

Web Working Papers
by

The Italian Group of Environmental Statistics



Gruppo di Ricerca per le Applicazioni della Statistica
ai Problemi Ambientali

www.graspa.org

Functional Cluster Analysis per l'ottimizzazione della rete di
monitoraggio della qualità dell'aria in Piemonte

Stefania Ghigo, Elisabetta Giovenali e Rosaria Ignaccolo

GRASPA Working paper n.26, Dicembre 2006

Functional Cluster Analysis per l'ottimizzazione della rete di monitoraggio della qualità dell'aria in Piemonte*

S. Ghigo, E. Giovenali e R. Ignaccolo[†]

18 dicembre 2006

Dipartimento di Statistica e Matematica Applicata “Diego de Castro”
Università degli Studi di Torino
e-mail: {ghigo, giovenali, ignaccolo}@econ.unito.it

Abstract: *Nel presente lavoro viene analizzata la rete piemontese di monitoraggio della qualità dell'aria a partire da informazioni caratterizzanti le centraline e da dati di concentrazione di tre tra gli inquinanti atmosferici a maggiore criticità (PM_{10} , NO_2 e O_3). Le serie storiche delle concentrazioni possono essere viste come dati funzionali e analizzate tramite metodi di functional data analysis. Attraverso la classificazione della rete si identificano livelli e evoluzioni temporali simili degli inquinanti nelle diverse stazioni di misura; in questo modo si possono determinare località rappresentative e, al tempo stesso, pattern spaziali per tali inquinanti in una data area.*

Keywords: *B-spline, Clustering, Dati Funzionali, Inquinanti Atmosferici.*

1 Introduzione

Il monitoraggio della qualità dell'aria rappresenta lo strumento principale di cui si avvalgono gli enti territoriali per una corretta gestione e valutazione dello stato di qualità dell'aria in ottemperanza ai propri compiti istituzionali. La rete di monitoraggio della qualità dell'aria di una regione è generalmente costituita da centraline di rilevamento di inquinanti atmosferici e di variabili meteorologiche distribuite nel territorio. La classificazione di tali stazioni di monitoraggio rappresenta quindi un metodo di valutazione della rete stessa, così da avere a disposizione un supporto adeguato per scelte miranti la sua ottimizzazione come, ad esempio, la collocazione o lo spostamento di strumenti di misura. Essa permette l'identificazione di centraline rappresentative relative ad ogni inquinante, poichè evidenzia la similarità tra livelli e/o evoluzioni temporali, nonchè la visualizzazione di eventuali *pattern* spaziali tramite la rappresentazione su mappa dei gruppi formati.

Per la rete di monitoraggio nella zona di Milano, per esempio, Lavecchia *et al.* [Lav-96] utilizzano una procedura di *clustering* gerarchico che permette di valutare la similarità tra le centraline misuranti le concentrazioni di ozono in termini di livelli e di variabilità temporale. Un approccio simile è utilizzato da Gabusi e Volta [Gab-05] per la classificazione delle centraline di monitoraggio

*Questo lavoro è stato parzialmente supportato dal progetto ministeriale MIUR PRIN-COFIN2004, e dai progetti CIPE2003 e CIPE2004 finanziati dalla Regione Piemonte.

[†]Indirizzo per corrispondenza: R. Ignaccolo, Dipartimento di Statistica e Matematica Applicata, piazza Arbarello, 8, 10122 Torino (TO), Italy

di tutto il Nord Italia, al fine di determinare le più rappresentative per successive analisi di livelli di concentrazione dell'inquinante studiato (l'ozono). Per definire i *pattern* spaziali dei massimi giornalieri di ozono associati al superamento di valori limite nella zona della Baia di Monterey e di San Francisco, Ludwig *et al.* [Lud-95] impiegano la *cluster analysis* congiuntamente alla *empirical orthogonal function (EOF) analysis*. Ionescu *et al.* [Ion-00] propongono invece una classificazione, tramite l'algoritmo di Ward, delle stime dei campi di concentrazione di NO₂ ottenuti attraverso l'uso di *thin plate spline function*, adoperando tutti gli episodi di picco rilevati nel periodo 1994-97 dalla rete di monitoraggio AIRPARIF nella regione parigina. La *cluster analysis* è inoltre utilizzata da Gramsh *et al.* [Gra-06] sui dati di particolato e ozono misurati dalla rete di monitoraggio di Santiago del Cile nell'anno 2000, al fine di determinare *trend* stagionali e distribuzioni spaziali.

Questo lavoro si propone come scopo l'analisi della rete di monitoraggio della qualità dell'aria in Piemonte a partire dalle informazioni caratterizzanti le centraline e dai dati di concentrazione dei tre inquinanti a maggiore criticità, le polveri sottili (PM₁₀), il biossido d'azoto (NO₂) e l'ozono (O₃), misurati dalle stazioni di rilevamento al suolo nell'anno 2004 e forniti dal Sistema Regionale di Rilevamento della Qualità dell'Aria (SRQA) della Regione Piemonte [2], al fine di identificare centraline rappresentative dei livelli e delle evoluzioni temporali. Le serie storiche dei valori di concentrazione degli inquinanti atmosferici a disposizione sono state considerate come "dati funzionali", così da essere analizzate tramite la *functional data analysis* [Ram-02, Ram-05]. In primo luogo, le centraline di monitoraggio sono state classificate, per ogni inquinante, tramite la *functional cluster analysis*, introdotta da Abraham *et al.* [Abr-03] per la classificazione delle evoluzioni del pH nei processi di acidificazione per la produzione del formaggio. Successivamente, l'utilizzo congiunto dei dati di concentrazione degli inquinanti in forma funzionale con le informazioni relative alle caratteristiche delle centraline (quali caratteristiche della zona, tipo di stazione, tipo di zona, percentuale di edificato e densità di popolazione relativa) ha permesso un'ulteriore analisi della rete di monitoraggio. Le classificazioni così ottenute sono state in seguito confrontate con i risultati della *cluster analysis* classica eseguita sulle variabili caratterizzanti le centraline e su indicatori sintetici delle serie storiche degli inquinanti, quali la media, lo scarto quadratico medio ed un quantile appropriato.

Dopo aver introdotto brevemente la rete di monitoraggio oggetto di studio nel paragrafo 2, il paragrafo 3 illustra l'approccio funzionale applicato. Nel paragrafo 4 viene presentata l'analisi sui tre inquinanti presi in considerazione e, di seguito, sono tracciate le conclusioni e gli eventuali sviluppi futuri (paragrafo 5). Infine, in appendice sono riportate in tabella le classificazioni risultanti dalle tre metodologie applicate su ogni inquinante.

2 La rete di monitoraggio

La prima rete pubblica di monitoraggio in Piemonte nasce nel 1971 nella città di Torino, ad opera del Comune, in ossequio alla legge n. 615 del 13 luglio 1966, per rilevare i dati di anidride solforosa. Fin dal 1988, a seguito del D.P.R. n. 203 del 24 maggio 1988 (che recepisce le direttive europee 80/779/CEE, 82/884/CEE, 85/203/CEE e 84/360/CEE), la Regione Piemonte ha elaborato un progetto di ristrutturazione generale di tutte le reti pubbliche esistenti sul territorio, per creare il Sistema Regionale di Rilevamento della Qualità dell'Aria. La rete di monitoraggio della qualità dell'aria attuale è stata realizzata facendo seguito al Piano regionale per il risanamento e la tutela della qualità dell'aria che in Piemonte è stato adottato con la legge regionale 43/2000, in attuazione e coerenza con le disposizioni del Decreto Legislativo n. 351/99 (recepimento nell'ordinamento italiano della Direttiva 96/62/CE).

In questo studio si analizzano i dati di concentrazione di PM₁₀, NO₂ e O₃ rilevati dal "Sistema Regionale di Rilevamento della Qualità dell'Aria" della Regione Piemonte nel corso dell'anno 2004. La Fig. 1 mostra la localizzazione delle stazioni¹ di monitoraggio nell'anno analizzato, che sono per

¹1 A.Terme, 2 AL N.Orti, 3 AL D'Annunzio, 4 AL Libertà, 5 C.Monferrato Castello, 6 C.Monferrato Negri, 7

lo più poste nella parte pianeggiante della regione. Per quanto riguarda il particolato atmosferico sono state prese in considerazione 23 stazioni di misura dotate di strumento gravimetrico LV (*Low Volume*) e 5 di strumento TEOM (*Tapered Element Oscillating Microbalance*); 55 e 25 sono invece le centraline analizzate rispettivamente per il biossido di azoto e per l'ozono, le metodologie di misura utilizzate per questi due inquinanti sono rispettivamente a chemiluminescenza e ad assorbimento di ultravioletti (per ulteriori indicazioni relative al metodo per il campionamento e la misurazione di ogni inquinante si vedano le normative di riferimento - [DM.60/02] e [Dl.183/04] - e le norme EN 12341 per il PM₁₀, ISO 7996 per l'NO₂ e ISO FDIS 13964 per l'O₃).

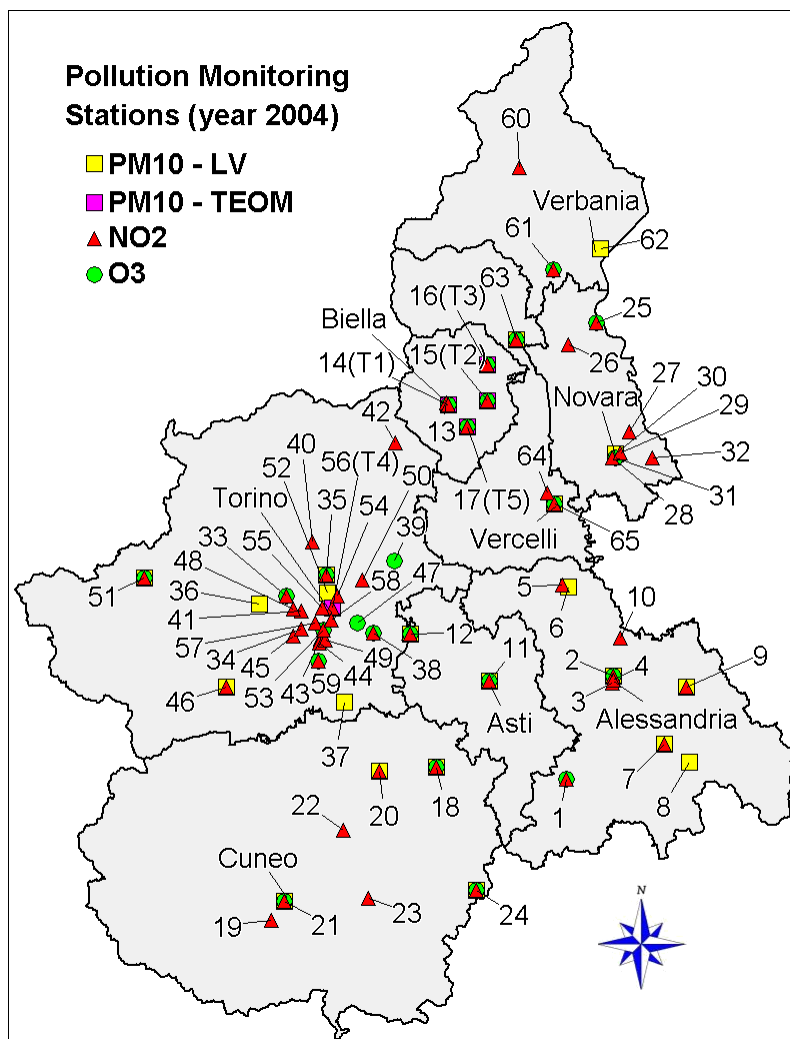


Figura 1: Localizzazione delle stazioni di monitoraggio nel 2004

Dal punto di vista normativo, il Decreto Ministeriale n. 60 del 2 aprile 2002 che recepisce la Direttiva 1999/30/CE concernente i valori limite di qualità dell'aria ambiente per biossido di zolfo,

N.Ligure, 8 S.Scrivia, 9 Tortona, 10 Valenza Po, 11 AT D'Acquisto, 12 Butt.d'Asti, 13 BI Lamarmora, 14 BI Sturzo, 15 Cossato, 16 Ponzone, 17 Verrone, 18 Alba, 19 B.S.Dalmazzo, 20 Bra, 21 CN Alpini, 22 Fossano, 23 Mondovì, 24 Saliceto, 25 Arona, 26 Borgomanero, 27 Cameri, 28 NO Leonardi, 29 NO Bovio, 30 NO Roma, 31 NO Verdi, 32 Trecate, 33 Alpignano, 34 Beinasco, 35 Borgaro, 36 Butt.Alta, 37 Carmagnola, 38 Chieri, 39 Chivasso, 40 Ciriè, 41 Grugliasco, 42 Ivrea, 43 Nichelino, 44 NichelinoAEM, 45 Orbassano, 46 Pinerolo, 47 P.Torinese, 48 Rivoli, 49 Sangone, 50 Settimo, 51 Susa, 52 TO Grassi, 53 TO Lingotto, 54 TO Rebaudengo, 55 TO Rivoli, 56 TO Consolata, 57 TO Gaidano, 58 TO M.Cristina, 59 Vinovo, 60 Domodossola, 61 Omegna, 62 Verbania, 63 Borgosesia, 64 Caresanablot, 65 VC Gastaldi, T1 BI Don Sturzo, T2 Cossato, T3 Ponzone, T4 TO Consolata, T5 Verrone

biossido di azoto, ossidi di azoto, particelle e piombo e il Decreto Legislativo n. 183 del 21 maggio 2004, attuazione della Direttiva 2002/3/CE relativa all'ozono nell'aria impongono che, per ogni stazione di rilevamento, il numero di dati effettivamente registrati sia pari almeno al 90% del totale da acquisire nel periodo di riferimento, allo scopo di rispettare gli obiettivi di qualità dei dati.

