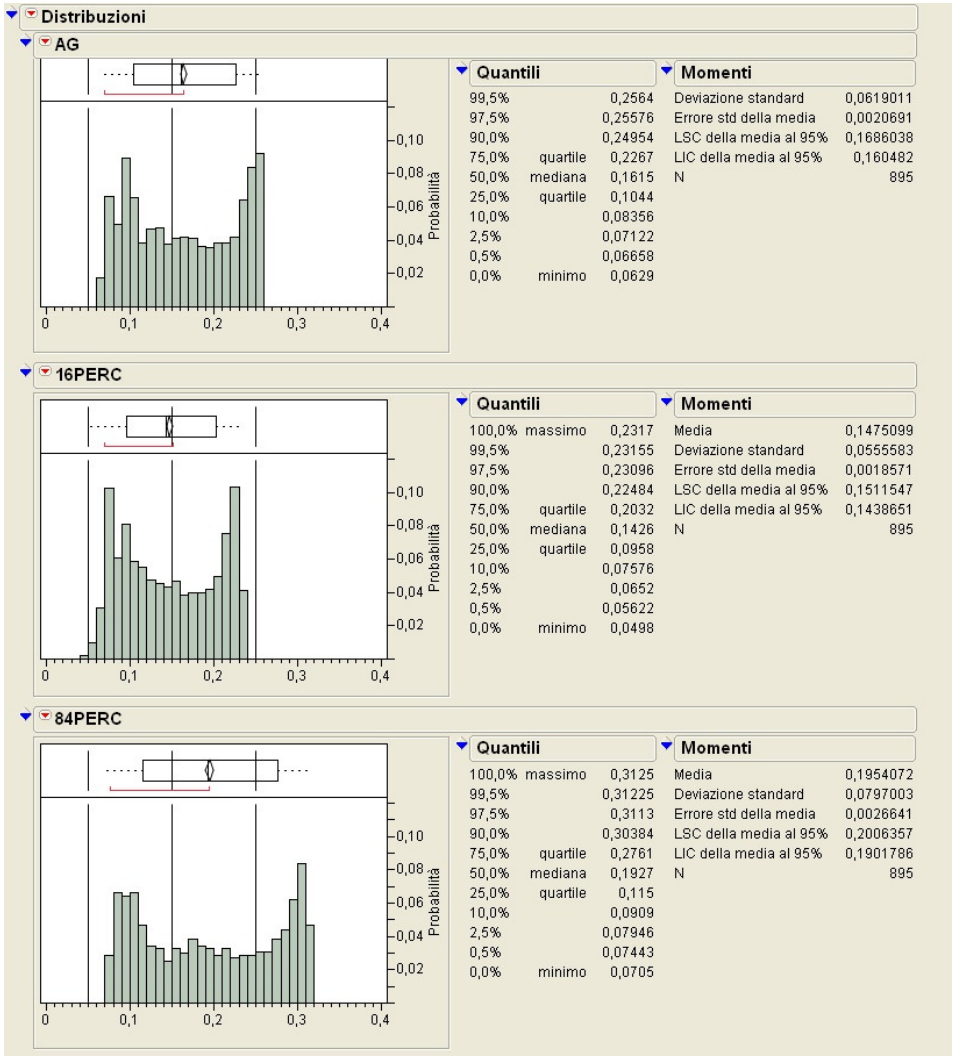
i dati di pericolosità sismica 2006 sono presentati con tre valori
16mo percentile, ag mediano, 84mo percentile della distribuzione dei valori di ag misurati ripetutamente in ciascun punto

questo significa che i valori di ag hanno un grado di incertezza

Belluno_ag_3003

Righe 895/0	ID	LON	LAT	AG	16PERC	84PERC
1	44546	12,0704	46,6687	0,0629	0,0498	0,0705
2	44560	12,4793	46,6727	0,0784	0,0734	0,0858
3	44561	12,5085	46,673	0,0795	0,0744	0,0872
4	45100	12,071	46,6486	0,0644	0,0526	0,0722

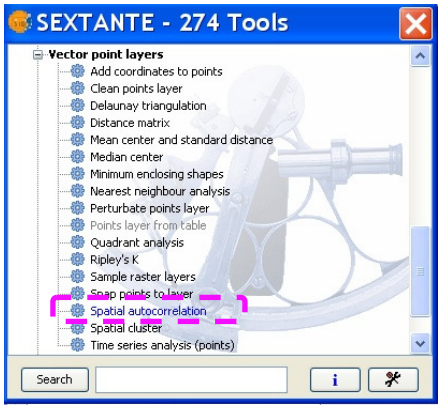


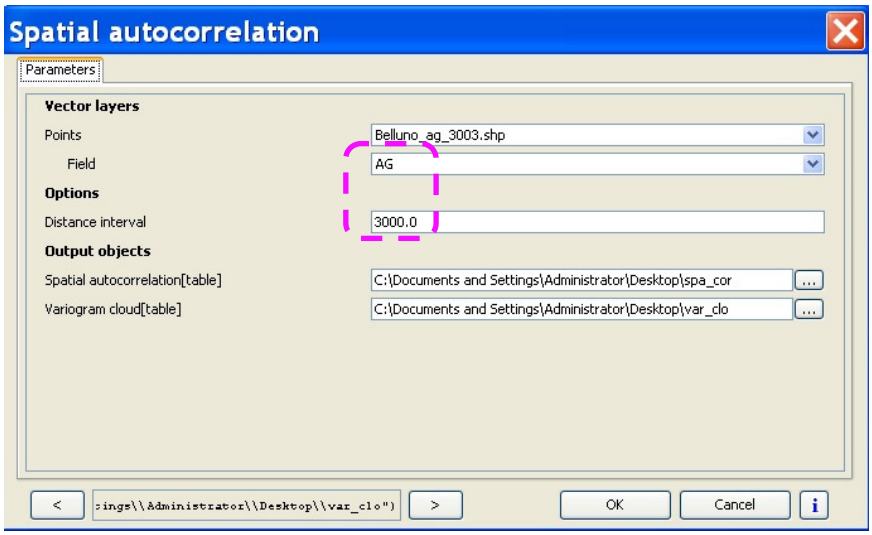


parte 2
autocorrelazione spaziale ?
studiamo il variogramma

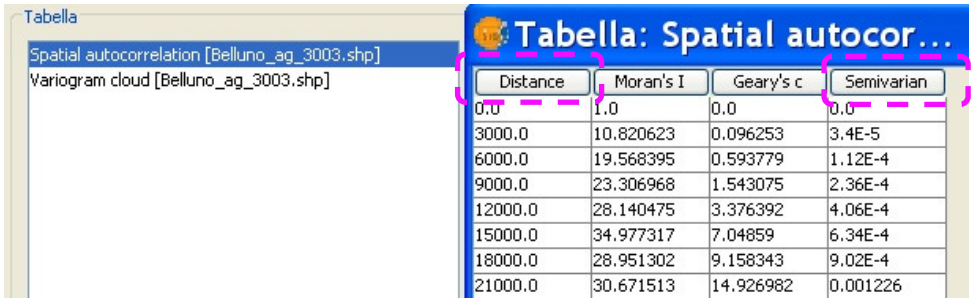
studiate le distanze tra i punti
qual è la distanza minima?
qual è un raggio ragionevole da usare come "lag" ("Distance interval") del variogramma? 3km? 4km?

