

Metodologie di analisi statistiche e geostatistiche per il supporto alla spazializzazione dei dati di stato sui Nitrati della Rete di Monitoraggio delle acque in Piemonte

Riccardo BALSOTTI*, Stefano BURATTO*, Teo FERRERO*, Tommaso NICCOLI**, Elio SESIA*, Alessandra TERRANDO*

**Area delle attività regionali per il coordinamento in materia ambientale*

***Area delle attività regionali per il rischio industriale e lo sviluppo economico compatibile*

Arpa Piemonte via della Rocca, 49 10123 Torino t.niccoli@arpa.piemonte.it <http://www.arpa.piemonte.it/>

Riassunto

Nell'ambito della revisione delle Zone Vulnerabili da Nitrati da parte della Regione Piemonte, Arpa Piemonte ha eseguito la spazializzazione dei dati di stato sulla contaminazione derivanti dalla Rete di monitoraggio delle acque sotterranee. Per affinare il risultato ottenuto in un precedente analogo lavoro è stato strutturato un processo di bonifica dei dati della rete basato su analisi statistiche, statistica spaziale e geostatistica propedeutiche ad una migliore definizione delle tecniche di spazializzazione dei dati.

Introduzione

Viene presentato il lavoro svolto dalle due strutture di ARPA in merito alla bonifica dei dati della Rete di monitoraggio delle acque sotterranee della Regione Piemonte gestita dall'Agenzia. Ad una prima analisi statistica dei dati, con conseguente bonifica di quelli ritenuti anomali, ha fatto seguito una analisi statistica spaziale e geostatistica per affinare le metodologie di rappresentazione dell'andamento della concentrazione del valore dei nitrati nella falda tramite tecniche di spazializzazione. Per le analisi e la bonifica dei dati sono stati considerati i dati dei nitrati, calcio e solfati acquisiti dai monitoraggi regionali per quanto riguarda la falda superficiale.

Per valutare la rappresentatività di ciascun pozzo sono stati esaminati due indicatori, la variabilità relativa % e il valore massimo dei nitrati per ciascun pozzo.

Si è poi effettuata una analisi statistica spaziale per conoscere le caratteristiche di distribuzione spaziale della popolazione, l'esistenza di anomalie nei dati risultato di un confronto con i pozzi limitrofi, l'appartenenza di un certo pozzo all'acquifero indagato, la presenza di un trend dei nitrati lungo la direzione di scorrimento della falda, l'autocorrelazione spaziale dei dati di calcio e solfati per evidenziare la presenza di *outlier*. Il tutto tramite *Parallel plot* sulle concentrazioni di solfati e calcio, regressioni di tipo lineare e analisi *LISA* e *I di Moran*. Questo ha consentito di scegliere tra tre tipologie di tecniche per la spazializzazione; il kriging Ordinario, il kriging Universale e l'IDW.

I dati spazializzati sono poi stati validati tramite cross-validazione, analisi dei valori predetti rispetto a quelli misurati effettuata in un raggio di circa 10 metri dal pozzo e analisi dei parametri di statistica descrittiva dei dati spazializzati messi a confronto con gli stessi parametri dei dati campionari.

Vengono presentati i risultati statistici, geostatistici e le spazializzazioni per due aree caratteristiche tra tutte le Aree Idrogeologicamente Separate nelle quali è suddivisa la pianura piemontese.

Costruzione della base dati alfanumerica

In merito alle concentrazioni di nitrati nella falda superficiale, determinate nell'ambito dei monitoraggi regionali del triennio 2001-2003, sono considerabili statisticamente rappresentativi i punti di monitoraggio con almeno quattro misure nell'arco del triennio di riferimento, per un totale di 449 pozzi.

Considerando le risultanze analitiche, ciascun punto di monitoraggio è stato classificato in funzione di due indicatori: grado di dispersione relativo (RSD) e valore massimo delle concentrazioni di

nitrati. Il primo indicatore evidenzia i pozzi con misure molto disperse intorno alla media e quindi ne esprime l'attendibilità, il secondo esprime invece la pericolosità del pozzo.

