

Approccio funzionale nella zonizzazione del territorio secondo la valutazione della qualità dell'aria

Stefano Bande¹, Stefania Ghigo², Rosaria Ignaccolo²

¹ Arpa Piemonte, Area Previsione e Monitoraggio Ambientale,
Via Pio VII 9, Torino, s.bande@arpa.piemonte.it

² Università degli Studi di Torino, Dipartimento di Statistica e Matematica Applicata,
Corso Unione Sovietica 218 bis, Torino, ghigo@econ.unito.it, ignaccolo@econ.unito.it

RIASSUNTO

In questo lavoro viene proposta la functional cluster analysis come metodologia per la classificazione del territorio in zone caratterizzate da differenti livelli di criticità in relazione alle concentrazioni di inquinanti atmosferici, come richiesto dalla normativa comunitaria e nazionale in materia di qualità dell'aria. Vengono prese in considerazione le serie temporali relative ai principali inquinanti, ottenute dal sistema modellistico implementato dall'Area Previsione e Monitoraggio Ambiente dell'Arpa Piemonte e misurate dalle stazioni di monitoraggio del Sistema Regionale di Rilevamento della Qualità dell'aria (S.R.R.Q.A). I dati osservati di concentrazione vengono introdotti nei campi al suolo prodotti dal sistema modellistico mediante tecniche di kriging con deriva esterna. Considerando quindi le serie storiche così ottenute come dati funzionali si effettua, attraverso la functional cluster analysis, la zonizzazione dell'intera regione servendosi dell'algoritmo PAM: la rappresentazione dei cluster sul territorio permette di realizzare mappe per zone di criticità a livello regionale, in funzione della concentrazione di ciascun inquinante analizzato. L'utilizzo di tecniche di up-scaling consente inoltre di associare ogni comune piemontese ad una determinata zona. Infine si presentano due soluzioni di sintesi delle serie storiche che permettono di considerare contemporaneamente tutti gli inquinanti analizzati per identificare una zonizzazione univoca del territorio.

INTRODUZIONE E DATI UTILIZZATI

La Direttiva europea Dir. 96/62/CE richiede che il territorio venga suddiviso, sulla base della valutazione della qualità dell'aria, in differenti zone ai fini di predisporre adeguate politiche di azione, risanamento o mantenimento. L'attuazione nazionale (D.Lgs. 351/99) di tale direttiva delega alle regioni italiane la responsabilità di identificare le aree della propria giurisdizione nelle quali i livelli di concentrazione di uno o più inquinanti comportano il rischio di superamento dei valori limite e delle soglie di allarme fissate dalla legge e l'individuazione dell'autorità competente alla gestione di tali situazioni di rischio. All'interno di tale quadro normativo lo scopo del lavoro presentato è quello di definire una metodologia che possa supportare gli enti istituzionali nella classificazione del territorio piemontese, mediante l'utilizzo di tecniche di clustering funzionale per le serie temporali relative agli inquinanti atmosferici NO₂, PM₁₀, CO ed SO₂.

I dati utilizzati fanno riferimento all'anno 2005 e provengono sia dalle stazioni di monitoraggio del Sistema Regionale di Rilevamento della Qualità dell'Aria della Regione Piemonte (S.R.R.Q.A), sia dal sistema modellistico euleriano di chimica e trasporto implementato dall'Area Previsione e Monitoraggio Ambientale dell'Arpa Piemonte. Tale sistema, utilizzato operativamente da Arpa a supporto delle attività di competenza della Regione Piemonte finalizzate alla valutazione (annuale) dello stato della qualità dell'aria, fornisce in output i campi di concentrazione dei principali inquinanti atmosferici su base oraria su un dominio di calcolo tridimensionale che comprende l'intera regione Piemonte con una risoluzione orizzontale di quattro chilometri. I risultati del sistema modellistico (Bande *et al.*, 2007) per il biossido di azoto sono soddisfacenti, anche se si individuano alcune sottostime dei picchi di concentrazione in caso di episodi acuti dovuti a condizioni meteorologiche avverse. Si osserva invece una sottostima dei livelli di concentrazione di PM₁₀ nel semestre invernale, specialmente per le stazioni ubicate al di fuori dall'area metropolitana di Torino: ciò è parzialmente dovuto ad una sottostima nelle emissioni fuori Torino e alla difficoltà di riprodurre su larga scala i fenomeni di accumulo di particolato atmosferico nel bacino padano. Per quanto riguarda le concentrazioni di CO e SO₂ il sistema modellistico tende a sovrastimare, tuttavia sia i dati osservati sia quelli simulati sono di molto al di sotto dei valori limite imposti dalle direttive europee.