Sextante → Vector point layers → Spatial autocorrelation





ottenete una tabella dati che vi consente di visualizzare il variogramma campionario



Distance	Moran's I	Geary's c	Semivarian
0.0	1.0	0.0	0.0
3000.0	10.820623	0.096253	3.4E-5
6000.0	19.568395	0.593779	1.12E-4
9000.0	23.306968	1.543075	2.36E-4
12000.0	28.140475	3.376392	4.06E-4
15000.0	34.977317	7.04859	6.34E-4
18000.0	28.951302	9.158343	9.02E-4
21000.0	30.671513	14.926982	0.001226

la colonna "Distance" rappresenta intervalli regolari di distanze, multipli del "Distance interval" fissato

la colonna "Semivarian" contiene la semi-varianza calcolata sui punti distanti quanto definito in riga

in pratica: vengono calcolate tutte le possibili distanze tra coppie di punti
poi queste distanze sono classificate in base agli intervalli scelti (lag)

es. prima riga: punti entro un raggio di 1.500 metri (nel nostro caso ce n'è sempre 1)
seconda riga: punti che distano tra loro dai 1.500 ai 4.500 metri
…

all'interno di ogni classe di distanza (riga) si calcola la semi-varianza sperimentale

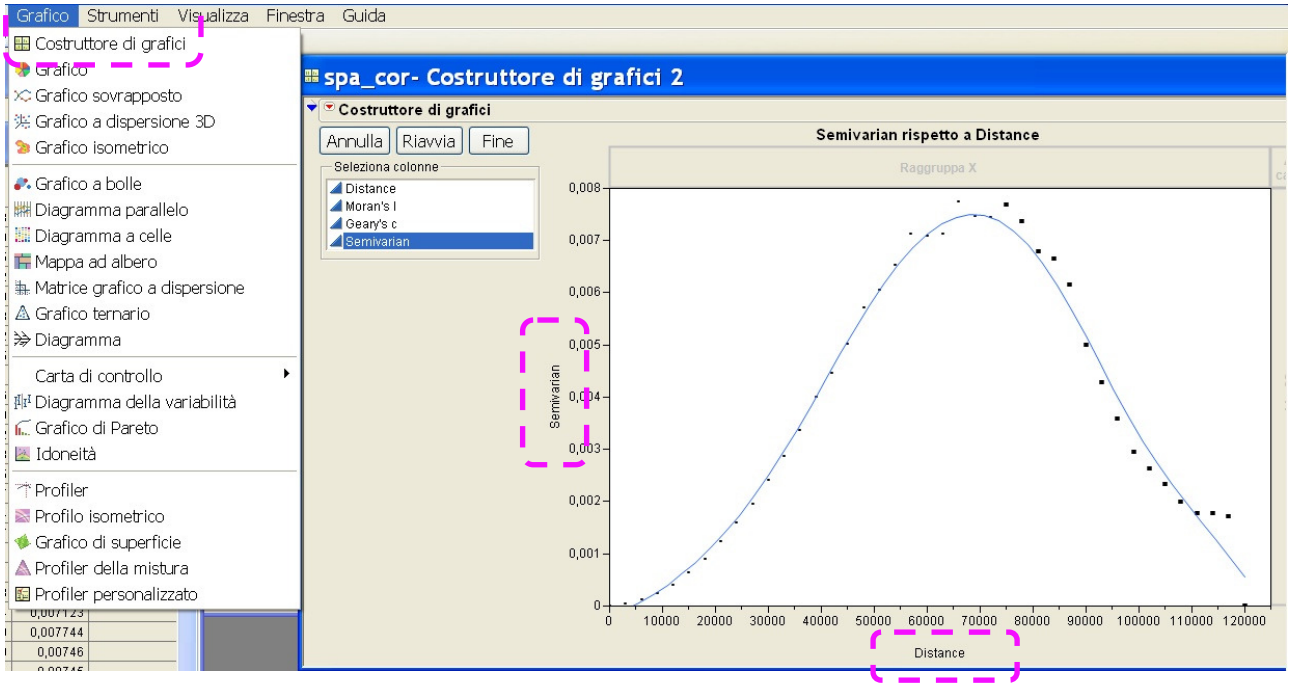
$$\gamma(h) = \frac{1}{2N(h)} \sum_{dij=h-\Delta/2}^{dij=h+\Delta/2} (z_i - z_j)^2$$

- si tiene conto della diversità di valore tra i punti (differenza)
- la si eleva al quadrato per rendere tutti i valori positivi e dare maggior enfasi a quelli grandi

- si divide per 2 per non contare due volte la stessa distanza calcolata tra due punti (da A a B – da B ad A)

aprite il file **spa_cor.dbf** con

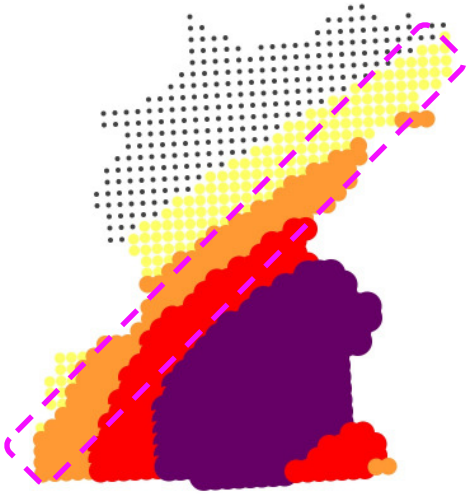
e create un grafico X-Y con i campi “Distance” e “Semivarian”