I punti di monitoraggio con bassa variabilità (RSD < 30%) sono considerabili statisticamente rappresentativi, mentre quelli con elevata variabilità (RSD ≥ 30%) necessitano di una preventiva analisi delle anomalie. Due i possibili esiti:

1. Le concentrazioni oscillano con ampiezza anomala e non stagionale: il punto di monitoraggio viene escluso perché la concentrazione media non è statisticamente rappresentativa.
2. Sono presenti alcune concentrazioni anomale, ma in numero non superiore a due: il pozzo è nel complesso accettabile e la concentrazione media viene calcolata escludendo le anomalie.

	RSD < 30%	RSD ≥ 30%
MAX < 25 µg/L	302 pozzi	69 pozzi
MAX ≥ 25 µg/L		78 pozzi

Tabella 1 – numero dei pozzi ripartiti nei due indicatori

L'analisi delle anomalie ha portato ad escludere 9 punti di monitoraggio, tutti contraddistinti da grande variabilità (RSD ≥ 30%) ed elevate concentrazioni (concentrazione MAX ≥ 25 µg/L). In conclusione la banca dati alfanumerica utilizzata per le attività successive è costituita da 440 punti di monitoraggio

Prime elaborazioni: studio dei parametri statistici e ricerca delle anomalie puntuali

Per ciascun pozzo è stata calcolata la media della concentrazione di nitrati, che costituisce la variabile fondamentale utilizzata per la rappresentazione statistico-spaziale (concentrazione puntuale dei nitrati). Successivamente si è studiata la distribuzione delle concentrazioni puntuali all'interno delle 37 singole aree idrogeologicamente separate (o brevemente "aree idrogeologiche"), in cui è stata suddivisa la pianura piemontese.

A tale scopo sono stati utilizzati i parametri statistici di base e sono state individuate e valutate le concentrazioni puntuali anomali, ovviamente laddove il numero di punti di monitoraggio è sufficientemente elevato.

Come esemplificazione, per le aree idrogeologiche codificate come CN02, CN03, VC02, vengono riportati alcuni diagrammi *Stem-and-Leaf*, utilizzati nella ricerca dei valori estremi (*outlier*).

Area Idrogeologica CN02

Minimo: 4.833
 Mediana: 27.000
 Massimo: 61.667

0 4
 0
 1
 1 H 777889
 2 44
 2 M 56779
 3 H 013344
 3
 4 0
 4
 5 1
 * * * *Outside Values* * * *

1
Outlier il valore 61.667 µg/L

Area Idrogeologica CN03

Minimo: 0.000
 Mediana: 32.883
 Massimo: 147.667

0 01779
 1 4778
 2 H 1122345556778
 3 M 05599
 4 11288
 5 H 00002469
 6 3
 7 3
 8 0
 9 01
 * * * *Outside Values* * * *

14 7
Outlier il valore 147.7 µg/L

Area Idrogeologica VC02

Minimo: 1.725
 Mediana: 7.103
 Massimo: 50.250

0 H 12223344
 0 M 556677779
 1 4
 1 H 5
 2 4
 2
 3 12
 * * * *Outside Values* * * *

3 57
 5 0
Outlier i valori 35, 35 e 50 µg/L

Analisi statistica spaziale per i dati sui nitrati, solfati e calcio nelle acque

Per i pozzi ricadenti in ogni singola area idrogeologica, si è effettuata un'analisi statistica spaziale, per verificare se esistano anomalie nei dati non rilevate dalla precedente analisi statistica classica. Le anomalie in questo caso risultano da un confronto dei parametri di un certo pozzo con i valori dei pozzi limitrofi, per rilevare forti incongruenze. Soprattutto si è puntato a verificare l'appartenenza di un certo pozzo all'acquifero indagato per escludere situazioni locali da non introdurre poi nella successiva spazializzazione. Si sono pertanto indagati, tramite *Parallel plot*, le concentrazioni di solfati e calcio dei singoli pozzi. Questo permette di individuare pozzi spazialmente vicini e correlarli al valore di concentrazione, per evidenziare dove vi siano casi di valori molto difforni tra loro.