Per ciascuno dei tre inquinanti abbiamo un collettivo costituito dalle serie storiche annuali registrate dalle stazioni di misura a disposizione. Per quanto riguarda il PM_{10} , gli strumenti LV e TEOM forniscono rispettivamente misure con frequenza giornaliera e oraria; vengono allora calcolate le medie giornaliere sui dati rilevati dal secondo strumento, così da poter analizzare in modo omogeneo la totalità delle serie storiche. Pertanto, ciascuna serie storica nella i -esima stazione, per $i = 1, \dots, n_{PM_{10}} = 28$, sarà costituita da $m_{(i, PM_{10})} = 366$ dati $\{y_{i,j}\}_{j=1, \dots, m_{(i, PM_{10})}}$, mentre ogni serie storica di NO_2 e O_3 , registrati con frequenze orarie, sarà data da $\{y_{i,j}\}_{j=1, \dots, m_{(i, NO_2)}}$, per $i = 1, \dots, n_{NO_2} = 55$ e $\{y_{i,j}\}_{j=1, \dots, m_{(i, O_3)}}$, per $i = 1, \dots, n_{O_3} = 25$, dove $m_{(i, NO_2)}$ e $m_{(i, O_3)}$ sono uguali a 8784. L'inserimento di dati dove le osservazioni sono mancanti non è necessaria in quanto la metodologia utilizzata ne effettua la stima (vedi par. 3)

3 L'approccio all'analisi dei dati

Il principio alla base della *Functional Data Analysis* è quello di considerare i dati registrati in tempi discreti come dati funzionali, assumendo cioè l'esistenza di una funzione continua che genera le osservazioni [Ram-02, Ram-05]. Così facendo, i valori di concentrazione che costituiscono le serie storiche di un inquinante possono essere visti come realizzazioni di processi continui monitorati in tempi discreti. Per adattare le osservazioni discrete $y_{i,j}$, del generico inquinante INQ , con $i = 1, \dots, n_{INQ}$ e $j = 1, \dots, m_{i, INQ}$, viene adottato il seguente modello:

$$y_{i,j} = G_i(t_j) + \epsilon_{i,j}, \quad (1)$$

dove G_i rappresenta una funzione sufficientemente liscia valutata al tempo t_j e gli $\epsilon_{i,j}$ sono errori casuali indipendenti. Il passaggio da dati discreti a funzioni continue avviene tramite tecniche di *smoothing*, utilizzando combinazioni lineari di funzioni *B-spline* che consentono di estendere i vantaggi dei polinomi pur garantendo alta flessibilità alle funzioni stimate [dBo-01].

Si consideri $t \in [a, b]$ e sia $(\xi_0 =) a < \xi_1 < \xi_2 < \dots < \xi_K < b (= \xi_{K+1})$ una suddivisione di $[a, b]$ tramite K punti distinti chiamati nodi. La funzione *spline* $s(t)$ è una polinomiale di grado d (o di ordine $d+1$) su ogni intervallo $[\xi_{i-1}, \xi_i]$, avente $d-1$ derivate continue nell'intervallo aperto (a, b) . Per una sequenza fissata di nodi $\xi = (\xi_1, \xi_2, \dots, \xi_K)$, una *B-spline* può esser scritta come

$$s(t, \beta) = \sum_{l=1}^{K+d+1} \beta_l B_l(t),$$

dove $\beta = (\beta_1, \dots, \beta_{K+d+1})'$ è il vettore dei coefficienti *spline*.

Denotando con B la matrice di dimensione $m_{i, INQ} \times \{K + d + 1\}$, che contiene i valori assunti dalle funzioni B_l in corrispondenza dei punti t_j , i coefficienti delle funzioni *B-spline* vengono stimati attraverso il criterio dei minimi quadrati da cui si ottiene, per ogni i ,

$$\hat{\beta}_i := \arg \min_{\beta_i} \frac{1}{m_{i, INQ}} \sum_{j=1}^{m_{i, INQ}} (y_{i,j} - s(t_j, \beta_i))^2 = [(B)'B]^{-1}(B)'y_i,$$

dove $B'B$ è supposta essere non singolare.

Quanto sopra permette di interpretare la serie storica come dato funzionale: i valori del modello (1) vengono infatti stimati come $\hat{y}_{i,j} = \widehat{G}_i(t_j)$ attraverso la combinazione lineare di osservazioni

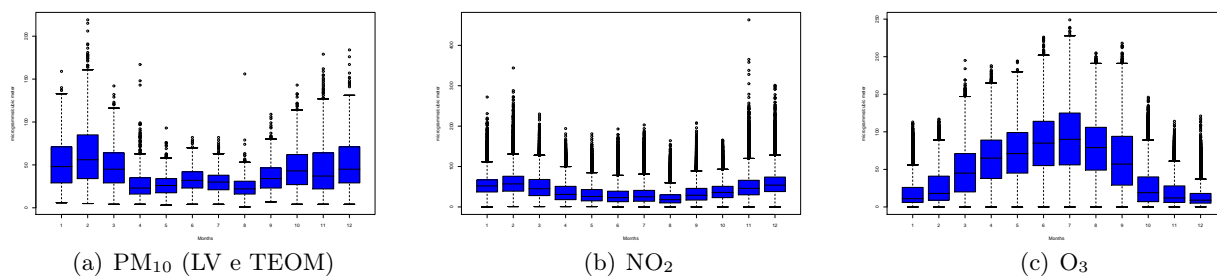


Figura 2: Boxplot per mese delle distribuzioni dei tre inquinanti nel 2004

discrete, ossia

$$\widehat{G}_i(t_j) = \sum_{k=1}^{m_{i,INQ}} s_j(t_k) y_{i,k},$$

dove $s_j(t_k)$ pesa i valori dei dati discreti per generare gli $\widehat{y}_{i,j}$, per ogni i e per ogni tempo t_j , anche in corrispondenza delle osservazioni mancanti.

L'insieme delle funzioni $\{G_1, \dots, G_{n_{INQ}}\}$ delle n_{INQ} serie storiche di ciascun inquinante è riassunto attraverso i coefficienti stimati $\{\widehat{\beta}_1, \dots, \widehat{\beta}_{n_{INQ}}\}$, avendo quindi un insieme di vettori in \mathbb{R}^{K+d+1} .

Applicando le procedure di *clustering* classiche ai dati funzionali così ottenuti, si effettua la *functional cluster analysis (FCA)* [Abr-03]. L'interesse principale della *FCA*, relativamente al presente studio, consiste nel definire gruppi di curve con simili livelli ed evoluzioni temporali di concentrazione per ogni inquinante e nell'esplorare i profili dei diversi *cluster* formatisi. Dal momento che in tutte le n_{INQ} serie storiche di ciascun inquinante vengono utilizzati lo stesso grado del polinomio e lo stesso vettore dei nodi, le funzioni (B_1, \dots, B_{K+d+1}) adottate rimangono sempre le stesse; pertanto le $K + d + 1$ componenti del vettore $\widehat{\beta}_i$ rappresentano lo stesso tipo di "carattere" per ogni curva G_i . Per suddividere le n_{INQ} curve relative a ciascun inquinante in k_{INQ} *cluster* è allora sufficiente classificare i loro coefficienti $\widehat{\beta}_i$. Tale classificazione viene effettuata tramite un algoritmo non gerarchico *K-means*, che richiede però la precisazione del numero k_{INQ} di gruppi che, nella presente analisi, è determinato applicando preventivamente il metodo di *Ward* (metodo gerarchico) agli stessi dati funzionali.

La scelta di applicare la procedura *K-means* è giustificata dai risultati di Abraham *et al.* [Abr-03], secondo i quali lo stimatore dei centri dei gruppi, fissato il numero di *cluster*, è fortemente consistente.

4 Analisi per i tre inquinanti in Piemonte

Poichè il metodo presentato si propone di realizzare una partizione dei dati per ciascun inquinante (PM_{10} , NO_2 e O_3) tenendo conto della loro forma funzionale mediante funzioni *B-spline*, è necessario scegliere opportunamente il grado e l'insieme ξ dei nodi interni delle *spline*, la cui dimensione è legata alla loro flessibilità. Si constata che una *spline* cubica si adatta molto bene alle serie storiche oggetto di studio e, dal momento che all'interno delle serie storiche non esistono particolari periodi che necessitano analisi più approfondite e quindi un aumento di numero di nodi, quest'ultimi sono mantenuti ugualmente spaziali. Tenendo conto dell'andamento nel tempo di ciascun inquinante in ogni centralina e dell'andamento sull'intera regione (visualizzato nei boxplot per mese di Fig. 2), i nodi vengono posizionati allo scadere di ogni mese. Si utilizzano quindi funzioni *B-spline* di ordine 4 e 11 nodi interni.

Una volta definite le basi e ottenuti i dati nella forma funzionale, si procede innanzitutto nella classificazione delle curve. Come già accennato in precedenza, si applica un metodo gerarchico (di *Ward*) ai coefficienti $\hat{\beta}_i$ ottenuti per ogni inquinante, in modo da poter scegliere il numero ottimale di raggruppamenti in cui suddividere successivamente il collettivo attraverso l'uso di un metodo non gerarchico (*K-means*).

Successivamente, si effettua un'ulteriore analisi, detta in seguito *FCA* "modificata" o *FCAm*, aggiungendo le informazioni relative alle caratteristiche delle centraline e della loro zona di collocazione, per esaminare ulteriormente la rete di monitoraggio e per comprendere quanto queste influenzino la classificazione. Si hanno a disposizione le seguenti informazioni:

- percentuale di edificato nell'intorno di 1 Km dal punto di collocazione della stazione (da cui usualmente si ricava il tipo di zona: *urbana*, *periferica* e *rurale*);
- caratteristica della zona (*commerciale*, *industriale*, *residenziale* e *agricola*);
- tipo di stazione (*traffico* legata al tipo di zona, *fondo* legata all'intensità del traffico nell'intorno della stazione, e *industriale* - [2001/752] e [APAT] -);
- densità di popolazione relativa, data dal rapporto tra il numero di residenti presenti in un dato comune e la superficie edificata dello stesso (grandezza che in questo caso risulta più adatta rispetto all'area totale del comune poichè a molti di essi appartengono vaste zone pressochè disabitate, come per esempio le aree montane).

Queste variabili unite ai coefficienti $\hat{\beta}_i$ per ogni inquinante costituiscono la nuova matrice dati utilizzata nelle procedure *Ward* e *K-means* per effettuare la *FCAm*.

Infine, si vogliono confrontare le classificazioni ottenute dai due precedenti approcci con i risultati della *cluster analysis* semplice applicata alle stesse 4 variabili esogene presentate per la *FCAm* unitamente ai seguenti indicatori sintetici:

- media annuale della concentrazione di inquinante;
- scarto quadratico medio della concentrazione di inquinante;
- quantile appropriato.

4.1 *Il particolato atmosferico*

I dati di PM_{10} analizzati si riferiscono all'anno 2004 in cui la rete di monitoraggio attiva sul territorio piemontese era composta da 28 siti con strumenti di misura di tipo gravimetrico LV e 6 di tipo TEOM. In accordo con i criteri di raccolta minima dei dati [DM.60/02], ne sono stati scartati 5 e pertanto si analizzano i valori provenienti da 23 stazioni di misura di tipo LV e 5 TEOM. È da notare che a Torino in Via Consolata e a Biella in Via Don Sturzo coesistono strumenti LV e TEOM; ciò permette di avere indicazioni e conferme relative agli effetti del tipo di strumento sulle misurazioni [Fas-05]. La Fig. 3 mostra, ad esempio, la serie storica della stazione di TO Consolata misurata attraverso uno strumento di tipo LV.

I boxplot per mese (Fig. 2(a)) mettono in evidenza il comportamento della concentrazione di PM_{10} nell'anno sulla regione Piemonte. Come è noto, le polveri sottili si accumulano maggiormente durante i mesi invernali mentre raggiungono valori inferiori durante l'estate. La distribuzione in figura mostra che per l'anno 2004 il mese più critico risulta essere febbraio; questo risultato è di difficile interpretazione poichè, effettuando un controllo sui dati meteorologici, le precipitazioni appaiono nella media per gennaio e marzo, ma nettamente superiori ad essa per febbraio [Aria-04]. La causa di questo comportamento può esser allora dovuta ad una condizione sfavorevole delle altre variabili meteorologiche in febbraio rispetto agli altri due mesi.

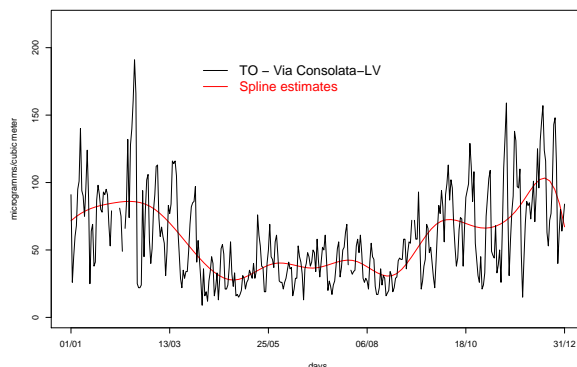


Figura 3: Serie storica e dato funzionale per il PM_{10} nella stazione di TO Consolata con strumento di misura del tipo LV

Come accennato in precedenza, i dati di particolato analizzati nel presente studio sono rilevati con frequenze giornaliere dallo strumento LV, orarie dal TEOM; per utilizzare congiuntamente la totalità di dati registrati son state calcolate le medie giornaliere sui valori misurati dal secondo strumento. I nodi interni scelti per la costruzione delle *B-spline* per il PM_{10} , come detto, suddividono l'intero anno in mesi e sono $\xi = (31, 61, 91, 121, 151, 181, 211, 241, 271, 302, 334)$. La Fig. 3 mostra un esempio (TO Consolata) di come la curva stimata si adatta alla serie storica iniziale, mentre i dati funzionali relativi a tutte le 28 stazioni di misura sono rappresentati in Fig. 4.