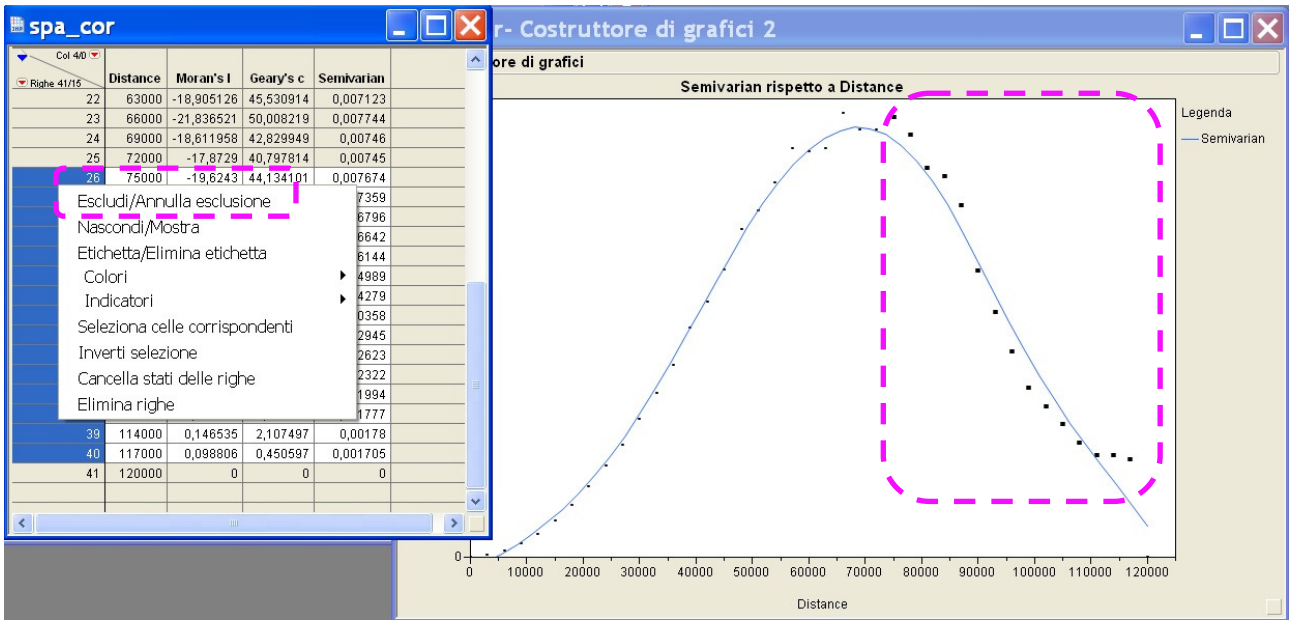
vi viene proposta in automatico una curva interpolante

osservate che oltre i 70 km la variabilità diminuisce … perché?
cercate la spiegazione nella visualizzazione tematizzata dei punti

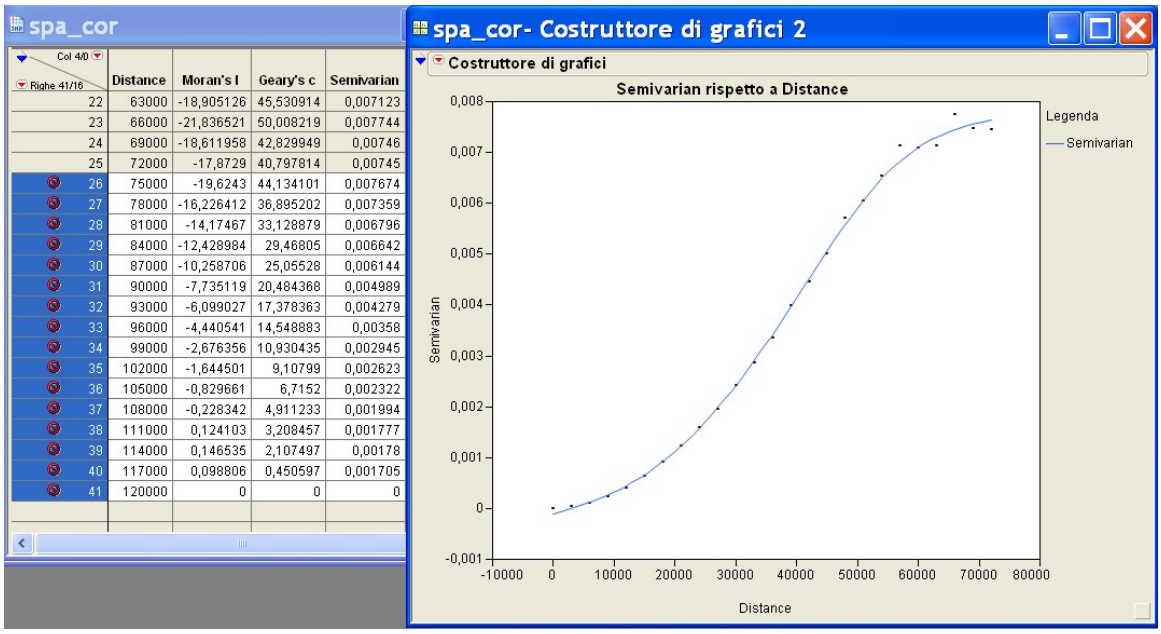
le distanze maggiori, oltre i 70 km, si hanno per punti che si collocano nella fascia di valori intermedi di ag,
tutti abbastanza simili tra loro



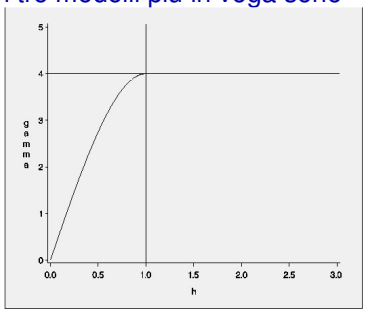
selezionate i punti nel grafico oltre i 75 km di distanza ed escludeteli dall'analisi
così modellerete la variabilità spaziale entro un raggio limitato; dovrete ricordarvene al momento di lanciare la procedura di kriging



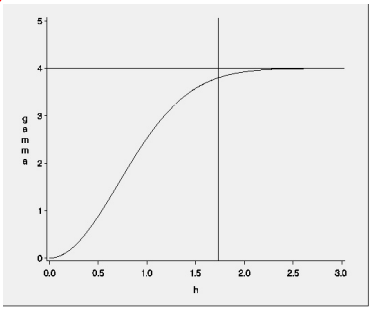
scegliete il modello teorico di semivariogramma più adatto a quello osservato
è un'equazione in funzione della distanza