Tramite analisi *LISA* e *I di Moran* si è verificata l'autocorrelazione spaziale dei dati di calcio e solfati. Con questa analisi si definisce un certo raggio di ricerca entro il quale viene valutato ogni singolo pozzo in confronto ai parametri medi della zona individuata e si producono diagrammi e cartografie, con i quali si evidenziano eventuali *outlier* rispetto al contesto e aggregazione di pozzi fortemente correlati tra loro. Si è inoltre analizzata la presenza o meno di un *trend* ben evidente tra la direzione di scorrimento della falda e l'andamento della concentrazione dei nitrati, tramite regressione di tipo lineare su una o due variabili indipendenti che indicano la direzione principale della falda (le coordinate x e y dei pozzi).

Spazializzazione dei dati

I dati sui nitrati, derivanti dai pozzi per ciascuna delle aree di riferimento e bonificati da eventuali anomalie, sono poi stati spazializzati per definire sul territorio la suddivisione di dettaglio delle aree che superano la soglia di contaminazione.

Le metodologie adottate dalla spazializzazione sono riconducibili a tre tipologie:

1. Kriging ordinario (OK): ha lo scopo di verificare se sussistono le condizioni per la definizione di un modello di distribuzione nello spazio attraverso analisi variografica, e, nel caso, la definizione dei pesi da applicare nella spazializzazione tramite l'inferenza del modello teorico al variogramma sperimentale. Per ogni spazializzazione effettuata è stato scelto il vicinaggio che minimizza la media standardizzata dell'errore di stima per tutta l'area.
2. regressione lineare per modellare il trend dei dati e krigaggio ordinario sui valori dei residui (Kriging Universale - UK): procedimento sostanzialmente analogo al precedente, ma opera in quei casi in cui, tramite l'analisi statistica spaziale precedente si è evidenziato un *trend* strutturato nei dati in una precisa direzione dello spazio. Questo fa sì che non sia possibile definire un variogramma per questi dati in quanto l'andamento della varianza non tende mai a divenire asintotico, impedendo di fatto la definizione di un *sill* e di un *range* di correlazione. L'eliminazione del *trend*, modellato dalla retta di regressione, consente di poter effettuare l'analisi variografica sui residui dei dati e l'applicazione delle tecniche di krigaggio per la spazializzazione di questi ultimi. Una volta effettuata la spazializzazione tramite OK si reintroduce i valori del *trend* sommandoli a quelli dei residui.

I risultati dell'applicazione delle due tecniche di krigaggio sono poi sottoposti ad operazioni di validazione attraverso:

- a) analisi dei dati prodotti dalla cross-validazione che consente di valutare per quali aree i modelli sono sicuramente applicabili (errori e deviazioni standard contenuti), non applicabili (errori o deviazioni standard troppo elevati), applicabili ma con evidente presenza di anomalie isolate.
- b) analisi dei valori predetti rispetto a quelli misurati effettuata in un raggio di circa 10 metri dal pozzo, considerando che in questa distanza si voglia riscontrare un valore di predizione del modello spazializzante ancora molto simile a quello derivante dall'analisi dei campioni.
- c) analisi dei parametri di statistica descrittiva dei dati spazializzati messi a confronto con gli stessi parametri dei dati campionari, per individuare se, con l'operazione di inferenza che si

compie con la spazializzazione, non si sia andati a variare di molto le caratteristiche della popolazione originaria.

3. spazializzazione tramite IDW che è stata applicata in quei casi in cui non vi erano sufficienti punti o dove, dall'analisi variografica, non si riusciva a strutturare un variogramma ritenuto accettabile. Per questo tecnica l'unica validazione proposta è l'analisi della media dell'errore e delle deviazioni standard ottenuta dalla cross validazione.

Si presentano di seguito i risultati più significativi per due aree

Area di riferimento CN03

Analisi statistica spaziale

Una prima analisi dell'area con il metodo *LISA* per valutare l'autocorrelazione spaziale dei dati ha permesso di identificare due pozzi con valori fortemente estremi rispetto al totale per quanto riguarda i solfati presenti. Gli stessi sono anche ben aggregati per gli altri pozzi. L'eliminazione dei due pozzi in questione aumenta il coefficiente di correlazione da 0.3 a 0.8.