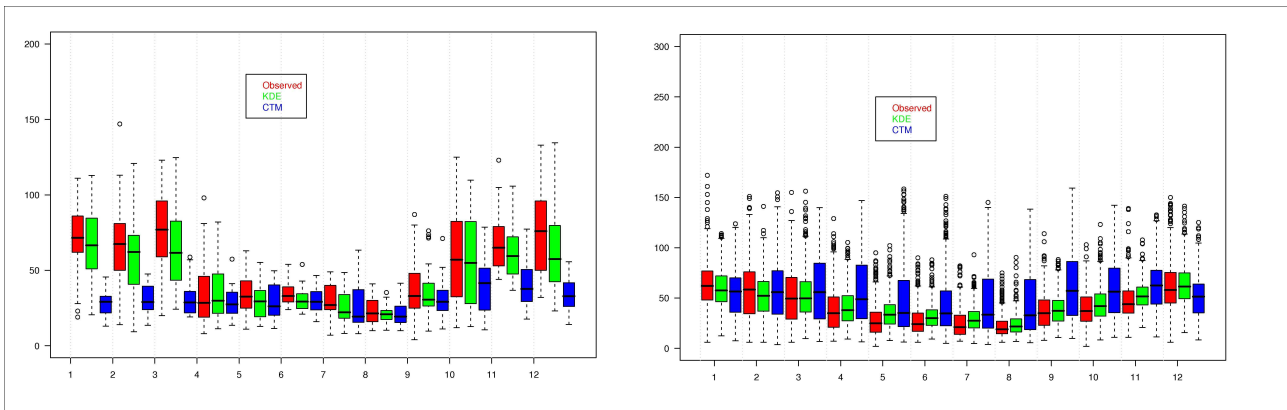
Pertanto, prima di effettuare la classificazione, si effettua una "correzione" dei campi di concentrazione mediante *kriging con deriva esterna (KED)* dei dati osservati dalle stazioni del Sistema Regionale di Rilevamento della Qualità dell'Aria della Regione Piemonte (S.R.R.Q.A). IL kriging con deriva esterna è un caso particolare del *kriging* universale (Wachernagel, 2003) e si differenzia da esso nelle componenti del *trend*. Infatti, nel kriging universale il *trend* è funzione delle coordinate della localizzazione spaziale in due direzioni ortogonali, mentre nel *KED* la deriva è definita esternamente, nel caso in esame in funzione dei campi di concentrazione del modello euleriano. Ciò equivale ad imporre che lo stimatore di *kriging* rispetti non solo l'andamento del *trend*, ma anche la variazione spaziale della deriva esterna. L'implementazione scelta del *KED* segue l'approccio di Van de Kassteele (Van de Kassteele *et.al.*, 2008). Per garantire la stazionarietà e la normalità, viene preliminarmente applicata una trasformazione *Box-Cox* ai dati originali. Per ogni istante temporale, le osservazioni trasformate sono interpretate come realizzazioni di una variabile gaussiana $Z(s)$ ad una localizzazione spaziale s . Questo processo spaziale è caratterizzato da una struttura del tipo:

$$Z(s) = \mu(s) + D(s) + \varepsilon$$

dove $\mu(s) = \xi_0 + X(s)\xi_1$ è la componente spaziale deterministica (*trend* o *deriva*), con X variabile deterministica che rappresenta il risultato della catena modellistica, e $\xi = \{\xi_0, \xi_1\}$ è il vettore dei parametri non noti. $D(s)$ è un processo casuale stazionario gaussiano, con media zero e *sill* σ^2 , che tiene conto della correlazione spaziale tra le osservazioni mediante la funzione di correlazione spaziale $\rho(\cdot)$ con *range* ϕ . Infine, ε è il termine d'errore caratterizzato da una varianza τ^2 (*nugget*). Per tutti gli inquinanti considerati viene utilizzata una covarianza spaziale di tipo esponenziale; i parametri della covarianza e della trasformazione *Box-Cox* sono stimati ad ogni istante temporale con il metodo della massima verosimiglianza.

I risultati del *KED* sono stati validati con metodi di *cross-validation* e confrontati con i risultati ottenuti direttamente dalla simulazione deterministica. A titolo di esempio, la fig.1 mostra i *boxplot* per mese delle distribuzioni di NO_2 e PM_{10} , osservate e stimate in due stazioni della rete di monitoraggio: si nota come, soprattutto per il PM_{10} nei mesi invernali, l'utilizzo del *KED* migliori notevolmente le stime di concentrazione.