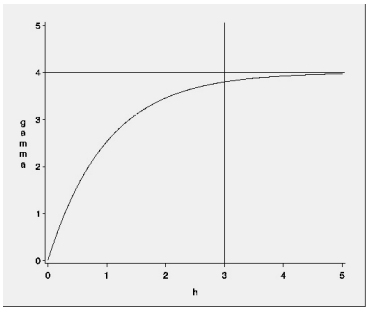
i tre modelli più in voga sono



sferico

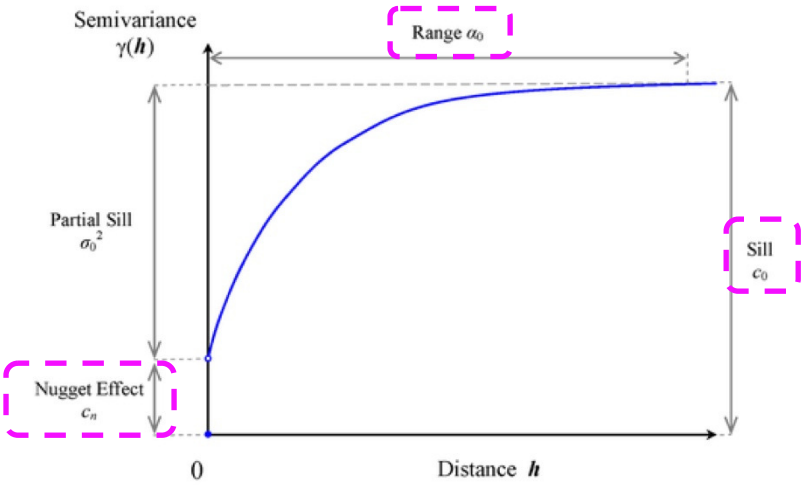


gaussiano



esponenziale

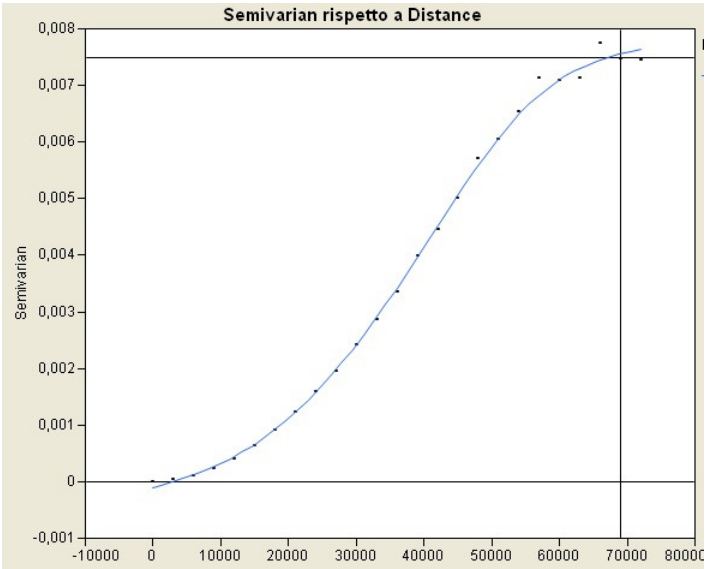
determinate i parametri chiave del semivariogramma



range distanza oltre la quale non vi è più incremento di variabilità

sill valore massimo di variabilità

nugget variabilità osservata a distanza 0
- si attribuisce a errori di misura o a variabilità non legata alla compontente spaziale del fenomeno
- un valore elevato rende i kriging non significativo
- se nugget > 0 il kriging non è un interpolante esatto



spa_cor

Col 4/0 / Righe 41/16	Distance	Moran's I	Geary's c	Semivarian
1	0	1	0	0
2	3000	10,820623	0,096253	0,000034

range 69000 metri

sill 0,0075

nugget 0 ?
la distanza minima tra due punti è 2km

non conoscete la variabilità a distanze più basse

ma certamente non è pari a 0
perché anche i valori puntuali di AG vengono presentati in 3 modi (16mo percent. – mediana – 84mo percent.)
e hanno dell'incertezza

prendete la metà del valore osservato nel primo intervallo di distanze in cui ci sia più di un punto
0,000017

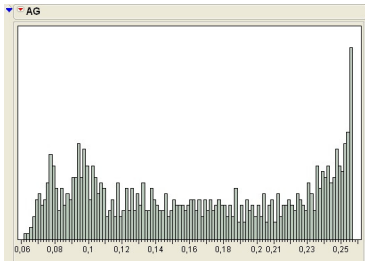
ora avete "modellato" la variabilità spaziale osservata per la variabile da interpolare e avete tutti gli elementi per

- procedere con l'algoritmo di kriging
- interpolando i valori di un fenomeno continuo nello spazio
- attraverso una procedura statistica, che tiene conto della variabilità spaziale del fenomeno
- che consente di calcolare gli errori di stima
- perché ipotizza che i valori puntuali osservati siano un campione statistico da una superficie ignota

parte 3
kriging ordinario

i valori nei punti da stimare sono una media ponderata dei punti noti vicini; i pesi non vengono scelti dall'utente, ma dipendono dalla struttura dell'autocorrelazione spaziale osservata, espressa dal modello di variogramma scelto (es. maggiore è la varianza "nugget", minori sono i pesi attribuiti ai valori più vicini)

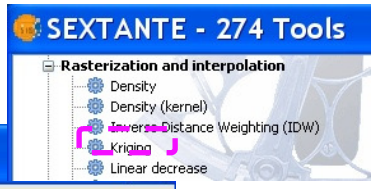
$$z_p = \sum_{j=1}^{n} \lambda_j z_j \quad \text{con la condizione che} \quad \sum_{j=1}^{n} \lambda_j$$