Si è poi evidenziato come si possa tranquillamente suddividere l'area in questione in almeno due sotto-aree distinte. Un'area che è stata definita CN03SW che ha comunque una buona correlazione dei solfati sui pozzi di appartenenza. La rimanente parte dei pozzi è confluita in un'area che è stata denominata CN03TAN, che ha un elevato coefficiente di correlazione per i solfati.

Il diagramma in figura 2 mostra come per la sottoarea CN03SW vi sia una forte omogeneità dell'andamento dei solfati con l'andamento della quota piezometrica (linee scure) e come invece vi sia una minore omogeneità per l'area CN03TAN, con le due linee chiare all'estremo destro del diagramma che rappresentano i due pozzi evidenziati come anomali e che probabilmente fanno storia a se.

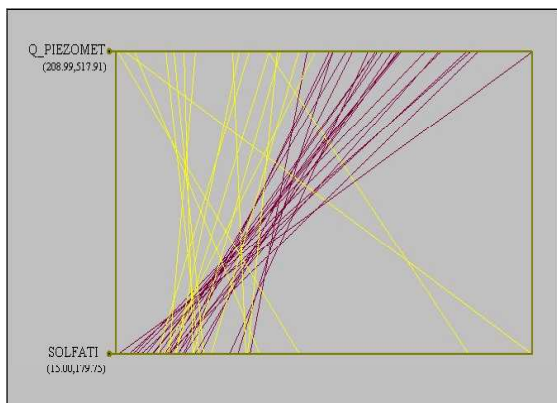
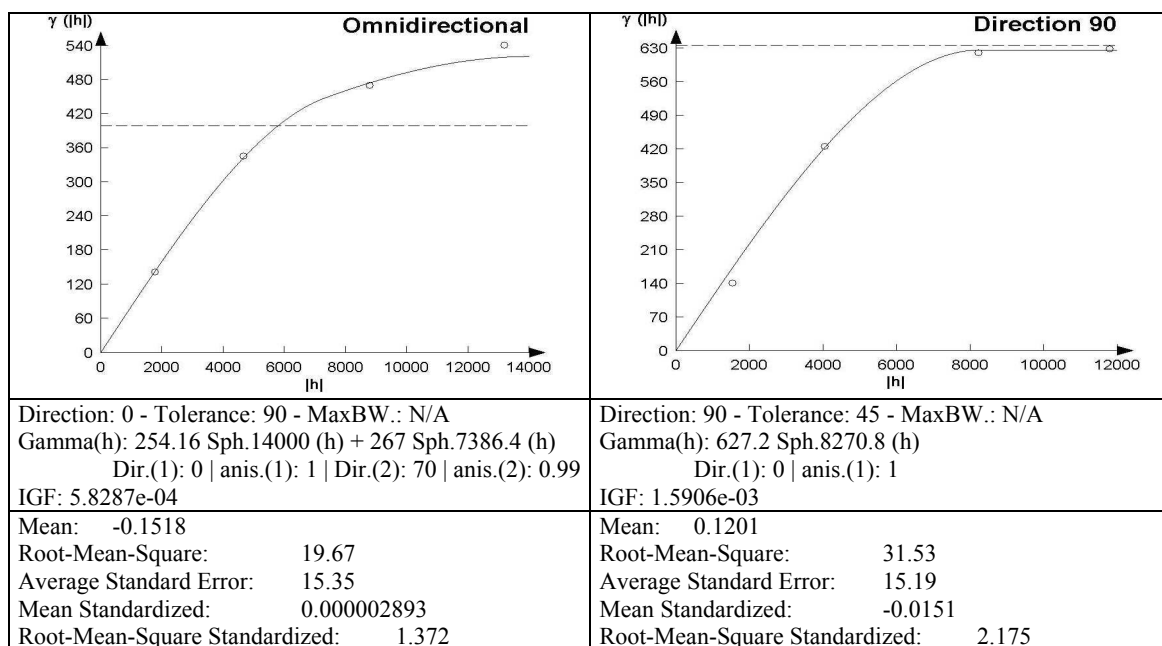


Figura 2 – parallel plot solfati/quota piezometrica

Spazializzazione

Si sono suddivisi i pozzi nelle due sottoaree evidenziate e utilizzato l'OK per la spazializzazione; di queste si riportano i variogrammi con i parametri del modello e della *cross-validation*.