Figura 1 – Confronto tra output del *KED* e output diretto della simulazione modellistica



Boxplot per mese delle distribuzioni di dati osservati (in rosso) e simulati (*KED* in verde e output diretto della simulazione modellistica in blu) per NO_2 nella stazione di Orbassano e PM_{10} nella stazione di Borgaro

METODOLOGIA

FUNCTIONAL CLUSTER ANALYSIS

Le serie storiche degli inquinanti possono essere considerate come realizzazioni di processi continui registrati in tempi discreti, cioè le n coppie (t_j, y_j) di valori y_j osservati ai tempi t_j sono istantanee di una curva continua, il dato funzionale. La conversione da dati discreti a curve avviene tramite tecniche di *smoothing*, impiegando combinazioni lineari di funzioni *B-spline* (Ramsay and Silverman, 2005).

Al primo stadio della procedura di *functional cluster analysis* (*FCA*), ogni curva viene sintetizzata attraverso un vettore di coefficienti in R^{K+d+1} , dove K è il numero di nodi e d il grado delle *B-spline*. In questo lavoro il grado è scelto pari a 3, poiché una *spline* cubica è abbastanza flessibile e si adatta bene alle serie storiche degli inquinanti, e i nodi vengono posizionati a distanza fissa, allo scadere di ogni mese ($K=11$). In tal modo si riesce a cogliere il *pattern* temporale predominante, evitando allo stesso tempo l'influenza di variazioni eccessivamente locali (settimanali o giornaliere) sulle curve stimate.

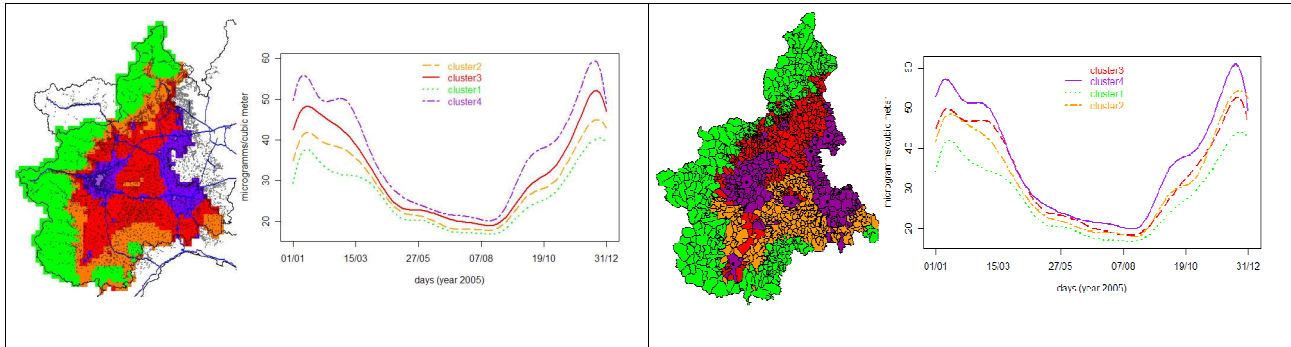
Ottenuti i coefficienti delle spline per ogni curva, diversamente dalla *FCA* classica (Abraham et al., 2003), qui viene utilizzato l'algoritmo non gerarchico *PAM* (*Partitioning Around Medoid*) per classificare gli oggetti (Kaufman and Rousseeuw, 1990), come già proposto in Ignaccolo et al. (2008). *PAM* identifica oggetti rappresentativi detti "medoidi" per ogni cluster, che nel caso funzionale sono curve in cui il pattern temporale può essere osservato come rappresentante l'evoluzione temporale dell'inquinante nell'area corrispondente (vedi fig. 2).

UP-SCALING

La classificazione del territorio piemontese su scala comunale viene effettuata applicando la *FCA* sulle serie storiche su ogni unità amministrativa (il comune), costruite mediante tecniche di *up-scaling* a partire dai campi prodotti da *KED*. Tali tecniche ricercano il valore equivalente ad una variabile su larga scala conoscendone il valore in un certo numero di punti ad una scala più piccola. Mediante il software G.I.S. G.R.A.S.S., il campo di concentrazione al suolo di un dato inquinante prodotto da *KED* viene sovrapposto

alle cartografie vettoriali tematiche che descrivono il territorio regionale. Basandosi sull'area di intersezione tra il poligono che rappresenta i confini comunali e le celle del grigliato di calcolo, viene valutata la superficie e l'area edificata di ogni comune che ricadono all'interno di ogni cella. L'aggregazione dei dati a livello comunale, per ogni istante e per ogni inquinante, viene quindi effettuata basandosi su tre differenti algoritmi: la media pesata delle concentrazioni che ricadono all'interno del territorio comunale con i pesi dati dalla percentuale di superficie comunale all'interno di una data cella, la media pesata con i pesi rappresentati dalla porzione di area edificata comunale all'interno di una data cella e, in un'ottica più cautelativa, il 90-esimo percentile della serie spaziale dei punti griglia che ricadono all'interno di ogni territorio comunale.