Ciascuna curva è riassunta attraverso $K + d + 1 = 15$ coefficienti delle *B-spline* $\hat{\beta}_i$. Gli $n_{PM_{10}} = 28$ vettori $\hat{\beta}_i$ possono quindi essere classificati utilizzando preliminarmente il metodo gerarchico di *Ward* per avere un'indicazione sul numero di raggruppamenti utili per suddividere in modo ottimale le curve. Il dendrogramma di Fig. 5 è ottenuto dall'applicazione di tale metodo ai dati a disposizione. La figura mostra che, ad una altezza notevole corrispondente ad una distanza di Manhattan all'incirca di 300, vengono a formarsi due *cluster*: al primo gruppo appartengono centraline aventi livelli medi di concentrazione più alti, mentre nel secondo si collocano quelle aventi valori medi inferiori. Si nota inoltre che la stazione TO ITIS Grassi fa gruppo a sè, unendosi alle altre centraline ad un'altezza pari all'incirca a 370, forse a causa di valori molto elevati dell'inquinante per tutto il periodo di studio che portano ad una media annuale pari a $69,46 \mu g/m^3$. Si osserva inoltre che, come si poteva supporre, le stazioni di tipo TEOM sono raggruppate tra di loro ad esclusione di TO Consolata; probabilmente ciò accade non solo per l'impiego dello stesso tipo di strumento, ma anche perchè queste centraline sono collocate tutte in provincia di Biella e quindi misurano livelli simili. Per quanto riguarda TO Consolata, si può notare come le due stazioni presenti in quel punto appartengano allo stesso *cluster* nonostante lo strumento TEOM sia meno sensibile alla variabilità dell'inquinante come mostra la Fig. 6. Al contario, probabilmente proprio questa caratteristica di sottostima di tali centraline fa sì che i due strumenti collocati in BI Sturzo non siano raggruppati insieme.

Dopo aver osservato che gli strumenti TEOM si "isolano" in un gruppo a sè stante, si procede nell'analisi prendendo in considerazione solamente le centraline di misura di tipo LV ed escludendo inoltre la stazione TO ITIS Grassi, avendo già constatato che si "stacca" dal collettivo per creare un gruppo a parte. La Fig. 7(a) mostra il grafico ad albero così ottenuto con le 22 stazioni di misura rimanenti. Considerando i due *cluster* che si formano nel dendrogramma ad un'altezza rispettivamente di circa 300 e circa 400, è possibile vedere come nel primo gruppo si siano collocate le stazioni con valori di media annuale più elevata e nel secondo quelle con media inferiore.

Dall'analisi del dendrogramma ottenuto, il numero ottimale di *cluster* in cui suddividere i 22 vettori $\hat{\beta}_i$ sembrerebbe essere 4. È dunque possibile procedere utilizzando il metodo *K-means*.

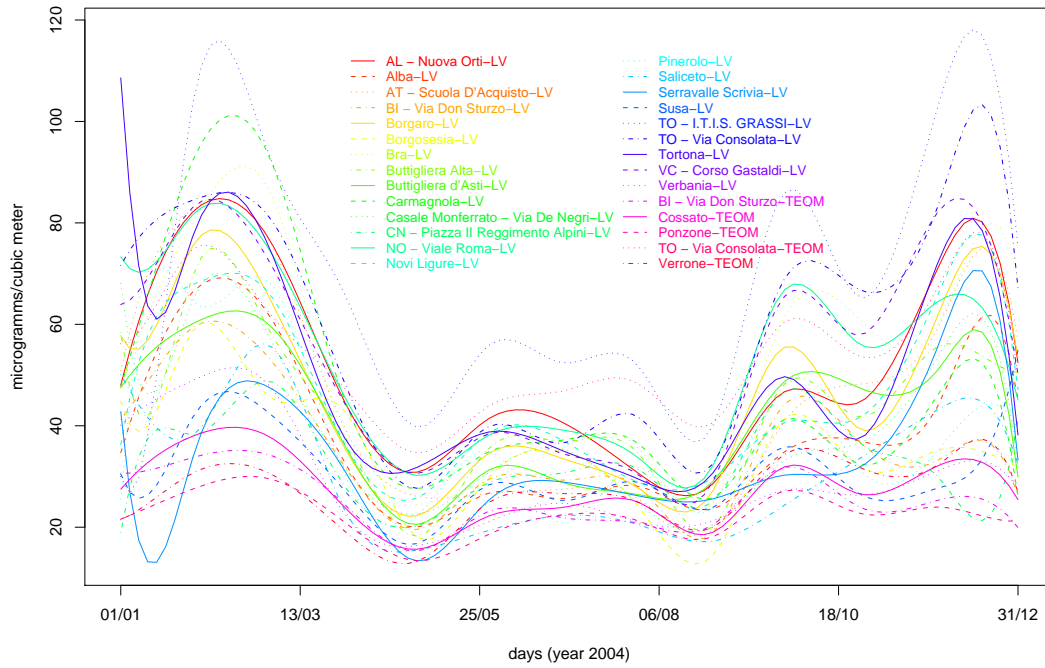


Figura 4: Dati funzionali per il PM₁₀ nel 2004

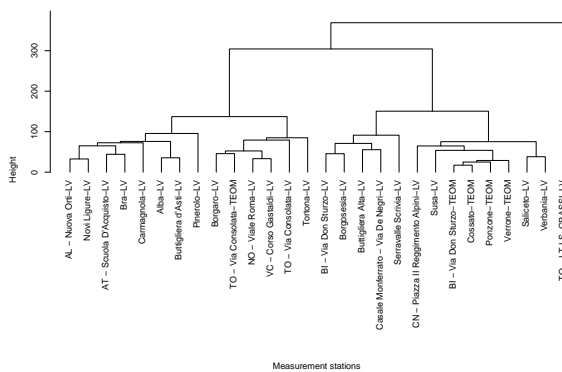


Figura 5: Dendrogramma ottenuto dal metodo di Ward per le stazioni di PM₁₀ di tipo LV e TEOM nel 2004

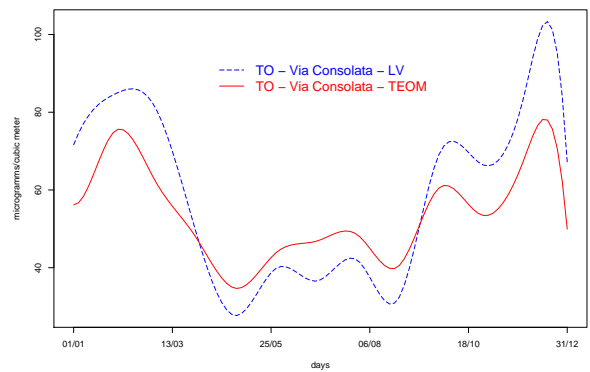


Figura 6: Dati funzionali nella stazione di TO Consolata per il PM₁₀ misurato dagli strumenti di tipo LV e TEOM nel 2004

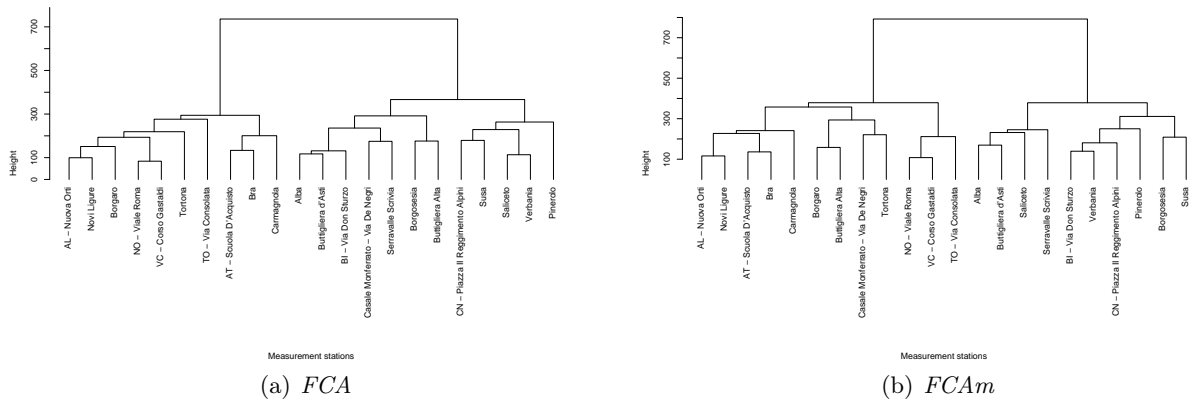


Figura 7: Dendrogrammi ottenuti dal metodo di *Ward* per il PM_{10} di tipo LV nel 2004

Esaminando il grafico delle curve raggruppate (Fig. 8(a)) e quello relativo ai centroidi dei *cluster* (Fig. 8(b)), si può affermare che l'appartenenza ad un dato gruppo derivi dalla similarità delle serie storiche secondo la loro evoluzione temporale e il livello di concentrazione dell'inquinante.

PM_{10}	Mean	St. Dev.	1st Qu.	Median	3rd Qu.
Cluster 1	31,31	19,89	17,00	27,00	40,00
Cluster 2	36,95	23,46	20,00	32,00	48,00
Cluster 3	42,71	26,64	24,00	35,00	55,00
Cluster 4	51,15	30,04	29,00	43,00	67,00
Cluster 5	69,46	36,62	43,00	62,00	90,50

Tabella 1: Indicatori sintetici dei *cluster* ottenuti tramite *FCA* per il PM_{10} nel 2004

È possibile infatti identificare (Fig. 8(b)) un gruppo di curve con livelli di inquinante inferiore (azzurro), la curva relativa alla stazione di misura TO Grassi (viola) avente livelli di concentrazione nettamente superiori rispetto agli altri *cluster* ottenuti, e tre *cluster* intermedi (blu, rosso e verde) con valori diversi durante l'arco di tutto l'anno.

La Tabella 1 sintetizza le caratteristiche dei *cluster* formati. Mettendo in relazione i risultati ottenuti dalla procedura *K-means* con le medie annuali per stazione di misura e le caratteristiche delle centraline (vedi Tabella 7 in Appendice), sembrerebbe che l'aspetto più influente sulla formazione dei gruppi sia stata la media annuale. Se, per esempio, si prende in considerazione il *cluster* 1, nonostante sia prevalentemente costituito da stazioni di tipo urbano (*urban*), come CN Alpini o BI Sturzo, questo risulta essere il *cluster* con media inferiore. Da notare, inoltre, che la stazione di Buttigliera d'Asti, di tipo rurale (*rural*), fa parte di un gruppo (*cluster* 3) con sola presenza di stazioni urbane; la stessa situazione si verifica nel *cluster* 4 per la centralina di tipo suburbano (*suburban*) di Borgaro. Riferendoci tuttavia alle medie annuali, Buttigliera risulta simile alle altre componenti del suo *cluster* di appartenenza, e ciò vale allo stesso modo per la seconda stazione presa in considerazione. Da ciò risulta evidente che le medie annuali più elevate si verificano nelle stazioni poste in città maggiormente soggette a traffico veicolare e ad altre sorgenti emmissive, mentre le situazioni di medie inferiori si riscontrano per le centraline in prossimità delle Alpi. Interpretando allora quanto detto in relazione alla Fig. 9(a), sembrerebbe che l'appartenenza ad un dato gruppo sia anche legata alla morfologia e alle caratteristiche del territorio.

Si procede ora nell'analisi utilizzando congiuntamente i dati di concentrazione (riassunti nei vettori $\hat{\beta}_i$), le categorie assegnate ad ogni centralina secondo quanto previsto dalla normativa e le informazioni relative alle caratteristiche della zona di ubicazione delle stazioni di misura; questo

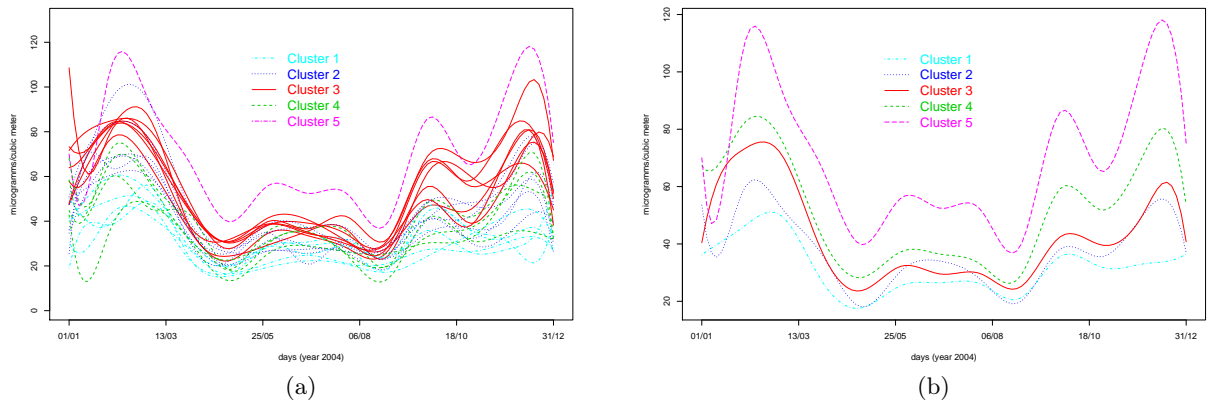


Figura 8: Dati funzionali raggruppati tramite *FCA* e relativi centroidi per il PM_{10} nel 2004

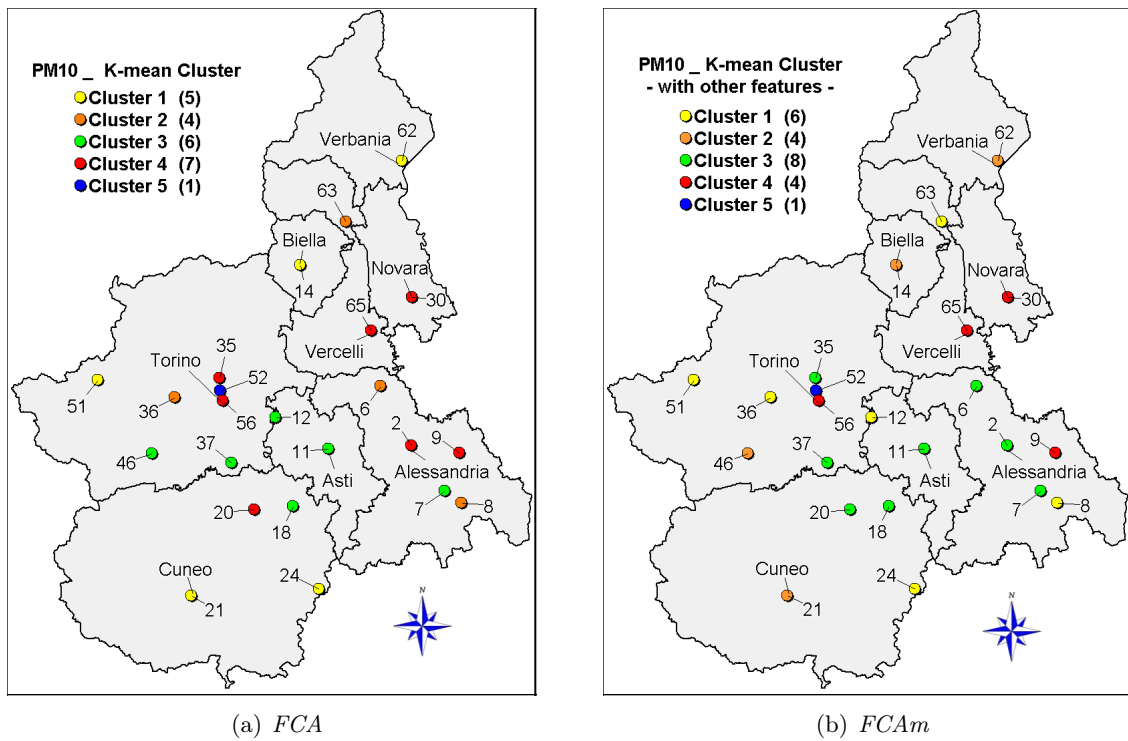


Figura 9: Localizzazione dei gruppi risultanti dalla *FCA* e dalla *FCAm* per il PM_{10} nel 2004

secondo approccio dovrebbe permettere di comprendere quanto le variabili esogene influenzino la classificazione.