Analisi statistica descrittiva dei dati spazializzati

Parametri	Kriging	Popolazione campionaria
Media	30.790	32.170
Varianza	177.850	416.810
dev. Standard	13.340	20.420
asimmetria	1.640	1.080
curtosi	2.860	0.520
minimo	3.860	3.500
massimo	84.800	85.050
campo di variazione	80.940	81.550
coeff. di variazione	0.430	0.630

Tabella 3 – Parametri statistici per l'area CN03SW

Parametri	Kriging	Popolazione campionaria
Media	46.600	49.790
Varianza	252.570	677.980
dev. Standard	15.890	26.040
asimmetria	0.150	0.420
curtosi	-0.120	1.030
minimo	2.000	1.500
massimo	113.660	114.880
campo di variazione	111.660	113.380
coeff. di variazione	0.340	0.520

Tabella 4 – Parametri statistici per l'area CN03TAN

Dall'analisi dei risultati di cross-validazione e dal confronto tra i parametri statistici, la spazializzazione è ritenuta accettabile separando in due l'area in esame, con buoni risultati per la sottoarea CN03SW e maggiori problemi per la sottoarea CN03TAN con due pozzi che apparentemente sembrano descrivere situazioni locali particolari ma che non vanno comunque ad influenzare i risultati di spazializzazione.

Area di riferimento VC02 VC03

Analisi statistica spaziale

Una prima analisi spaziale è stata l'analisi tramite la *I di Moran* per stimare l'autocorrelazione spaziale dei dati. Risulta una correlazione alta – alta nella zona Ovest dell'area di riferimento, con correlazione positiva tra valori alti di concentrazione dei nitrati e una Sud-Est a correlazione sempre positiva di tipo bassa – bassa. Si rafforza la tesi di due sottopopolazioni di dati presenti nell'area di riferimento emersa nel precedente lavoro, anche se questo non implica una separazione fisica dei dati ma, probabilmente, solo una distribuzione fortemente localizzata dell'influenza dei carichi di azoto sulla falda. La stessa analisi sui valori di calcio e solfati d'altronde non porta a risultati che facciano pensare ad una separazione fisica delle due aree. Si rileva invece un pozzo che rappresenta un *outlier* rispetto al contesto geografico in cui è insediato di tipo alto –basso, cioè un valore elevato di nitrati in un contesto di valori decisamente più bassi nel vicinaggio di ricerca.

Una seconda analisi ha cercato di individuare un trend lineare nella distribuzione dei dati per affinare la successiva spazializzazione dei dati. Si riportano di fianco i principali parametri della funzione di regressione lineare individuata sui dati di concentrazione di nitrati rispetto alla coordinata Est dei pozzi (direzione principale della falda).

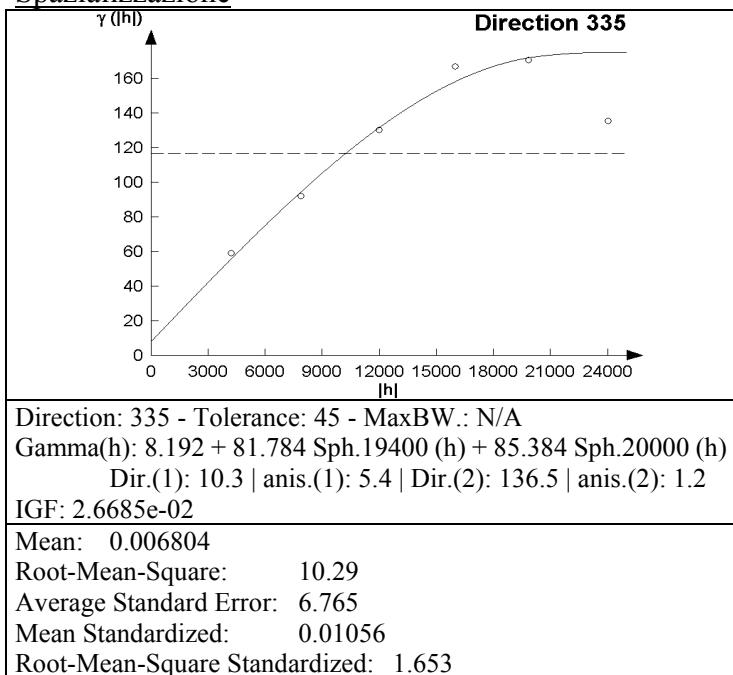
Variable	Coefficient	Std. Error	t-Statistic	Probability
CONSTANT	378.4696	63.52116	5.958165	0.0000005
X	-0.000816	0.000142	-5.71045	0.0000010