Figura 2 – Risultati della *Functional cluster analysis* prima e dopo *up-scaling*



Mappe di rischio e mediodi per l'NO₂ nella classificazione effettuata a partire dai campi di concentrazione (destra) e dalle serie comunali ottenute con tecniche di *up-scaling* (sinistra).

ZONIZZAZIONE MULTI-INQUINANTE

Per ottenere una zonizzazione del territorio univoca, che tenga conto dei diversi inquinanti, si propongono due differenti soluzioni di sintesi delle serie storiche: l'indice di qualità dell'aria BC introdotto da Bruno e Cocchi (2002), e la *functional principal component analysis (FPCA)*. La suddivisione in zone viene successivamente effettuata sulla base di queste sintesi.

La proposta di Bruno e Cocchi si basa sulla costruzione di una famiglia di indici che aggregano dati nel tempo, nello spazio e rispetto agli inquinanti. L'indice finale si ottiene con successive aggregazioni: il primo passo consiste nell'aggregare temporalmente i campi orari di concentrazione per ottenere un valore giornaliero, mediante funzioni di aggregazione indicate dalla normativa vigente D.M. N° 60/2002. Successivamente si effettua la standardizzazione del valore ottenuto dividendo ciascun dato su base giornaliera per il relativo valore limite fissato dalla normativa di riferimento. Ottenuto il dato relativizzato si procede all'aggregazione rispetto all'inquinante. Si sceglie di utilizzare il massimo come funzione di aggregazione, in modo da considerare la situazione più critica, caratterizzata da alti valori di inquinamento. Ottenute le serie storiche dell'indice BC, si procede alla classificazione funzionale dei siti.

In alternativa, le serie storiche delle concentrazioni su griglia possono essere sintetizzate attraverso l'analisi delle componenti principali (PCA) funzionali. Una PCA funzionale e multivariata può offrire una sintesi di informazioni tra variabili differenti evidenziando le principali componenti di variazione legate alla interazione tra di esse (Ramsay and Silverman, 2005). Nel contesto funzionale, gli *score* delle componenti principali, ossia le nuove variabili non correlate tra loro, sono specificati attraverso una combinazione lineare dei dati funzionali (le curve lisce per ogni sito e ogni inquinante). I pesi nella suddetta combinazione lineare sono autofunzioni $\alpha_p(t)$, definite sullo stesso intervallo temporale dei dati funzionali, mutuamente ortogonali ottenute in modo da massimizzare l'informazione (in termini di variabilità) contenuta nelle componenti principali. I siti vengono poi raggruppati sulla base degli *score* delle componenti principali usando nuovamente l'algoritmo PAM.

RISULTATI

In questo lavoro si propone la classificazione del territorio piemontese adottando tecniche di *clustering* funzionale sulla base dei campi di concentrazione di NO₂, PM₁₀, CO e SO₂ forniti dalla catena modellistica euleriana implementata dall'Area Previsione e Monitoraggio Ambientale dell'ARPA Piemonte e dei dati di concentrazioni misurati dal S.R.R.Q.A, introdotti nei campi di concentrazione con tecniche di *kriging* con deriva esterna. Sono stati presentati approcci diversi alla zonizzazione.