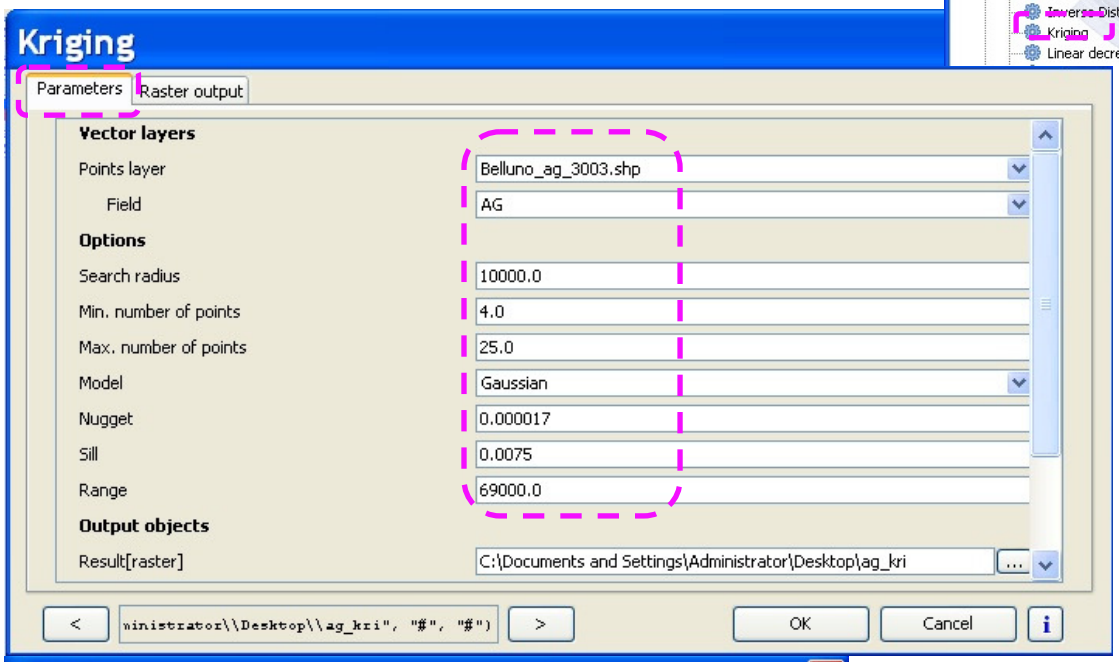


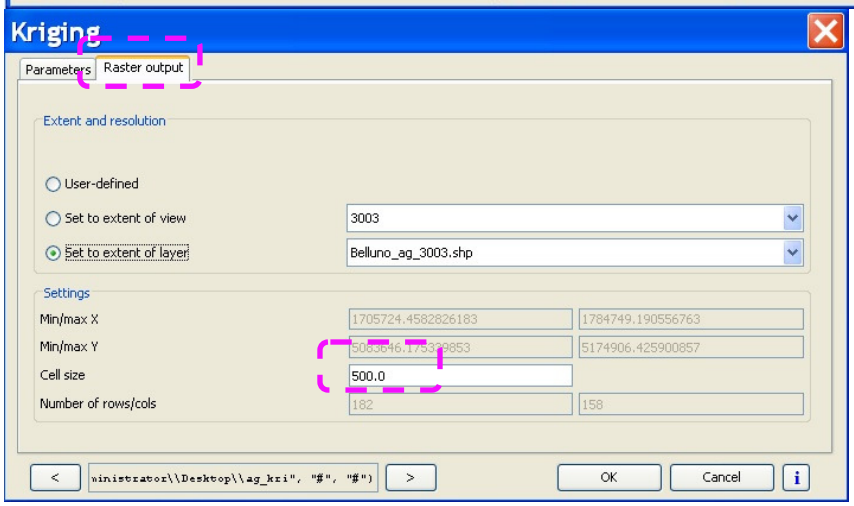
si basa sull'ipotesi che i dati siano distribuiti "normalmente"
nel nostro caso non è proprio così

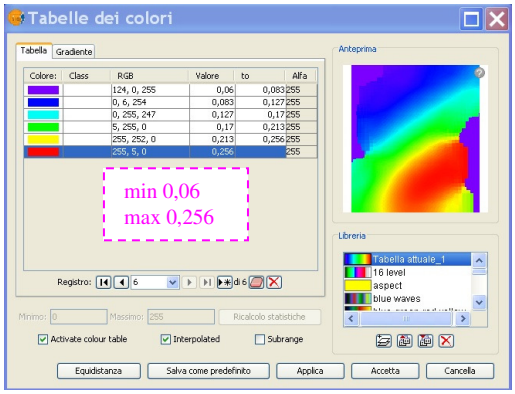
per fortuna questa "mancanza" influisce soltanto sulla bontà della mappa degli errori che accompagna la mappa delle stime

Sextante → Rasterization and interpolation → Kriging

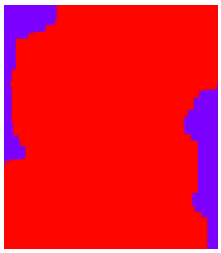


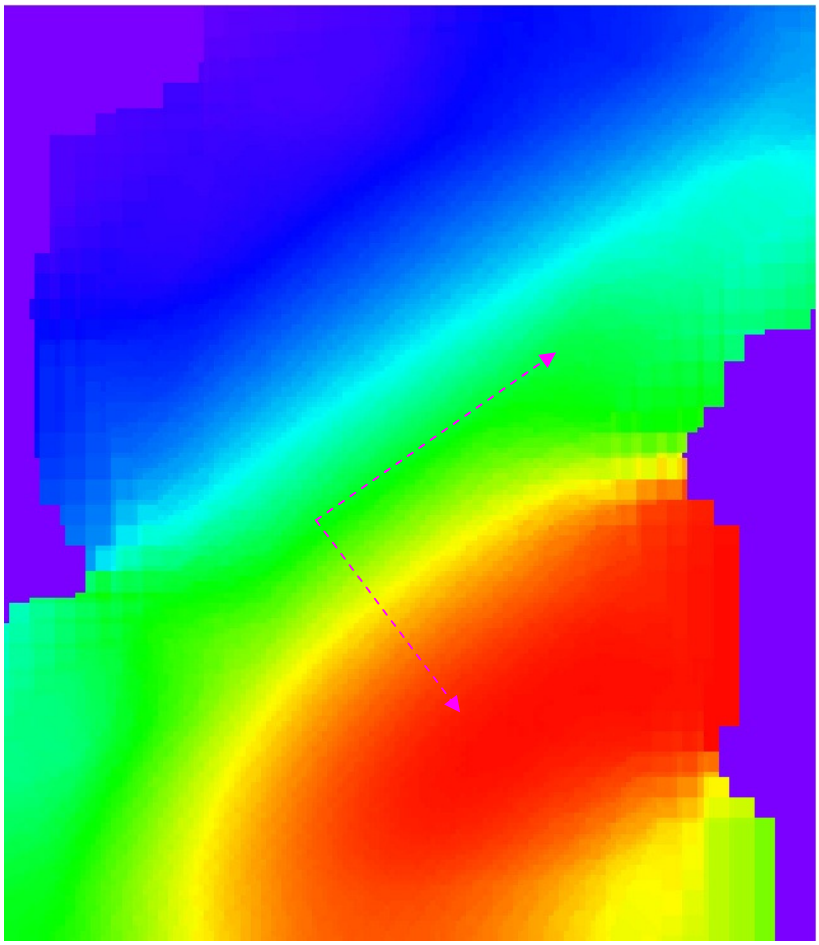






la mappa della variabilità non è molto significativa, perché i punti sono equidistanti, non ci sono zone più dense, in cui la qualità della stima è maggiore, e zone con punti più radi, in cui la probabilità di commettere un errore di stima è più elevate; inoltre, la distribuzione dei valori originali non è normale