REGRESSION DIAGNOSTICS
 MULTICOLLINEARITY CONDITION NUMBER 76.22091
 (Extreme Multicollinearity)

Vista l'estrema rappresentatività dei dati ad opera della funzione lineare ($Prob(F\text{-statistic}) < 0.001$ e *Probability* decisamente bassa) si è associato alle coordinate del pozzo il valore dei residui calcolati dalla retta di regressione ed utilizzato questi valori per l'analisi variografica e la successiva spazializzazione dei dati.

Figura 5 – Parametri della regressione

Spazializzazione



La spazializzazione è il risultato della somma dei due livelli (UK), ottenuti il primo dalla funzione di regressione lineare dei dati rispetto alla coordinata Est dei pozzi che modella il *trend* e il secondo tramite krigaggio ordinario dei valori dei residui dei dati grezzi dalla retta di regressione utilizzata per il primo *layer*. Si riportano di fianco il variogramma con i parametri del modello e della *cross-validation*.

Analisi statistica descrittiva dei dati spazializzati

Parametri	Stima da Kriging	Stima da Regress.	Valori di campionam.
Media	17.82	18.34	15.86
Varianza	123.10	164.16	212.02
dev. Standard	11.09	12.81	14.56
asimmetria	0.67	0.62	0.78
curtosi	-0.89	-0.91	-0.70
minimo	1.24	0.56	0.00
massimo	43.68	50.05	48.40
campo di variazione	42.44	49.49	48.40
coeff. di variazione	0.62	0.70	0.92

Tabella 6 – Parametri statistici per l’area VC02 VC03

Valori predetti in confronto ai valori misurati in un raggio di 10 metri

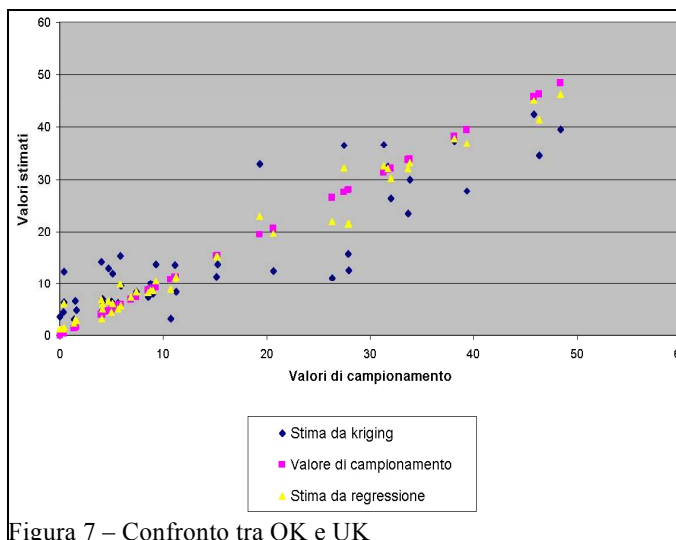


Figura 7 – Confronto tra OK e UK

Dall’analisi dei risultati di cross-validazione e dal confronto tra i parametri statistici, la spazializzazione offre buoni risultati. Per evidenziare l’efficacia di spazializzare tramite UK si riporta il confronto tra i valori statistici e un diagramma che rappresenta il valore stimato dal risultato dei due metodi a 10 metri dal pozzo campionato. Si noti come la stima fatta attraverso la regressione sia molto attinente ai dati campionati per valori di concentrazione bassi e tenda poi a differire con valori inferiori al misurato per concentrazioni più elevate. La stima tramite OK invece appare da subito molto meno lineare e con valori fortemente discosti da quelli misurati.

Riferimenti bibliografici

Anselin, L. (2003). *GeoDa 0.9 User’s Guide*, Spatial Analysis Laboratory (SAL). Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign, IL.