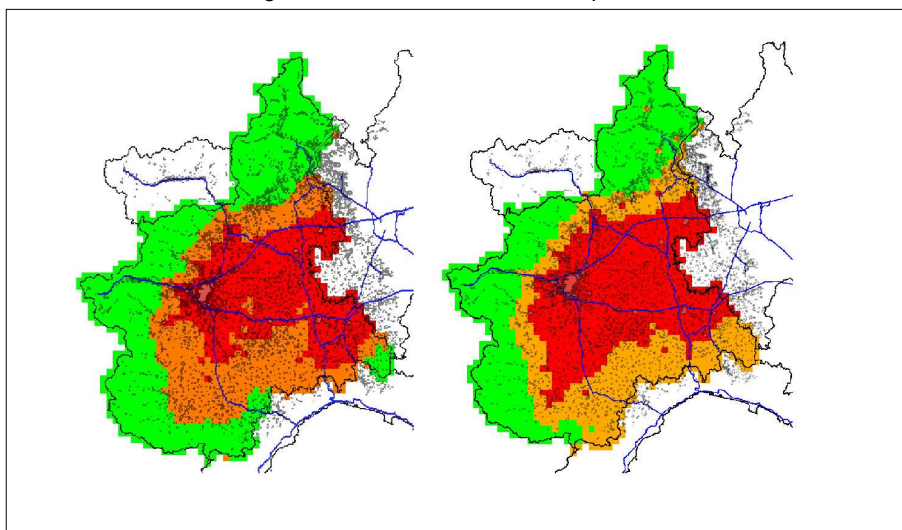
La classificazione effettuata a partire dai campi di concentrazione con una risoluzione di 4 km presenta un quadro dettagliato della situazione per ciascun inquinante in analisi. Dalla mappa per l'NO₂ in fig. 2 (sinistra) è ben visibile come i gruppi si "allungano" in funzione della rete stradale e dei principali agglomerati piemontesi. In particolare il cluster 4 vede al suo interno Torino, Alessandria, Novara e le rispettive periferie, ossia le zone più urbanizzate e antropizzate della regione e per questo le più critiche. Il gruppo 3 comprende

l'intorno dei percorsi autostradali e delle vie di comunicazione più trafficate, mentre il cluster 2 racchiude la restante pianura e le zone pedemontane e il numero 1 è caratterizzato dal territorio montuoso della regione. I medoidi sottolineano la grande variazione che, soprattutto nei mesi invernali, esiste tra punti griglia appartenenti a gruppi diversi, che raggiunge anche 30 microgrammi/m³.

Qualora si debbano fornire informazioni sulla criticità della qualità dell'aria nel territorio ai decisori, agli epidemiologi e alla popolazione piemontese, è invece opportuno fare riferimento ad entità amministrative, quali i comuni. Le zonizzazioni ottenute attraverso i tre metodi di *up-scaling* presentano per ogni inquinante strutture molto simili, anche se quella basata sul calcolo del 90-esimo percentile delle serie spaziali risulta leggermente più cautelativa. Per ragioni di spazio, viene rappresentata in fig. 2 (destra) solo la mappa relativa al biossido di azoto.

Per avere un quadro complessivo del territorio in relazione allo stato della qualità dell'aria è possibile considerare la zonizzazione multi-inquinante. Sono stati proposti due metodi di sintesi delle serie storiche, attraverso l'uso dell'indice BC di qualità dell'aria e della FPCA: la classificazione unica si è ottenuta sulla base di queste sintesi. Si riscontra che il contributo del particolato atmosferico sulla zonizzazione è predominante rispetto a quello degli altri inquinanti. Entrambi i risultati ottenuti si possono definire precauzionali, anche se la zonizzazione multi-inquinante ricavata tramite FPCA risulta la più cautelativa in assoluto, anche rispetto a quella basata sul calcolo del 90-esimo percentile delle serie spaziali, come mostrano le mappe in fig. 3.

Figura 3 – Zonizzazione multi-inquinante



Mappe di rischio per la zonizzazione multi-inquinante ottenute utilizzando l'indice BC (sinistra) e la FPCA (destra) come metodi di sintesi delle serie storiche.

Bibliografia

- Abraham C.; Cornillon P. A.; Matzner-Løber E. e Molinari N., *Unsupervised curve clustering using B-splines*, 2003, Scandinavian Journal of Statistics, 30: 581-595.
- Bande S.; Clemente M.; De Maria R.; Muraro M.; Piccolo M.; Arduino G.; Calori G.; Fi nardi S.; Radice P.; Silibello C.; Brusasca G., *The modelling system supporting piemonte region yearly air quality assessment*, 2007, Proceedings of the 6th International Conference on Urban Air Quality.
- Bruno F., Cocchi D., *A unified strategy for building simple air quality indexes*, 2002, Environmetrics, 13: 243-261.
- Ignaccolo R., Ghigo S., Giovenali E., *Analysis of air quality monitoring networks by functional clustering*, 2008, Environmetrics, 19: 672 – 686.
- Kaufman L., Rousseeuw P.J., *Finding Groups in Data. An introduction to cluster analysis*, 1990, Wiley.
- Ramsay J.O., Silverman B., *Functional Data Analysis*, 2005, Springer Verlag.
- H. Wackernagel, *Multivariate geostatistics: an introduction with applications*, 2003, Springer.
- Van de Kasstele J.; Stein A.; Dekkers A.; Velders G. *External drift kriging of NOx concentrations with dispersion model output in a reduced air quality monitoring network*, 2008, Environmental and Ecological Statistics, DOI 10.1007/s10651-007-0052-x.