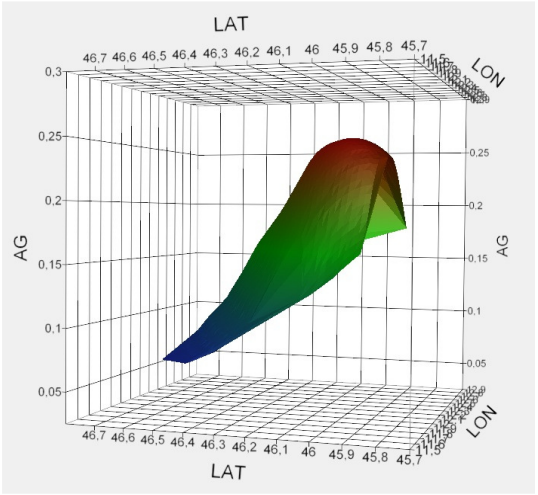


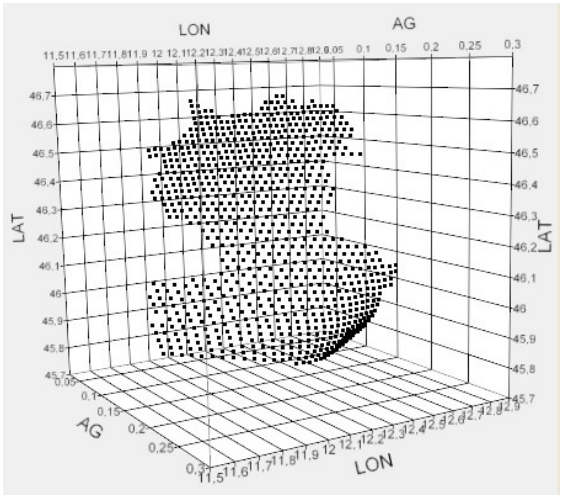
bello eh?
ma è anche significativo?

un'ipotesi fondamentale alla base della procedura appena applicata è la stazionarietà

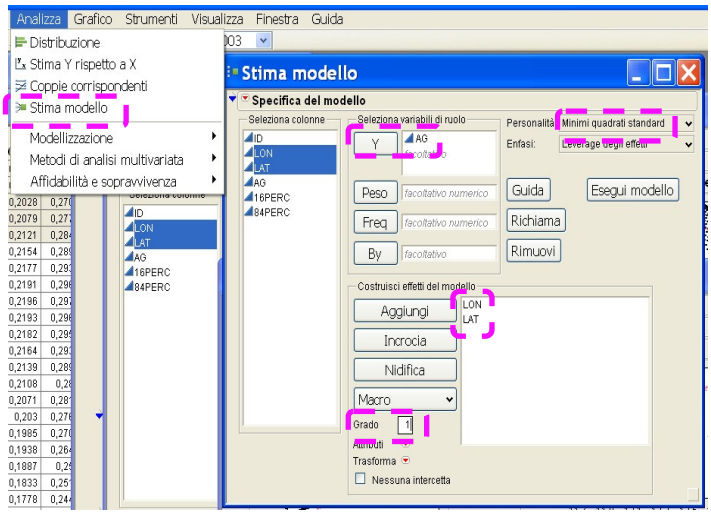
- di primo ordine: la media del fenomeno è costante nello spazio considerato, non dipende dalla posizione
- di secondo ordine: l'autocorrelazione dipende soltanto dalla distanza tra le osservazioni, non dalla direzione del vettore che le separa

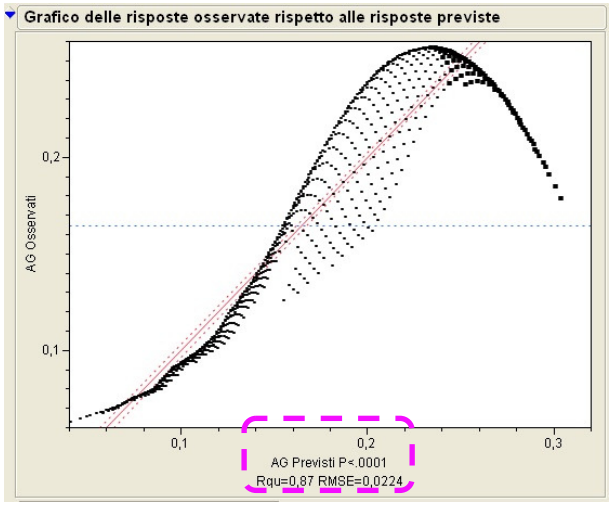
ma … i valori di ag mostrano un evidente trend a crescere, con un piccolo trend in discesa in una zona particolare



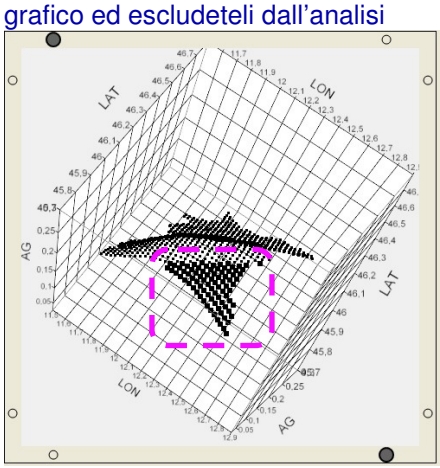


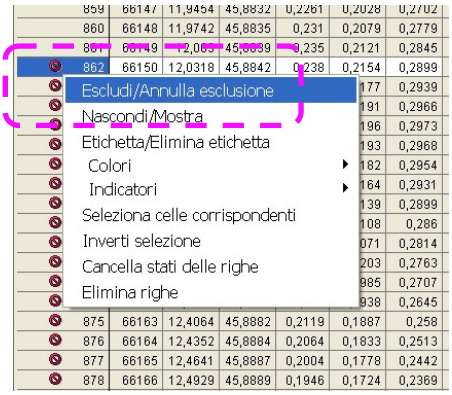
calcolate il modello della superficie di trend



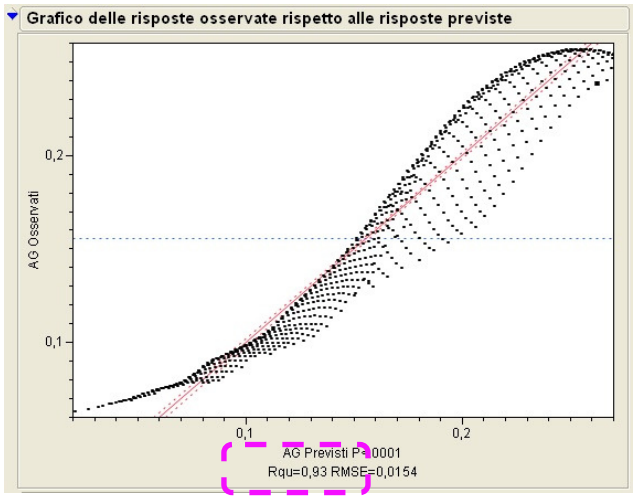


concentrate l'analisi sull'area per la quale i valori di ag sono sempre in crescita: selezionate gli altri punti nel grafico ed escludeteli dall'analisi





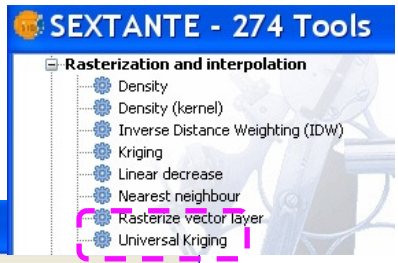
ripetete l'analisi: vi è un piano di trend molto ben definito