Dall'applicazione del metodo gerarchico (Fig. 7(b)), il numero di *cluster* in cui suddividere le stazioni di misura rimane lo stesso. La sintesi dei *cluster* ottenuti con l'applicazione del metodo *K-means*, è riportata in Tabella 2. Rispetto ai risultati della prima parte dell'analisi (ossia l'applicazione della *functional cluster analysis* ai soli livelli di concentrazione) si può notare (vedi Tabella 8 in Appendice) che il *cluster* 1 mantiene solamente due elementi (Saliceto e Susa); i componenti del secondo gruppo per la *FCA* (ad eccezione di Casale Monferrato Negri) sono ora passati al primo *cluster* insieme alla stazione di Buttigliera d'Asti. Il *cluster* 2 è invece composto dai 4 rimanenti elementi del *cluster* 1 per la *FCA*; non essendo più presenti infatti due componenti con media annuale bassa, la media del gruppo così formato si è naturalmente alzata, portando ad uno scambio di posizione dei due raggruppamenti. Probabilmente la variabile "percentuale di edificato" ha giocato un ruolo importante in questo passaggio, anche se i livelli medi di concentrazione delle stazioni di Saliceto e Susa sono più compatibili con i rimanenti elementi del *cluster* 2. Si può inoltre osservare che anche il *cluster* 4 si è ridotto di numero, mantenendo 4 stazioni di tipo *urban*, aventi sia percentuali di edificato che valori medi di inquinante tra i più elevati.

Dalle mappe risultanti si può osservare che nella prima parte dell'analisi (Fig. 9(a)) le stazioni appartenenti al *cluster* 1 sono quelle più vicine all'arco alpino con una quota altimetrica media all'incirca pari a 400 m s.l.m.; per la loro collocazione esse potrebbero quindi risentire meno gli effetti del trasporto di emissioni provenienti dalle grandi città. È inoltre importante notare che le concentrazioni più elevate si riscontrano nelle stazioni poste nelle grandi città o nelle loro immediate periferie. La Fig. 9(b), relativa alla *FCAm*, mostra che, in linea di massima, i gruppi formati non sono così lontani da quelli della mappa precedente; si ampliano solamente i *cluster* intermedi. Tuttavia, sembrerebbe che quelli ottenuti dalla prima analisi, più che quelli creati dalla seconda, siano in relazione con la conformazione del territorio e la collocazione degli strumenti di misura in zone più o meno urbanizzate.

PM ₁₀	Mean	St. Dev.	1st Qu.	Median	3rd Qu.	Built % Mean
Cluster 1	34,98	22,64	19,00	29,00	45,00	78,08
Cluster 2	33,68	21,74	19,00	29,00	43,00	27,98
Cluster 3	45,02	27,70	25,00	37,00	58,00	61,34
Cluster 4	53,27	30,58	30,00	45,00	70,00	89,93
Cluster 5	69,46	36,62	43,00	62,00	90,50	98,90

Tabella 2: Indicatori sintetici dei *cluster* ottenuti tramite *FCAm* per il PM₁₀ nel 2004

Prendendo in considerazione l'applicazione della *cluster analysis* alle variabili elencate in precedenza (pag. 6) si ottengono i gruppi visualizzati in Fig. 10; il quantile scelto è quello di ordine 0,904 che secondo la normativa dovrebbe essere pari a $50 \mu\text{g}/\text{m}^3$, essendo tale valore la soglia da non superare per più di 35 volte in un anno (vedi art. 17, All. III, [DM.60/02]). È subito evidente la formazione di 4 gruppi anziché 5, come nei casi precedenti. Analizzando poi i componenti dei vari *cluster*, si trovano dei "rimiscolamenti" rispetto al passo precedente, causati dalla maggiore influenza che le caratteristiche delle centraline hanno rispetto ai valori di media annuale sulla suddivisione.

4.2 Il biossido di azoto

Nel 2004 la rete di monitoraggio per il biossido di azoto era composta da 63 stazioni di misura, di cui 55 hanno rilevato un numero di dati in accordo con i criteri di raccolta minima [DM.60/02]. L'andamento della concentrazione di questo inquinante nell'intero anno in Piemonte è presentato nei boxplot per mese (Fig. 2(b) pag. 5): il grafico evidenzia come i mesi invernali rappresentino il periodo di maggior criticità.

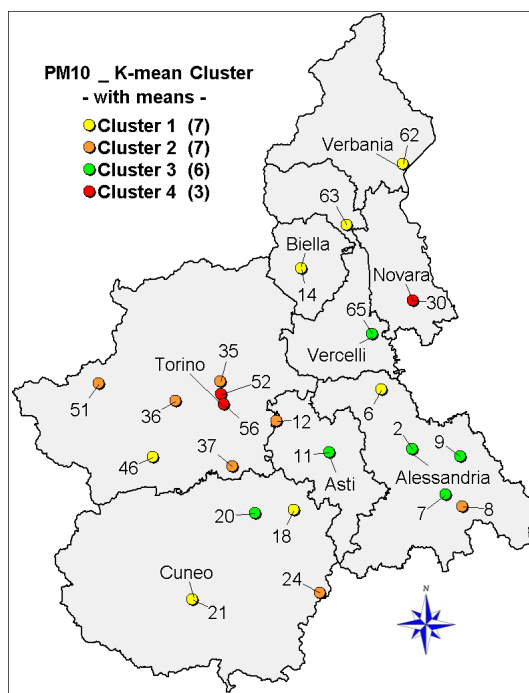


Figura 10: Localizzazione dei gruppi risultanti dalla *cluster analysis* per il PM₁₀ nel 2004

A differenza del PM₁₀, il biossido di azoto è rilevato con frequenza oraria, dando luogo ad una serie con 8784 valori possibili; i nodi interni scelti per la costruzione dei dati funzionali tramite *B-spline* (Fig. 11) diventano allora $\xi = (732, 1464, 2196, 2928, 3660, 4392, 5124, 5856, 6588, 7320, 8052)$, avendo scelto anche per questo inquinante un nodo per mese.

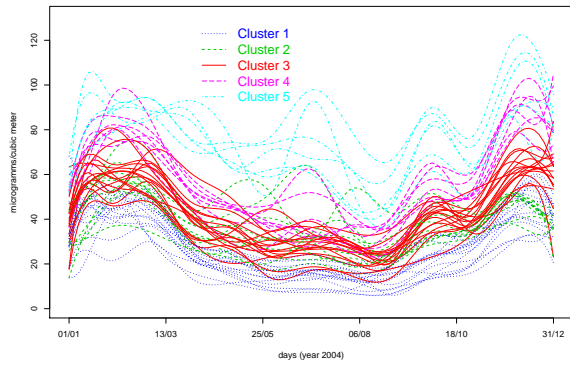
Dalla procedura di *Ward* (Fig. 12(a)) ad un'altezza di 1000 si osserva la formazione di due gruppi, le stazioni aventi media annuale più elevata appartengono al secondo di essi. In relazione a tale dendrogramma la scelta di suddividere il collettivo in 5 *cluster* sembra essere la più adeguata. Applicando la procedura *K-means* si ottiene la partizione rappresentata in Fig. 13(a). Similmente a quanto ottenuto per il particolato, dalla figura è evidente che la suddivisione è in relazione con le medie annuali di ciascuna stazione di misura e con l'evoluzione temporale di ogni serie storica: i centri dei *cluster* (Fig. 13(b)) identificano i diversi tipi di curve presenti.

NO ₂	Mean	St. Dev.	1st Qu.	Median	3rd Qu.
Cluster 1	27,14	19,21	12,00	22,00	38,00
Cluster 2	37,71	21,79	22,00	34,00	50,00
Cluster 3	40,23	24,59	21,00	36,00	55,00
Cluster 4	55,45	31,78	31,00	51,00	73,00
Cluster 5	75,35	30,93	54,00	72,00	93,00

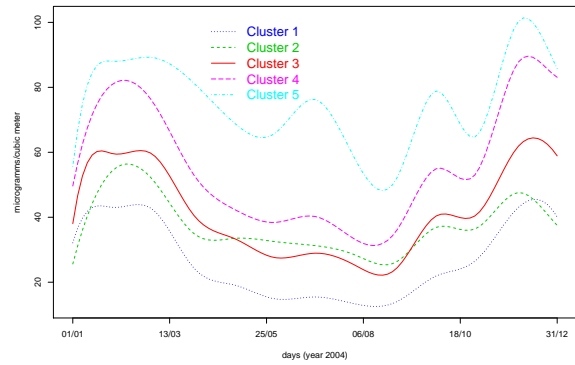
Tabella 3: Indicatori sintetici dei *cluster* ottenuti tramite *FCA* per l'NO₂ nel 2004

La Tabella 3 mette in evidenza le informazioni relative ai *cluster* ottenuti. Il gruppo più omogeneo per quanto riguarda le caratteristiche relative al tipo di zona e al tipo di stazione è il *cluster* 5 (vedi Tabella 10 in Appendice) in cui sono riunite tutte le stazioni collocate in Torino (tranne TO Lingotto). Prendendo inoltre in considerazione la mappa risultante dalla *FCA* (Fig. 14(a)), è possibile vedere come gli elementi dei *cluster* aventi medie annuali più elevate appartengono al centro di Torino e alla sua prima cintura, ad eccezione di AL D'Annunzio e NO Leonardi, mentre le stazioni di tipo *rural* fanno parte del *cluster* 1.

È da osservare che le due stazioni di misura collocate nel comune di Nichelino (Nichelino e

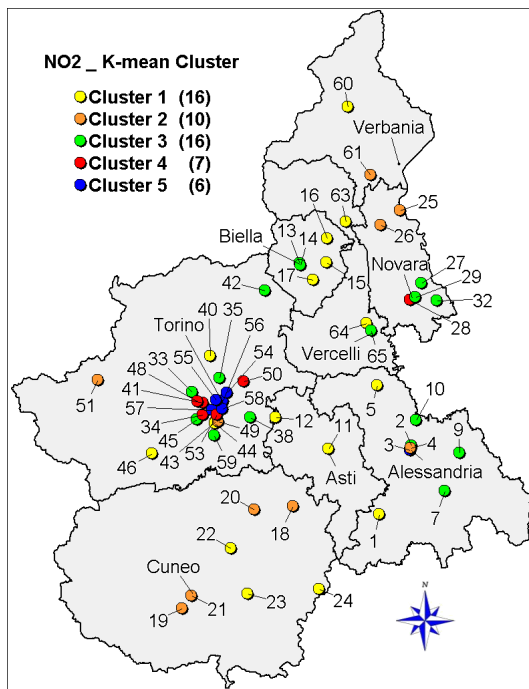


(a)

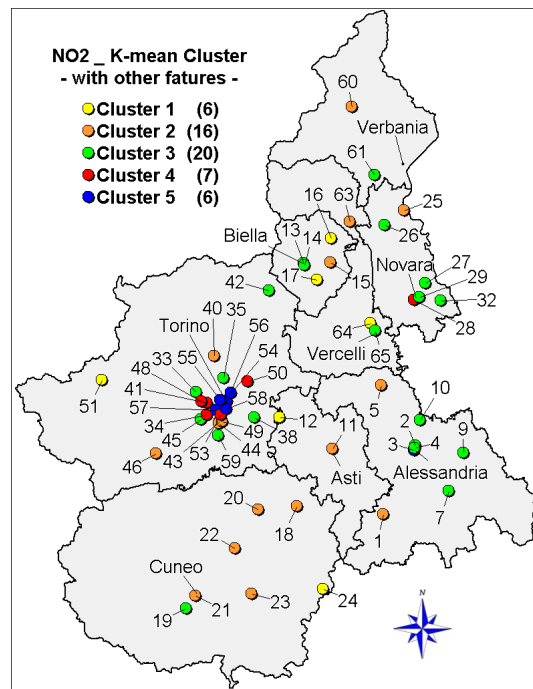


(b)

Figura 13: Dati funzionali raggruppati tramite *FCA* e relativi centroidi per l'NO₂ nel 2004



(a) *FCA*



(b) *FCAm*

Figura 14: Localizzazione dei gruppi risultanti dalla *FCA* e dalla *FCAm* per l'NO₂ nel 2004

Nichelino AEM) hanno livelli di concentrazione decisamente diversi; ciò può esser attribuito al fatto che Nichelino AEM è collocata in un parco, nonostante la sua percentuale di edificato (63,20%) risulti maggiore di quella dell'altra stazione (55%). Si noti ancora come le stazioni di Susa, CN Alpini e Borgo San Dalmazzo, inserite in zone pedemontane con simili caratteristiche territoriali e facenti tutte parte del secondo *cluster*, non abbiano valori annuali medi tra i più bassi.

Procedendo nell'analisi, il metodo gerarchico (Fig. 12(b)) sembrerebbe portare a suddividere i 55 vettori $\hat{\beta}_i$ sempre nello stesso numero di *cluster*. Una prima sintesi dei risultati ottenuti dall'applicazione della *FCAm* sono mostrati in Tabella 4. È subito evidente che i *cluster* 4 e 5 rimangono invariati (vedi Tabella 11 in Appendice). Al contrario si riduce il numero di elementi appartenenti al *cluster* 1: rimangono infatti solamente le stazioni di misura di tipo *suburban* e *rural*, aventi una percentuale di edificato molto bassa. Inoltre, quest'ultima variabile probabilmente ha influenzato il passaggio di undici unità dal *cluster* 1 al 2. Quattro siti che nel primo passo dell'analisi facevano parte del *cluster* 2 (AL Libertà, Borgo San Dalmazzo, Borgomanero e Omegna) ora si trovano nel terzo gruppo: le loro medie annuali risultano meglio collocate in questo gruppo rispetto al precedente.

NO ₂	Mean	St. Dev.	1st Qu.	Median	3rd Qu.	Built % Mean
Cluster 1	25,76	18,54	11,00	21,00	36,00	23,33
Cluster 2	29,68	19,60	15,00	26,00	41,00	66,73
Cluster 3	41,92	23,98	23,00	38,00	56,00	65,39
Cluster 4	55,45	31,78	31,00	51,00	73,00	77,79
Cluster 5	75,35	30,93	54,00	72,00	93,00	92,57

Tabella 4: Indicatori sintetici dei *cluster* ottenuti tramite *FCAm* per l'NO₂ nel 2004

Come accade nella *FCA*, dalla Fig. 14(b) si può vedere che i siti posti in Torino sono inclusi in un unico *cluster*; inoltre, le stazioni situate nella “prima cintura” di Torino (similmente a quelle collocate nella seconda) sono raggruppate insieme. Nel passaggio dal primo passo dell'analisi al secondo, quasi tutta la provincia di Cuneo viene a far parte di uno stesso gruppo.

Dalle mappe ottenute si può osservare che il risultato principale dell'analisi sul biossido di azoto è la formazione di “poli di attrazione”: Torino, Novara e Alessandria sono dei punti di maggior concentrazione dell'inquinante. A partire da questi poli, il livello dell'inquinante si riduce in modo più o meno concentrico.

Infine, si confrontano i risultati ottenuti con quelli della *cluster* semplice. Viene presa in considerazione la variabile quantile di ordine 0,998 che secondo la normativa dovrebbe essere pari a 200 $\mu\text{g}/\text{m}^3$, soglia da non superare per più di 18 volte in un anno (vedi art. 4, All. II, [99/30/CE]). Si nota che il *cluster* 1 è lo stesso ottenuto dall'applicazione della *FCAm* che, come detto, include anche le variabili esogene. Per quanto riguarda gli altri gruppi sono avvenuti dei “rimiscolamenti” che sembrerebbero esser influenzati soprattutto dalle categorie del “tipo di stazione” e del “tipo di zona”. Dalla Fig. 15 si può osservare che tra i punti con livelli di concentrazione elevati si aggiungono anche stazioni collocate in zone periferiche aventi caratteristiche morfologiche differenti dalle grandi città, come ad esempio Borgo San Dalmazzo e Ivrea.

4.3 L'ozono

Per quanto riguarda i dati di O₃, nel 2004 la rete di monitoraggio era costituita da 30 strumenti di misura, di cui 25 rispondono ai requisiti di raccolta minima dei dati [Dlg.183/04]. La Fig. 2(c) (pag. 5) presenta i boxplot per mese nel periodo oggetto di studio. Come è noto, i valori di O₃ crescono in concomitanza di un intenso irraggiamento solare, condizione presente soprattutto nei mesi estivi; il grafico mostra infatti che nel 2004 i valori più alti sono stati osservati in luglio.

Come il biossido di azoto, l'ozono presenta 8784 tempi di osservazione possibili, poichè lo stru-

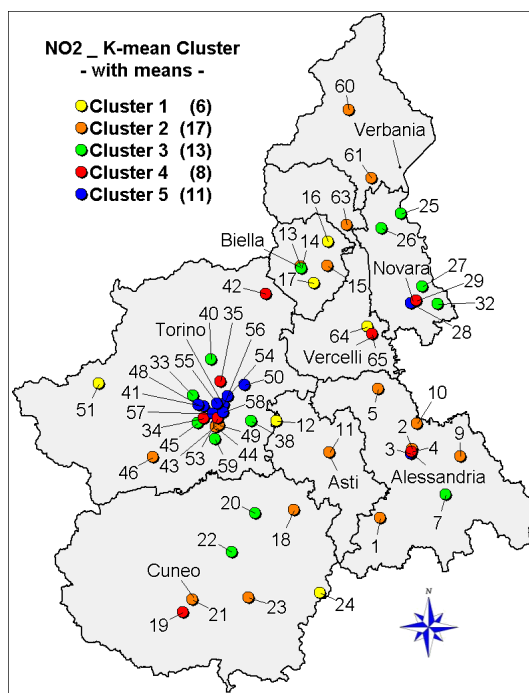


Figura 15: Localizzazione dei gruppi risultanti dalla *cluster analysis* per l'NO₂ nel 2004

mento fornisce una misura oraria; per la costruzione delle curve funzionali vengono dunque utilizzati gli stessi nodi interni (Fig. 16).

La Fig. 17(a) mostra il grafo ad albero ottenuto dall'applicazione del metodo gerarchico; mettendo in relazione i raggruppamenti che vengono a formarsi con un taglio ad altezza 450 del dendrogramma con la media annuale dell'inquinante studiato, sembrerebbe che, anche per quanto riguarda l'ozono, la partizione avvenga tenendo conto di questa caratteristica.

In riferimento a tale dendrogramma, il numero ottimale di *cluster* in cui ripartire i 25 vettori $\hat{\beta}_i$ sembrerebbe 3. È ora possibile applicare la procedura *K-means* ottenendo la suddivisione presentata in Fig. 18(a) e i relativi centroidi (Fig. 18(b)). Esaminando l'output della *FCA*, risulta ancora più evidente rispetto al PM₁₀ e all'NO₂ che la suddivisione in *cluster* avvenga secondo una "somiglianza" del livello di concentrazione dell'inquinante nelle varie stazioni di misura. È infatti possibile individuare tre tipi di curve (Fig. 18(a)): quelle aventi valori di inquinante più bassi (rosso), quelle con un livello più elevato (blu) e il gruppo caratterizzato da valori intermedi (verde).

Nella Tabella 5 vengono indicate le caratteristiche dei *cluster* ottenuti. Le stazioni di tipo *rural* si vanno a collocare nel terzo gruppo (avente valori di concentrazione maggiori) dove sono presenti stazioni con percentuali di edificato relativamente basse e con altitudini tra le più alte riscontrate (vedi Tabella 13 in Appendice); ciò è dovuto alla natura secondaria dell'ozono che si forma, in condizioni di sufficiente irraggiamento, a partire dagli ossidi di azoto e dai composti organici volatili. La Fig. 19(a) mette in risalto il fatto che le stazioni collocate in zone più vicine all'arco alpino siano quelle con medie più alte, in contrasto con quanto succedeva per il biossido di azoto.

O ₃	Mean	St. Dev.	1st Qu.	Median	3rd Qu.
Cluster 1	45,01	40,71	10,00	33,00	71,00
Cluster 2	54,31	43,36	15,00	45,00	84,00
Cluster 3	64,68	42,19	28,00	62,00	94,00

Tabella 5: Indicatori sintetici dei *cluster* ottenuti tramite *FCA* per l'O₃ nel 2004

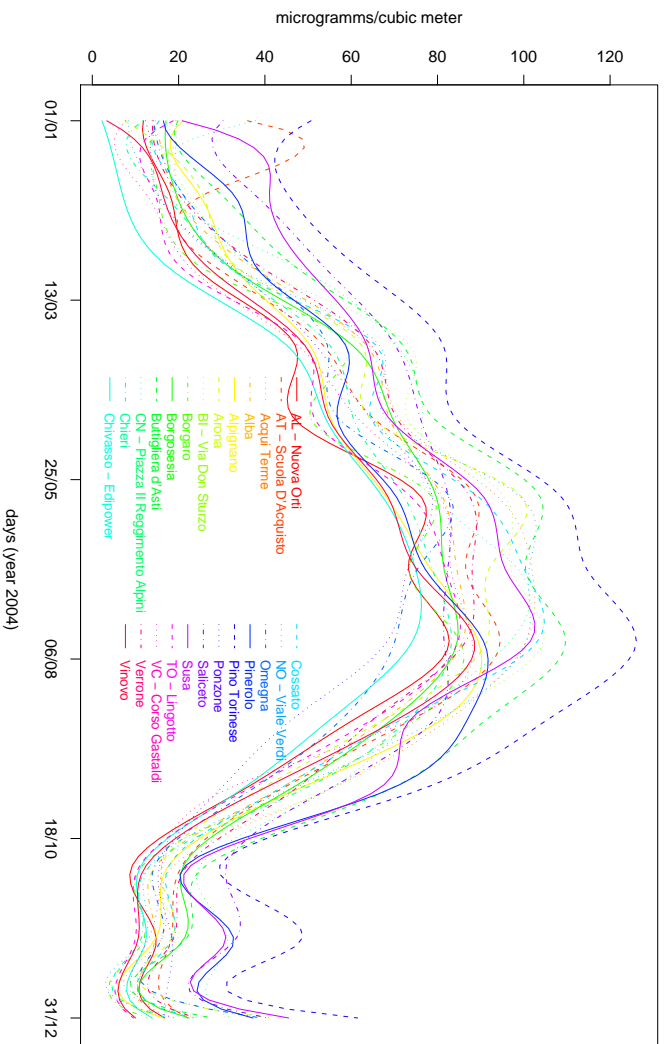


Figura 16: Dati funzionali per l'O₃ nel 2004

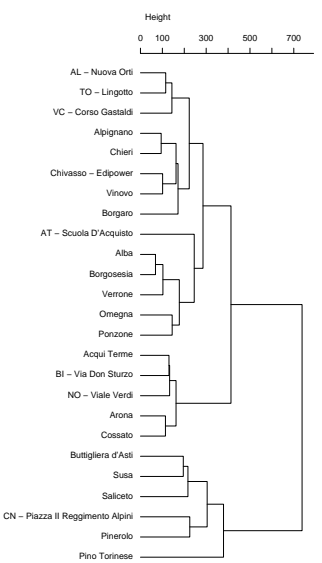
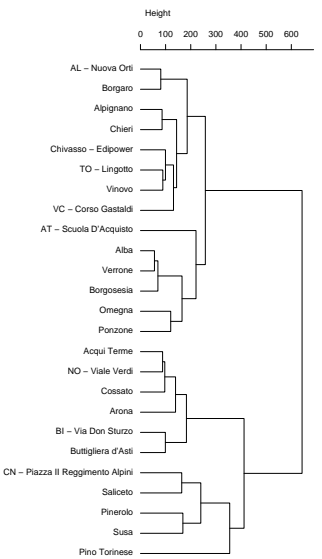
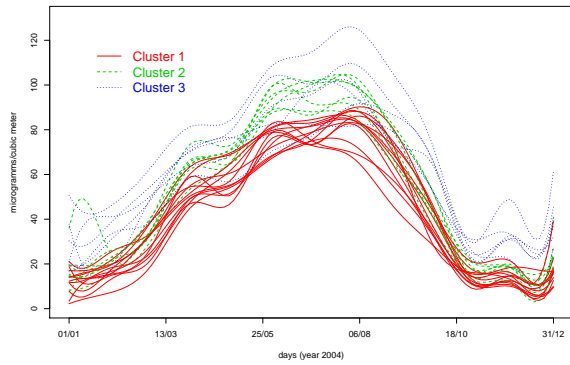
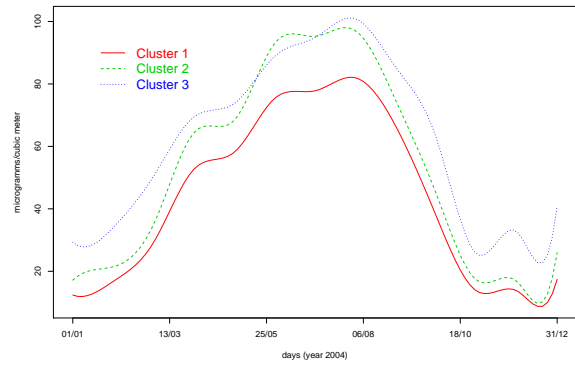


Figura 17: Dendrogrammi ottenuti dal metodo di *Ward* per le stazioni di O₃ nel 2004

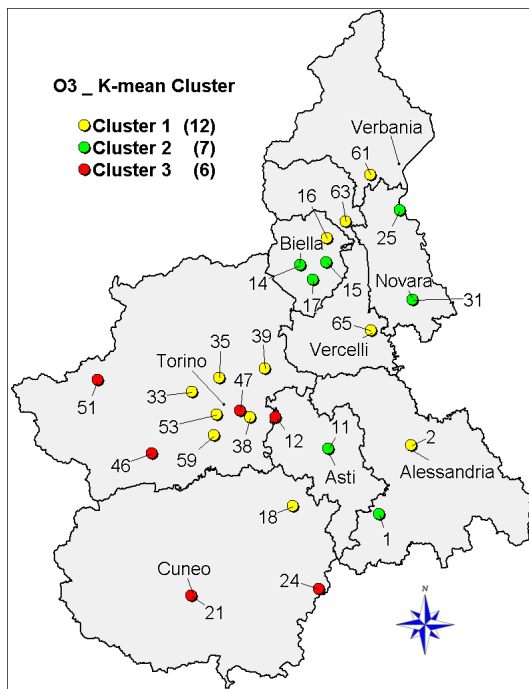


(a)