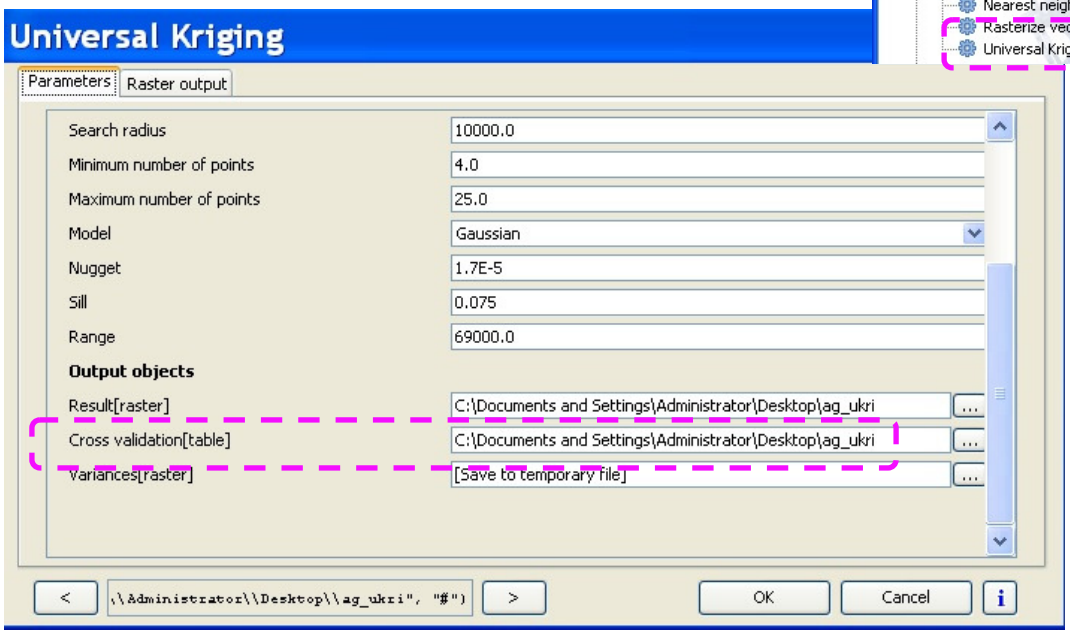


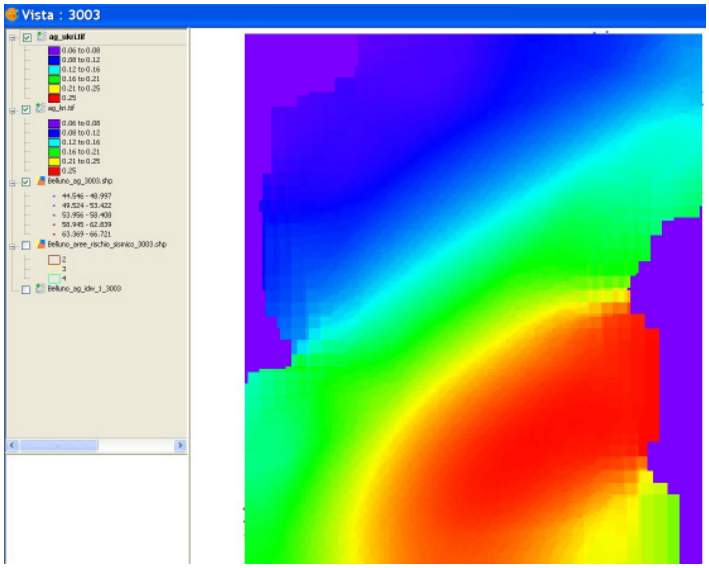
parte 4
kriging universale

tiene conto del trend presente nei dati, introducendo nella procedura un modello di regressione per il trend

Sextante → Rasterization and interpolation → Universal Kriging







confrontate i due kriging: ordinario e universale; vedete qualche differenza?

che cosa accadrebbe se

- il raggio che delimita i punti da usare per la stima fosse maggiore?
- la maglia dei pixel per cui stimare i valori fosse più fitta?

importate la tabella di Cross-validation in

per ogni punto (riga) contiene (in colonna) i valori osservati e quelli stimati con un kriging sul resto dei punti

ag_ukri

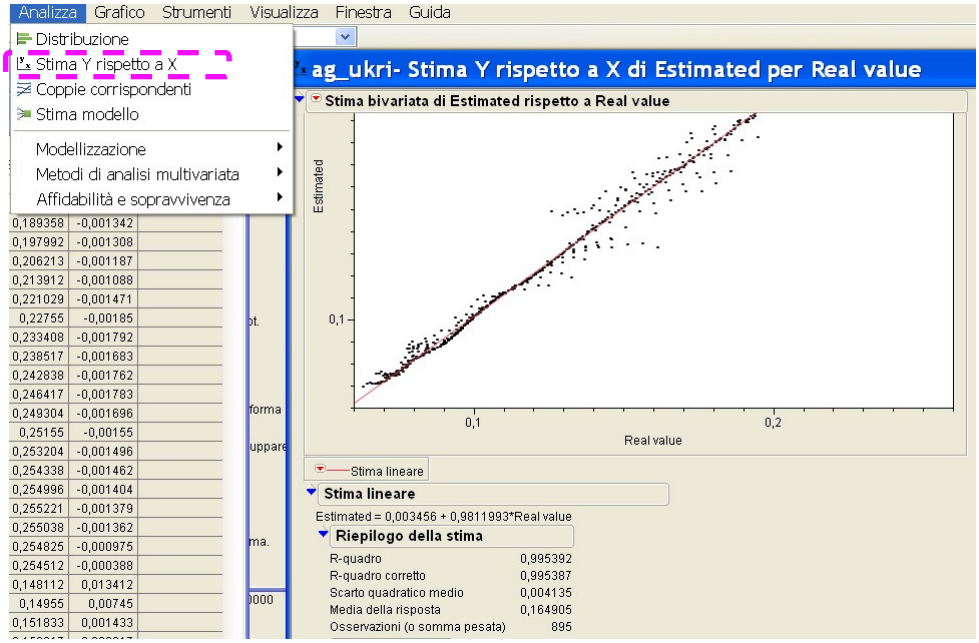
Righe 895/0	X	Y	Real value	Estimated	Difference
1	1734872,48	5173027,64	0,0629	0,069481	0,006581
2	1766129,83	5174773,89	0,0784	0,088417	0,010017
3	1768361,68	5174906,42	0,0795	0,089954	0,010454
4	1735005,52	5170796,04	0,0644	0,070668	0,006268
5	1737238,58	5170916,96	0,0654	0,071229	0,005829
6	1761792,79	5172301,53	0,0784	0,086575	0,008175
7	1764026,01	5172421,3	0,0796	0,087517	0,007917
8	1766259,2	5172541,9	0,0813	0,088379	0,007079
9	1768491,87	5172674,44	0,0841	0,089787	0,005687

che relazione c'è tra i valori osservati e quelli stimati?