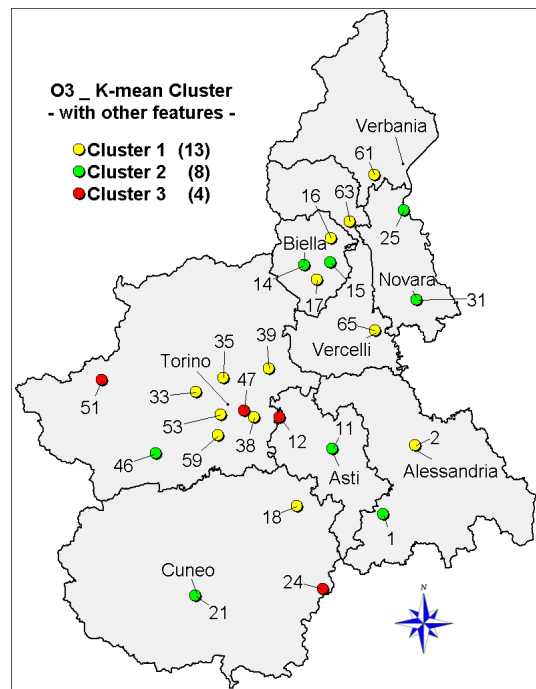


(b)

Figura 18: Dati funzionali raggruppati tramite *FCA* e relativi centroidi per l'O₃ nel 2004



(a) *FCA*



(b) *FCAm*

Figura 19: Localizzazione dei gruppi risultanti dalla *FCA* e dalla *FCAm* per l'O₃ nel 2004

Come per il particolato e il biossido di azoto, si procede ora ad una nuova analisi inserendo dati aggiuntivi, relativi alla zona di ubicazione delle centraline e alle proprietà delle stesse. Anche per quanto riguarda questo inquinante, il numero di *cluster* in cui vengono ripartiti i 25 vettori $\hat{\beta}_i$ rimane invariato rispetto alla prima parte dell'analisi (Fig. 17(a)); la Tabella 6 sintetizza le principali caratteristiche dei gruppi formatisi.

O ₃	Mean	St. Dev.	1st Qu.	Median	3rd Qu.	Built % Mean
Cluster 1	44,48	40,55	10,00	32,00	70,00	53,63
Cluster 2	53,59	44,49	13,00	43,00	84,00	74,61
Cluster 3	61,64	41,15	26,00	57,00	90,00	17,03

Tabella 6: Indicatori sintetici dei *cluster* ottenuti tramite *FCAm* per l'O₃ nel 2004

Il passaggio dalla *FCA* alla *FCAm* vede l'aggiunta al *cluster* 1 della stazione Verrone (che in un primo momento faceva parte del gruppo 2): analizzando i valori misurati da tale stazione si osserva che i livelli medi di concentrazione sono più compatibili con questo nuovo *cluster* (vedi Tabella 14 in Appendice). Le stazioni di Cuneo e Pinerolo passano dal *cluster* 3 al 2, anche se Cuneo si inseriva meglio nel gruppo precedente per quanto riguarda la media annuale dell'inquinante. Probabilmente questo cambiamento è dovuto all'influenza della variabile "percentuale di edificato"; in questo modo il *cluster* 3 risulta formato da stazioni aventi un basso valore di quest'ultima variabile che porta ad un dato medio per gruppo pari a 17,03%.

Come è possibile osservare anche dalla Fig. 19(b), con l'aggiunta nell'analisi delle variabili esogene le stazioni situate nelle strette vicinanze dell'arco alpino non fanno più parte di uno stesso *cluster*, mentre la maggior parte di quelle collocate nel torinese sono raggruppate insieme. Proprio per questo motivo sembrerebbe che i risultati della *FCA* rispecchino meglio la morfologia del territorio.

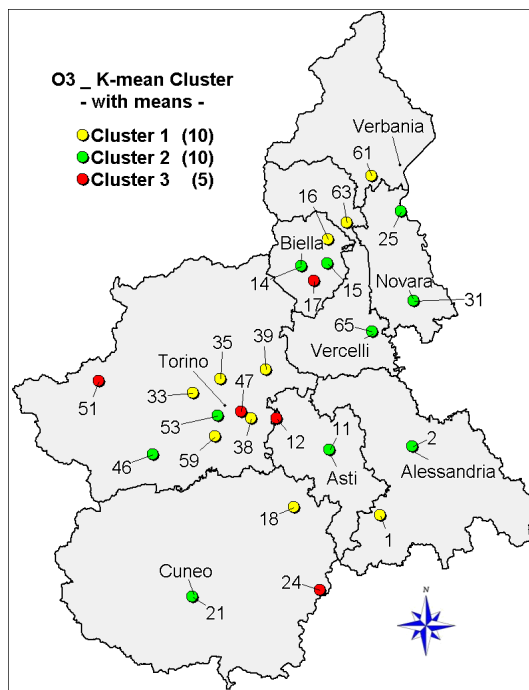


Figura 20: Localizzazione dei gruppi risultanti dalla *cluster analysis* per l'O₃ nel 2004

Per quanto riguarda l'applicazione dell'analisi *cluster* semplice, in questo caso si utilizza la variabile quantile di ordine 0,98. Per questo inquinante, la normativa non fornisce un valore limite

come per PM_{10} e NO_2 . Riferendoci al valore bersaglio per la protezione della salute umana (vedi art. 3, All. I, [02/3/CE]) che stabilisce che non si dovrebbe superare per più di 25 giorni per anno civile come media su 3 anni per la massima media mobile giornaliera la soglia pari a $120 \mu g/m^3$, ed in relazione a quanto calcolato per gli altri due inquinanti, si è scelto questo valore di controllo. Come è possibile notare dalla Fig. 20 la formazione dei gruppi ha risentito molto dell'influenza della variabile "percentuale di edificato". Anche in questa analisi le stazioni di tipo *rural* continuano a far parte del *cluster 3* avente i livelli di concentrazione più elevati.

5 Conclusioni e sviluppi futuri

In questo lavoro si è presentato un approccio metodologico per la classificazione delle stazioni di misura di una rete di monitoraggio della qualità dell'aria. A tal fine sono state effettuate sintesi di serie storiche, da un lato tramite coefficienti di *B-spline* e dall'altro tramite indicatori sintetici. In primo luogo, allora, al fine di rispettare la forma funzionale dei dati di concentrazione degli inquinanti atmosferici, è stata effettuata una parametrizzazione di funzioni usando le *B-spline*, che seguono lo stesso principio dell'*empirical orthogonal function (EOF) analysis*, impiegata da Ludwig *et al.* [Lud-95], di calcolare parametri che ben rappresentino i dati iniziali. In questo modo è stato possibile riassumere l'informazione contenuta nelle serie storiche attraverso un esiguo numero di coefficienti e quindi raggruppare le curve funzionali. In secondo luogo, si sono adottati indicatori sintetici, quali media annuale, scarto quadratico medio e quantile della concentrazione di inquinante. Ai due approcci è stata in seguito applicata la stessa procedura di *clustering*, in cui come metodo di classificazione è stato scelto di utilizzare l'algoritmo *K-means* (metodo non gerarchico), mentre per stabilire il numero di *cluster* ottimale k in cui suddividere il collettivo, si è applicato il metodo di *Ward* (gerarchico).

L'applicazione della *functional cluster analysis* alle serie storiche di PM_{10} , NO_2 e O_3 porta ad un raggruppamento che tende a seguire l'evoluzione temporale delle curve e la similarità tra i livelli di concentrazione. In questo caso la geografia del territorio e le altre variabili esogene non giocano un ruolo attivo nella classificazione, ma vengono utilizzate per interpretare i risultati ottenuti. Per quanto concerne il particolato e il biossido di azoto, le stazioni appartenenti ai *cluster* con valori medi annuali di inquinante più elevati corrispondono a zone con livelli di urbanizzazione e/o industrializzazione abbastanza elevati. L'aumento di ozono, al contrario, corrisponde ad una riduzione della densità di popolazione e di traffico, e ad un cambiamento del "tipo di zona".

La rappresentazione su mappa dei *cluster* formati ha permesso di evidenziare la presenza di *pattern* spaziali: le rappresentazioni relative al particolato e al biossido di azoto rilevano una situazione simile, nonostante la grande differenza del numero di stazioni di misura tra i due inquinanti. Le mappe dell' NO_2 infatti presentano, in modo evidente più che per il PM_{10} , Torino, Novara e Alessandria come punti di maggior concentrazione dell'inquinante: a partire da questi "poli di attrazione" il livello dell'inquinante si riduce in modo più o meno concentrico, così che le centraline collocate in zone più vicine all'arco alpino vengono a collocarsi nel gruppo avente media più bassa. La natura secondaria dell'ozono produce invece come effetto valori di media annuale tra i più alti per le stesse stazioni di misura "pedemontane".

Confrontando i risultati ottenuti attraverso la *FCA* e quelli relativi alla *FCAm* per osservare l'entità dell'influenza delle informazioni esterne sulla suddivisione in gruppi, sembrerebbe che l'applicazione sui soli livelli di concentrazione rispecchi di più la conformazione del territorio piemontese e la collocazione degli strumenti di misura in zone più o meno urbanizzate. Invece la *cluster analysis* conduce ad una "stabilità nei gruppi" tra i diversi inquinanti che evidenzia come le caratteristiche delle stazioni e delle loro zone di ubicazione pesino maggiormente rispetto ai valori di media annuale, deviazione standard e quantile. Pertanto, per ciascun inquinante, diminuisce l'importanza dei livelli di concentrazione in ogni serie storica e del suo andamento nel tempo.

I risultati delle tre analisi adottate rappresentano comunque un'utile informazione per chi, per

esempio, deve operare scelte di razionalizzazione della rete o indicare centraline rappresentative per analisi a livello nazionale.

Ringraziamenti

I nostri ringraziamenti vanno al CSI Piemonte per aver fornito i dati utilizzati in questo studio e per il supporto tecnico, e all'Ing. Giorgio Arduino della Direzione Ambiente della Regione Piemonte per la collaborazione ed i preziosi suggerimenti.

Riferimenti bibliografici

- [Aria-04] AA. VV. (2004), ‘Relazione annuale *Uno sguardo all’aria*’, Provincia di Torino.
- [Abr-03] Abraham C., Cornillon P.A., Matzner-Løber E., Molinari N. (2003) ‘Unsupervised curve clustering using B-splines’, *Scandinavian journal of statistics*, **30**, 581–595.
- [Ban-05] Bande S., Bertaccini P. Ghigo S., Ignaccolo R., Nicolis O. (2005), ‘Something about the Piemonte air - Che aria tira in Piemonte?’, *Poster presentato alla Graspa Conference 2005 - Statistics for Environmental Decisions (Bertinoro, 21–23 aprile 2005)* .
- [dBo-01] de Boor C. (2001) *A Practical Guide to Splines*, Springer-Verlag New York.
- [Fas-05] Fassò A., Nicolis O. (2005), ‘Space-Time Integration of Heterogeneous Networks in Air Quality Monitoring’, *Statistics and Environment, SIS Invited Papers* CLEUP, 265–276.
- [Gab-05] Gabusi V., Volta M. (2005) ‘A methodology for seasonal photochemical model simulation assessment’, *Journal of Environment and Pollution*, **21**, 11–21.
- [Gra-06] Gramsh E., Cereceda-Balic F., Oyola P., von Baer D. (2006) ‘Examination of pollution trends in Santiago de Chile with cluster analysis of PM₁₀ and Ozone data’, *Atmospheric environment*, **40**, 5464–5475.
- [Ion-00] Ionescu A., Candau Y., Mayer E., Colda I. (200), “Analytical determination and classification of pollutant concentration fields using air pollution monitoring network data - Methodology and application in the Paris area, during episodes with peak nitrogen dioxide levels”, *Environmental Modelling & Software*, **15**, 565–573.
- [Kau-90] Kaufman L., Rousseeuw P. J. (1990) *Finding Groups in Data. An Introduction to Cluster Analysis*, John Wiley & Sons.
- [Lav-96] Lavecchia C., Angelino E., Bedogni M., Brevetti E., Gualdi R., Lanzani G., Musitelli A., Valentini M. (1996) ‘The ozone patterns in the aerological basin of Milan (Italy)’, *Environmental Software*, **11**, 73–80.
- [Lud-95] Ludwig, F.L., J. Jiang and J. Chen (1995), ‘Classification of ozone and heather patterns associated with high ozone concentrations in the San Francisco and Monterey Bay areas’, *Atmospheric Environment*, **29**, 2915–2928.
- [Mig-94] Mignani S., Montanari A. (1994) *Appunti di Analisi Statistica Multivariata*, Società Editrice Esculapio.
- [Ram-02] Ramsay J. O., Silverman B. (2002) *Applied functional data analysis*, Springer-Verlag.
- [Ram-05] Ramsay J. O., Silverman B. (2005) *Functional data analysis*, Springer-Verlag.
- [Weg-83] Wegman E.J., Wright I.W. (1983) ‘Splines in statistics’, *Journal of the American Statistical Association*, **78**, 351–365.

SITI CONSULTATI:

- [1] <http://extranet.regione.piemonte.it>
- [2] <http://www.sistemapiemonte.it/ambiente/srqa/>
- [3] <http://ego.psych.mcgill.ca/misc/fda/>

RIFERIMENTI LEGISLATIVI:

- [APAT] Classificazione rivista dalle ARPA partecipanti sulla base delle LINEE GUIDA PER LA PREDISPOSIZIONE DELLE RETI DI QUALITÀ DELL'ARIA IN ITALIA APAT CTN_ ACE.
- [Ozone] Inquinamento da ozono nella valle Padana - Atti del convegno - Fondazione Lombardia per l'Ambiente, Regione Lombardia, Milano, 25-26 giugno 1997, N. 39.
- [615/66] LEGGE 13 luglio 1966, n. 615 Provvedimenti contro l'inquinamento atmosferico (G.U. n. 201 del 13 agosto 1966)
- [D.P.R.203/88] Attuazione delle direttive CEE numeri 80/779, 82/884, 84/360 e 85/203 concernenti norme in materia di qualità dell'aria, relativamente a specifici agenti inquinanti, e di inquinamento prodotto dagli impianti industriali, ai sensi dell'art.15 della legge 16 aprile 1987, numero 183 (G.U. n. 140 del 16 giugno 1988, S.O.)
- [96/62/CE] DIRETTIVA DEL CONSIGLIO 27 settembre 1996, n.96 - In materia di valutazione e di gestione della qualità dell'aria ambiente. (G.U.C.E n. L 296 del 21 novembre 1996).
- [99/30/CE] DIRETTIVA DEL CONSIGLIO 22 aprile 1999, n.30 - Direttiva del Consiglio concernente i valori limite di qualità dell'aria ambiente per il biossido di zolfo, il biossido di azoto, gli ossidi di azoto, le particelle e il piombo. (G.U.C.E. n. L 163 del 29 giugno 1999)
- [Dlg.351/99] DECRETO LEGISLATIVO 4 agosto 1999, n.351 - Attuazione della direttiva 96/62/CE in materia di valutazione e di gestione della qualità dell'aria ambiente. (G.U. n. 241 del 13 ottobre 1999)
- [Reg.43/00] LEGGE REGIONALE 7 aprile 2000, n.43 - Disposizioni per la tutela dell'ambiente in materia di inquinamento atmosferico. Prima attuazione del piano regionale per il risanamento e la tutela della qualità dell'aria. (Suppl. n. 2 al B.U. n. 15 del 12 aprile 2000)
- [2001/752] DECISIONE 2001/752/CE - Exchange of Information (EOI), Criteria for EUROAIRNET.
- [02/3/CE] DIRETTIVA DEL PARLAMENTO EUROPEO E DEL CONSIGLIO 12 febbraio 2002, n.3 - Direttiva del Parlamento Europeo e del Consiglio relativa all'ozono nell'aria. (G.U.C.E. n. L 67 del 9 marzo 2002)
- [DM.60/02] DECRETO MINISTERIALE 2 aprile 2002, n.60 - Recepimento della direttiva 1999/30/CE concernente i valori limite di qualità dell'aria ambiente per il biossido di zolfo, il biossido di azoto, gli ossidi di azoto, le particelle e il piombo e della direttiva 2000/69/CE relativa ai valori limite di qualità dell'aria ambiente per il benzene ed il monossido di carbonio. (Suppl. n. 77 alla G.U. n. 87 del 13 aprile 2002)
- [Dlg.183/04] DECRETO LEGISLATIVO 21 maggio 2004, n.183 - Attuazione della direttiva 2002/3/CE relativa all'ozono nell'aria. (Suppl. alla G.U. n. 171 del 23 luglio 2004)
- [EN12341] EN 12341 'Air quality - Determination of the PM₁₀ fraction of suspended particulate matter Reference method and field test procedure to demonstrate reference equivalence of measurement methods'
- [ISO7996] ISO 7996: 1985 - Ambient Air - Determination of the mass concentration of nitrogen oxides - Chemiluminescence Method
- [ISO13964] ISO FDIS 13964: 1998 Air quality - Determination of ozone in ambient air - Ultraviolet photometric method

Appendice: Tabelle PM₁₀ - NO₂ - O₃

Monitoring Stations	Mean	St. Dev.	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>						
14 BI Sturzo	35,27	21,02	85,70	Urban	Background	406,2
21 CN Alpini	32,91	21,25	75,60	Urban	Background	551,6
24 Saliceto	29,26	18,46	9,60	Rural	Background	388,5
51 Susa	30,17	18,63	30,40	Suburban	Traffic	494,0
62 Verbania	28,95	19,21	73,00	Urban	Background	201,4
<i>CLUSTER 2</i>						
63 Borgosesia	34,51	23,43	53,40	Urban	Background	345,2
36 Butt.Alta	42,89	25,62	29,20	Suburban	Background	331,0
6 C.Monferrato Negri	39,10	23,85	82,50	Suburban	Traffic	112,2
8 S.Scrivia	31,55	19,05	33,60	Suburban	Background	226,6
<i>CLUSTER 3</i>						
18 Alba	38,67	23,32	59,00	Urban	Background	164,1
11 AT D'Acquisto	45,52	29,95	66,70	Urban	Background	149,9
12 Butt.d'Asti	46,58	28,40	11,70	Rural	Background	280,4
37 Carmagnola	45,12	26,08	42,30	Suburban	Traffic	238,8
7 N.Ligure	52,86	29,62	59,70	Urban	Traffic	201,7
46 Pinerolo	38,00	24,50	78,00	Urban	Traffic	374,9
<i>CLUSTER 4</i>						
2 AL N.Orti	48,77	27,92	76,20	Urban	Traffic	91,7
35 Borgaro	45,51	27,77	35,80	Suburban	Background	257,5
20 Bra	41,87	25,65	68,50	Urban	Background	274,9
30 NO Roma	50,77	31,25	100,00	Urban	Traffic	156,3
56 TO Consolata	58,29	33,47	96,40	Urban	Traffic	243,6
9 Tortona	48,08	27,21	85,40	Urban	Background	118,5
65 VC Gastaldi	53,60	30,81	77,90	Urban	Traffic	131,0
<i>CLUSTER 5</i>						
52 TO Grassi	69,46	36,62	98,90	Urban	Traffic	245,0

Tabella 7: Gruppi ottenuti dall'applicazione della *functional cluster analysis* in relazione con ulteriori variabili esogene relative al PM₁₀ nel 2004

Monitoring Stations	Mean	St. Dev.	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>						
63 Borgosesia	34,51	23,43	53,40	Urban	Background	345,2
36 Butt. Alta	42,89	25,62	29,20	Suburban	Background	331,0
12 Butt. d' Asti	46,58	28,40	11,70	Rural	Background	280,4
24 Saliceto	29,26	18,46	9,60	Rural	Background	388,5
8 S. Scrivia	31,55	19,05	33,60	Suburban	Background	226,6
51 Susa	30,17	18,63	30,40	Suburban	Traffic	494,0
<i>CLUSTER 2</i>						
14 BI Sturzo	35,27	21,02	85,70	Urban	Background	406,2
21 CN Alpini	32,91	21,25	75,60	Urban	Background	551,6
46 Pinerolo	38,00	24,50	78,00	Urban	Traffic	374,9
62 Verbania	28,95	19,21	73,00	Urban	Background	201,4
<i>CLUSTER 3</i>						
2 AL N. Orti	48,77	27,92	76,20	Urban	Traffic	91,7
18 Alba	38,67	23,32	59,00	Urban	Background	164,1
11 AT D'Acquisto	45,52	29,95	66,70	Urban	Background	149,9
35 Borgaro	45,51	27,77	35,80	Suburban	Background	257,5
20 Bra	41,87	25,65	68,50	Urban	Background	274,9
37 Carmagnola	45,12	26,08	42,30	Suburban	Traffic	238,8
6 C. Monferrato Negri	39,10	23,85	82,50	Suburban	Traffic	112,2
7 N. Ligure	52,86	29,62	59,70	Urban	Traffic	201,7
<i>CLUSTER 4</i>						
30 NO Roma	50,77	31,25	100,00	Urban	Traffic	156,3
56 TO Consolata	58,29	33,47	96,40	Urban	Traffic	243,6
9 Tortona	48,08	27,21	85,40	Urban	Background	118,5
65 VC Gastaldi	53,60	30,81	77,90	Urban	Traffic	131
<i>CLUSTER 5</i>						
52 TO Grassi	69,46	36,62	98,90	Urban	Traffic	245,0

Tabella 8: Gruppi ottenuti dall'applicazione della *functional cluster analysis* 'modificata' al PM₁₀ nel 2004

Monitoring Stations	Mean	St. Dev.	Quant904	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>							
18 Alba	38,67	23,32	72,00	59,00	Urban	Background	164,1
14 BI Sturzo	35,27	21,02	62,11	85,70	Urban	Background	406,2
63 Borgosesia	34,51	23,43	62,00	53,40	Urban	Background	345,2
6 C. Monferrato Negri	39,10	23,85	74,00	82,50	Suburban	Traffic	112,2
21 CN Alpini	32,91	21,25	58,00	75,60	Urban	Background	551,6
46 Pinerolo	38,00	24,50	68,32	78,00	Urban	Traffic	374,9
62 Verbania	28,95	19,21	52,00	73,00	Urban	Background	201,4
<i>CLUSTER 2</i>							
35 Borgaro	45,51	27,77	81,00	35,80	Suburban	Background	257,5
36 Butt. Alta	42,89	25,62	77,00	29,20	Suburban	Background	331,0
12 Butt. d' Asti	46,58	28,40	81,00	11,70	Rural	Background	280,4
37 Carmagnola	45,12	26,08	87,26	42,30	Suburban	Traffic	238,8
24 Saliceto	29,26	18,46	56,00	9,60	Rural	Background	388,5
8 S. Scrivia	31,55	19,05	56,00	33,60	Suburban	Background	226,6
51 Susa	30,17	18,63	55,00	30,40	Suburban	Traffic	494,0
<i>CLUSTER 3</i>							
2 AL N. Orti	48,77	27,92	89,00	76,20	Urban	Traffic	91,7
11 AT D'Acquisto	45,52	29,95	90,00	66,70	Urban	Background	149,9
20 Bra	41,87	25,65	99,78	68,50	Urban	Background	274,9
7 N. Ligure	52,86	29,62	85,00	59,70	Urban	Traffic	201,7
9 Tortona	48,08	27,21	86,53	85,40	Urban	Background	118,5
65 VC Gastaldi	53,60	30,81	97,00	77,90	Urban	Traffic	131,0
<i>CLUSTER 4</i>							
30 NO Roma	50,77	31,25	95,59	100,00	Urban	Traffic	156,3
56 TO Consolata	58,29	33,47	121,00	96,40	Urban	Traffic	243,6
52 TO Grassi	69,46	36,62	104,46	98,90	Urban	Traffic	245,0

Tabella 9: Gruppi ottenuti dall'applicazione della *cluster analysis* alla media annuale, deviazione standard, al quantile di ordine 0.904 e ad alcune variabili esogene relative al PM₁₀ nel 2004

Monitoring Stations	Mean	St. Dev.	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>						
11 AT D'Acquisto	28,40	20,89	66,70	Urban	Background	149,9
1 A.Terme	25,17	17,77	52,50	Urban	Background	157,8
63 Borgosesia	27,11	18,11	53,40	Urban	Background	345,2
12 Butt.d'Asti	24,46	20,58	11,70	Rural	Background	280,4
64 Caresanablot	30,88	19,14	26,60	Suburban	Traffic	134,0
5 C.Monferrato Castello	19,46	16,68	68,10	Urban	Background	118,3
40 Ciriè	31,00	22,19	56,50	Urban	Background	340,5
15 Cossato	29,46	19,65	71,50	Urban	Background	271,7
60 Domodossola	27,69	21,79	82,40	Urban	Background	292,8
22 Fossano	33,98	20,14	59,70	Urban	Traffic	371,3
23 Mondovì	28,06	15,54	63,10	Urban	Background	419,6
44 NichelinoAEM	21,80	14,44	63,20	Urban	Background	231,0
46 Pinerolo	30,94	19,84	78,00	Urban	Traffic	374,9
16 Ponzone	20,15	13,51	36,10	Suburban	Background	487,2
24 Saliceto	18,84	12,65	9,60	Rural	Background	388,5
17 Verrone	26,46	18,25	25,60	Suburban	Background	279,0
<i>CLUSTER 2</i>						
4 AL Libertà	49,69	24,01	96,00	Urban	Traffic	95,2
18 Alba	33,56	16,57	59,00	Urban	Background	164,1
25 Arona	32,83	21,34	69,40	Urban	Background	209,2
19 B.S.Dalmazzo	44,18	25,44	55,00	Urban	Traffic	631,6
26 Borgomanero	39,74	20,80	77,90	Urban	Traffic	313,2
20 Bra	36,04	20,54	68,50	Urban	Background	274,9
21 CN Alpini	35,81	21,62	75,60	Urban	Background	551,6
61 Omegna	39,42	20,72	57,50	Urban	Traffic	305,1
49 Sangone	32,05	18,45	80,10	Urban	Background	222,3
51 Susa	34,02	20,61	30,40	Suburban	Traffic	494,0
<i>CLUSTER 3</i>						
2 AL N.Orti	39,85	18,11	76,20	Urban	Traffic	91,7
33 Alpignano	38,62	24,87	55,40	Urban	Background	326,0
13 BI Lamarmora	46,06	22,55	79,30	Urban	Traffic	419,9
14 BI Sturzo	33,29	24,18	85,70	Urban	Background	406,2
35 Borgaro	41,86	27,13	35,80	Suburban	Background	257,5
27 Cameri	40,43	22,12	53,10	Urban	Background	159,3
38 Chieri	43,84	27,88	49,90	Suburban	Traffic	302,2
42 Ivrea	47,34	28,02	67,30	Urban	Traffic	242,0
29 NO Bovio	41,84	27,47	92,80	Urban	Background	146,0
7 N.Ligure	43,77	22,31	59,70	Urban	Traffic	201,7
45 Orbassano	44,47	24,67	39,60	Suburban	Background	269,1
9 Tortona	48,07	19,47	85,40	Urban	Background	118,5
32 Trecate	40,37	24,36	61,10	Urban	Background	135,0
65 VC Gastaldi	43,93	23,54	77,90	Urban	Traffic	131,0
10 Valenza Po	34,68	21,66	59,20	Urban	Background	126,0
59 Vinovo	37,52	20,67	42,90	Suburban	Background	234,5
<i>CLUSTER 4</i>						
34 Beinasco	54,66	35,12	64,30	Urban	Background	268,4
41 Grugliasco	49,29	31,31	96,30	Urban	Traffic	295,0
28 NO Leonardi	55,47	27,45	95,30	Urban	Traffic	149,0
43 Nichelino	59,39	32,95	55,00	Urban	Traffic	228,3
48 Rivoli	60,35	30,22	73,00	Urban	Traffic	315,0
50 Settimo	57,45	32,55	67,30	Urban	Traffic	201,2
53 TO Lingotto	51,32	31,03	93,30	Urban	Background	243,6
<i>CLUSTER 5</i>						
3 AL D'Annunzio	68,84	27,68	100,00	Urban	Traffic	95,2
54 TO Rebaudengo	84,70	33,82	84,80	Urban	Traffic	233,0
55 TO Rivoli	80,78	31,59	99,80	Urban	Traffic	257,0
56 TO Consolata	71,92	28,64	96,40	Urban	Traffic	243,6
57 TO Gaidano	70,82	32,28	92,90	Urban	Traffic	259,9
58 TO M.Cristina	74,77	27,80	81,50	Urban	Traffic	234,9