il legame lineare tra valori osservati e stimati è quasi perfetto

il coefficiente R-quadro è quasi pari a 1

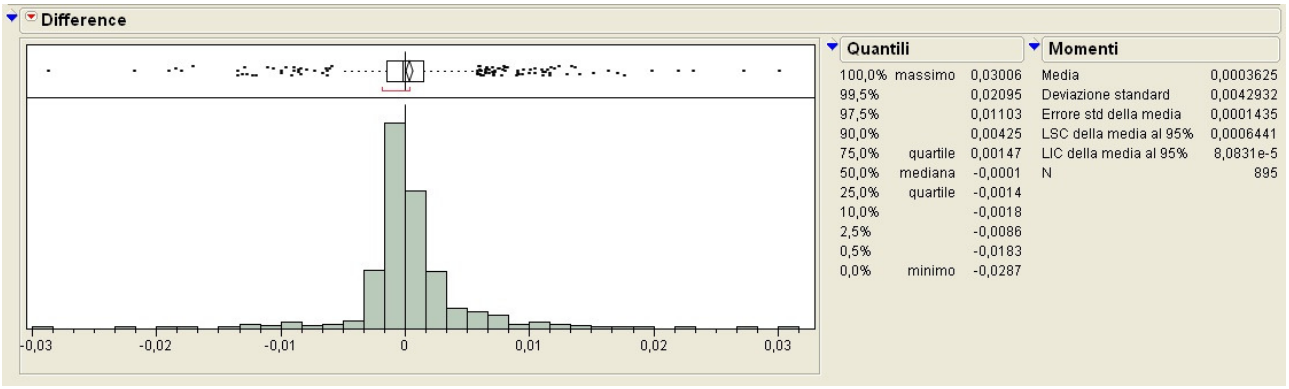
l'angolo con l'asse X della retta interpolante è molto vicino ai 45°



Difference = Estimated – Real-value
come sono distribuiti i residui? bene!

molto concentrati intorno allo zero, con distribuzione simmetrica

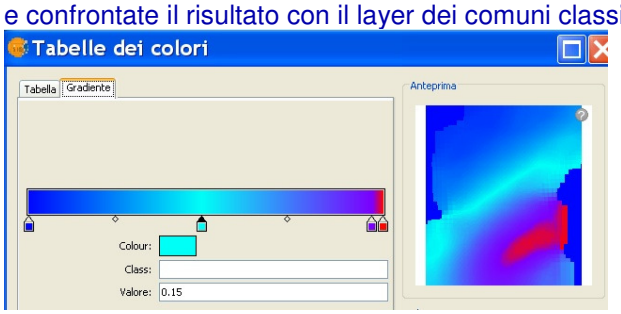
osservate come la moda sia sui valori negativi molto vicini allo zero → leggera tendenza alla sottostima

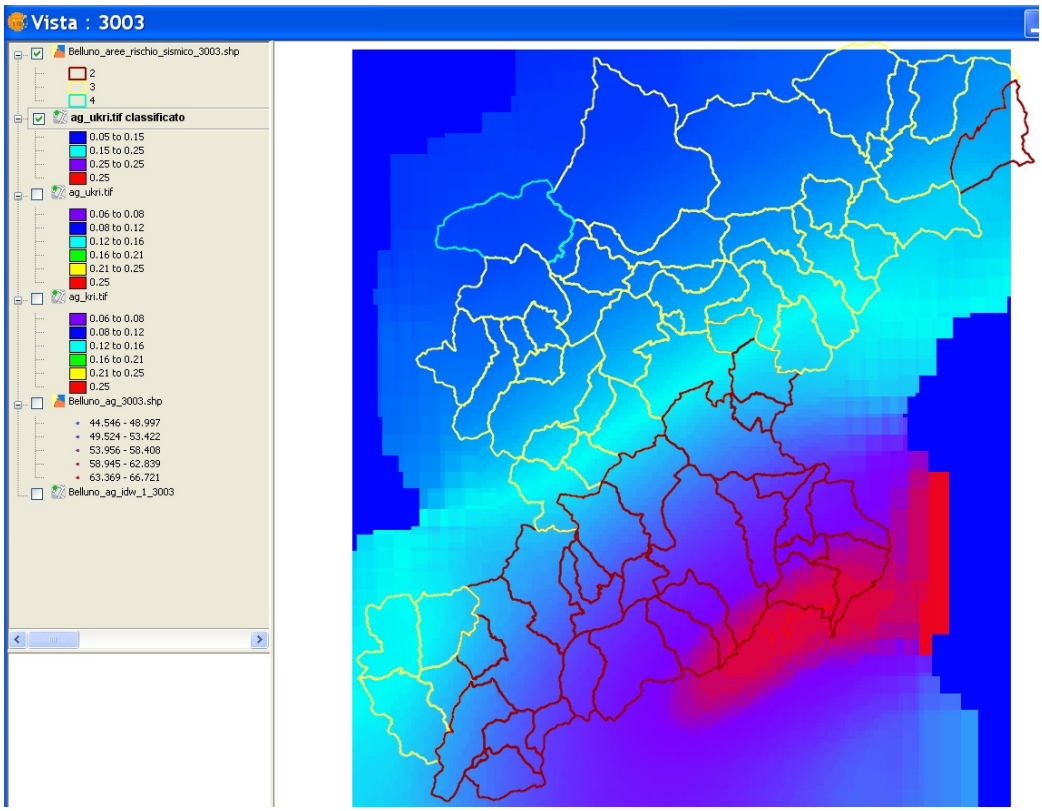


classificate i valori del layer interpolato
secondo le classi standard:

< 0.05	zona 4 pericolosità minima
0.05 - 0.15	zona 3
0.15 - 0.25	zona 2
> 0.25001	zona 1 pericolosità massima

e confrontate il risultato con il layer dei comuni classificati nel 2003





in questo esercizio non abbiamo trattato alcuni argomenti fondamentali

- modello concettuale della superficie da interpolare
- analisi preliminare dei **valori anomali** e studio della distribuzione dei valori dei punti
- eventuali trasformazioni "normalizzanti"
- metodi di verifica dell'ipotesi di stazionarietà
- variogrammi multipli in caso di anisotropia
- analisi della carta degli errori e valutazione della qualità delle stime
- altri tipi di kriging: lognormal, indicator, probabilità, disjuntive, co-kriging, …
- …

si tratta solo di un assaggio, giusto per cominciare ad addentrarsi in una materia che richiede numerosi approfondimenti

… che non vi venga in mente di mettervi a fare kriging senza aver approfondito l'argomento: il rischio di ottenere risultati "balordi" è notevole, soprattutto con i software che offrono una versione di default dell'algoritmo
ve la sentireste di fare un kriging di pericolosità sismica alla luce di quanto è successo a Finale Emilia la settimana scorsa?