Tabella 10: Gruppi ottenuti dall'applicazione della *functional cluster analysis* in relazione con ulteriori variabili esogene relative all'NO₂ nel 2004

Monitoring Stations	Mean	St. Dev.	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>						
12 Butt.d'Asti	24,46	20,58	11,70	Rural	Background	280,4
64 Caresanablot	30,88	19,14	26,60	Suburban	Traffic	134,0
16 Ponzzone	20,15	13,51	36,10	Suburban	Background	487,2
24 Saliceto	18,84	12,65	9,60	Rural	Background	388,5
51 Susa	34,02	20,61	30,40	Suburban	Traffic	494,0
17 Verrone	26,46	18,25	25,60	Suburban	Background	279,0
<i>CLUSTER 2</i>						
11 AT D'Acquisto	28,40	20,89	66,70	Urban	Background	149,9
1 A.Terme	25,17	17,77	52,50	Urban	Background	157,8
18 Alba	33,56	16,57	59,00	Urban	Background	164,1
25 Arona	32,83	21,34	69,40	Urban	Background	209,2
63 Borgosesia	27,11	18,11	53,40	Urban	Background	345,2
20 Bra	36,04	20,54	68,50	Urban	Background	274,9
21 CN Alpini	35,81	21,62	75,60	Urban	Background	551,6
5 C.Monferrato Castello	19,46	16,68	68,10	Urban	Background	118,3
40 Ciriè	31,00	22,19	56,50	Urban	Background	340,5
15 Cossato	29,46	19,65	71,50	Urban	Background	271,7
60 Domodossola	27,69	21,79	82,40	Urban	Background	292,8
22 Fossano	33,98	20,14	59,70	Urban	Traffic	371,3
23 Mondovì	28,06	15,54	63,10	Urban	Background	419,6
44 NichelinoAEM	21,80	14,44	63,20	Urban	Background	231,0
46 Pinerolo	30,94	19,84	78,00	Urban	Traffic	374,9
49 Sangone	32,05	18,45	80,10	Urban	Background	222,3
<i>CLUSTER 3</i>						
2 AL N.Orti	39,85	18,11	76,20	Urban	Traffic	91,7
4 AL Libertà	49,69	24,01	96,00	Urban	Traffic	95,2
33 Alpignano	38,62	24,87	55,40	Urban	Background	326,0
13 BI Lamarmora	46,06	22,55	79,30	Urban	Traffic	419,9
14 BI Sturzo	33,29	24,18	85,70	Urban	Background	406,2
35 Borgaro	41,86	27,13	35,80	Suburban	Background	257,5
19 B.S.Dalmazzo	44,18	25,44	55,00	Urban	Traffic	631,6
26 Borgomanero	39,74	20,80	77,90	Urban	Traffic	313,2
27 Cameri	40,43	22,12	53,10	Urban	Background	159,3
38 Chieri	43,84	27,88	49,90	Suburban	Traffic	302,2
42 Ivrea	47,34	28,02	67,30	Urban	Traffic	242,0
29 NO Bovio	41,84	27,47	92,80	Urban	Background	146,0
7 N.Ligure	43,77	22,31	59,70	Urban	Traffic	201,7
61 Omegna	39,42	20,72	57,50	Urban	Traffic	305,1
45 Orbassano	44,47	24,67	39,60	Suburban	Background	269,1
9 Tortona	48,07	19,47	85,40	Urban	Background	118,5
32 Trecate	40,37	24,36	61,10	Urban	Background	135,0
65 VC Gastaldi	43,93	23,54	77,90	Urban	Traffic	131,0
10 Valenza Po	34,68	21,66	59,20	Urban	Background	126,0
59 Vinovo	37,52	20,67	42,90	Suburban	Background	234,5
<i>CLUSTER 4</i>						
34 Beinasco	54,66	35,12	64,30	Urban	Background	268,4
41 Grugliasco	49,29	31,31	96,30	Urban	Traffic	295,0
28 NO Leonardi	55,47	27,45	95,30	Urban	Traffic	149,0
43 Nichelino	59,39	32,95	55,00	Urban	Traffic	228,3
48 Rivoli	60,35	30,22	73,00	Urban	Traffic	315,0
50 Settimo	57,45	32,55	67,30	Urban	Traffic	201,2
53 TO Lingotto	51,32	31,03	93,30	Urban	Background	243,6
<i>CLUSTER 5</i>						
3 AL D'Annunzio	68,84	27,68	100,00	Urban	Traffic	95,2
54 TO Rebaudengo	84,70	33,82	84,80	Urban	Traffic	233,0
55 TO Rivoli	80,78	31,59	99,80	Urban	Traffic	257,0
56 TO Consolata	71,92	28,64	96,40	Urban	Traffic	243,6
57 TO Gaidano	70,82	32,28	92,90	Urban	Traffic	259,9
58 TO M.Cristina	74,77	27,80	81,50	Urban	Traffic	234,9

Tabella 11: Gruppi ottenuti dall'applicazione della *functional cluster analysis* 'modificata' all'NO₂ nel 2004

Monitoring Stations	Mean	St. Dev.	Quant998	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>							
12 Butt.d'Asti	24,46	20,58	102,00	11,70	Rural	Background	280,4
64 Caresanablot	30,88	19,14	112,42	26,60	Suburban	Traffic	134,0
16 Ponzzone	20,15	13,51	73,00	36,10	Suburban	Background	487,2
24 Saliceto	18,84	12,65	75,89	9,60	Rural	Background	388,5
51 Susa	34,02	20,61	118,51	30,40	Suburban	Traffic	494,0
17 Verrone	26,46	18,25	91,00	25,60	Suburban	Background	279,0
<i>CLUSTER 2</i>							
2 AL N.Orti	39,85	18,11	108,00	76,20	Urban	Traffic	91,7
11 AT D'Acquisto	28,40	20,89	112,00	66,70	Urban	Background	149,9
1 A.Terme	25,17	17,77	90,00	52,50	Urban	Background	157,8
18 Alba	33,56	16,57	96,00	59,00	Urban	Background	164,1
13 BI Lamarmora	46,06	22,55	117,00	79,30	Urban	Traffic	419,9
63 Borgosesia	27,11	18,11	98,73	53,40	Urban	Background	345,2
21 CN Alpini	35,81	21,62	111,00	75,60	Urban	Background	551,6
5 C.Monferrato Castello	19,46	16,68	90,39	68,10	Urban	Background	118,3
15 Cossato	29,46	19,65	99,62	71,50	Urban	Background	271,7
60 Domodossola	27,69	21,79	112,00	82,40	Urban	Background	292,8
23 Mondovì	28,06	15,54	85,76	63,10	Urban	Background	419,6
44 NichelinoAEM	21,80	14,44	76,00	63,20	Urban	Background	231,0
61 Omegna	39,42	20,72	109,00	57,50	Urban	Traffic	305,1
46 Pinerolo	30,94	19,84	114,00	78,00	Urban	Traffic	374,9
49 Sangone	32,05	18,45	96,37	80,10	Urban	Background	222,3
9 Tortona	48,07	19,47	108,00	85,40	Urban	Background	118,5
10 Valenza Po	34,68	21,66	107,00	59,20	Urban	Background	126,0
<i>CLUSTER 3</i>							
33 Alpignano	38,62	24,87	128,00	55,40	Urban	Background	326,0
25 Arona	32,83	21,34	128,00	69,40	Urban	Background	209,2
14 BI Sturzo	33,29	24,18	127,50	85,70	Urban	Background	406,2
26 Borgomanero	39,74	20,80	130,32	77,90	Urban	Traffic	313,2
20 Bra	36,04	20,54	122,92	68,50	Urban	Background	274,9
27 Cameri	40,43	22,12	132,63	53,10	Urban	Background	159,3
38 Chieri	43,84	27,88	149,28	49,90	Suburban	Traffic	302,2
40 Ciriè	31,00	22,19	125,00	56,50	Urban	Background	340,5
22 Fossano	33,98	20,14	116,30	59,70	Urban	Traffic	371,3
7 N.Ligure	43,77	22,31	120,44	59,70	Urban	Traffic	201,7
45 Orbassano	44,47	24,67	136,94	39,60	Suburban	Background	269,1
32 Trecate	40,37	24,36	136,00	61,10	Urban	Background	135,0
59 Vinovo	37,52	20,67	118,00	42,90	Suburban	Background	234,5
<i>CLUSTER 4</i>							
4 AL Libertà	49,69	24,01	154,39	96,00	Urban	Traffic	95,2
34 Beinasco	54,66	35,12	182,00	64,30	Urban	Background	268,4
35 Borgaro	41,86	27,13	166,14	35,80	Suburban	Background	257,5
19 B.S.Dalmazzo	44,18	25,44	151,81	55,00	Urban	Traffic	631,6
42 Ivrea	47,34	28,02	164,89	67,30	Urban	Traffic	242,0
29 NO Bovio	41,84	27,47	164,18	92,80	Urban	Background	146,0
53 TO Lingotto	51,32	31,03	171,00	93,30	Urban	Background	243,6
65 VC Gastaldi	43,93	23,54	153,50	77,90	Urban	Traffic	131,0
<i>CLUSTER 5</i>							
3 AL D'Annunzio	68,84	27,68	201,18	100,00	Urban	Traffic	95,2
41 Grugliasco	49,29	31,31	186,70	96,30	Urban	Traffic	295,0
28 NO Leonardi	55,47	27,45	182,95	95,30	Urban	Traffic	149,0
43 Nichelino	59,39	32,95	203,19	55,00	Urban	Traffic	228,3
48 Rivoli	60,35	30,22	185,47	73,00	Urban	Traffic	315,0
50 Settimo	57,45	32,55	195,38	67,30	Urban	Traffic	201,2
54 TO Rebaudengo	84,70	33,82	247,00	84,80	Urban	Traffic	233,0
55 TO Rivoli	80,78	31,59	220,05	99,80	Urban	Traffic	257,0
56 TO Consolata	71,92	28,64	183,00	96,40	Urban	Traffic	243,6
57 TO Gaidano	70,82	32,28	218,53	92,90	Urban	Traffic	259,9
58 TO M.Cristina	74,77	27,80	178,00	81,50	Urban	Traffic	234,9

Tabella 12: Gruppi ottenuti dall'applicazione della *cluster analysis* alla media annuale, deviazione standard, al quantile di ordine 0.998 e ad alcune variabili esogene relative a NO₂ nel 2004

Monitoring Stations	Mean	St. Dev.	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>						
2 AL N.Orti	42,46	41,87	76,20	Urban	Traffic	91,7
18 Alba	49,63	43,62	59,00	Urban	Background	164,1
33 Alpignano	46,27	37,43	55,40	Urban	Background	326,0
35 Borgaro	45,87	40,38	35,80	Suburban	Background	257,5
63 Borgosesia	49,30	38,81	53,40	Urban	Background	345,2
38 Chieri	44,55	39,14	49,90	Suburban	Traffic	302,2
39 Chivasso	37,57	41,95	34,20	Suburban	Traffic	183,7
61 Omegna	44,42	33,40	57,50	Urban	Traffic	305,1
16 Ponzzone	42,79	39,68	36,10	Suburban	Background	487,2
53 TO Lingotto	43,42	43,80	93,30	Urban	Background	243,6
65 VC Gastaldi	44,61	38,88	77,90	Urban	Traffic	131,0
59 Vinovo	51,26	42,06	42,90	Suburban	Background	234,5
<i>CLUSTER 2</i>						
1 A.Terme	55,72	45,83	52,50	Urban	Background	157,8
25 Arona	52,08	44,07	69,40	Urban	Background	209,2
11 AT D'Acquisto	52,68	43,95	66,70	Urban	Background	149,9
14 BI Sturzo	57,13	42,11	85,70	Urban	Background	406,2
15 Cossato	52,71	43,55	71,50	Urban	Background	271,7
31 NO Verdi	53,47	49,15	97,50	Urban	Background	154,5
17 Verrone	42,91	44,93	25,60	Suburban	Background	279,0
<i>CLUSTER 3</i>						
12 Butt.d'Asti	63,35	45,08	11,70	Rural	Background	280,4
21 CN Alpini	58,27	41,12	75,60	Urban	Background	551,6
46 Pinerolo	52,14	34,46	78,00	Urban	Traffic	374,9
47 P.Torinese	76,67	39,95	16,40	Rural	Background	612,4
24 Saliceto	58,72	38,99	9,60	Rural	Background	388,5
51 Susa	59,81	41,98	30,40	Suburban	Traffic	494,0

Tabella 13: Gruppi ottenuti dall'applicazione della *functional cluster analysis* in relazione con ulteriori variabili esogene relative all'O₃ nel 2004

Monitoring Stations	Mean	St. Dev.	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>						
2 AL N.Orti	42,46	41,87	76,20	Urban	Traffic	91,7
18 Alba	49,63	43,62	59,00	Urban	Background	164,1
33 Alpignano	46,27	37,43	55,40	Urban	Background	326,0
35 Borgaro	45,87	40,38	35,80	Suburban	Background	257,5
63 Borgosesia	49,30	38,81	53,40	Urban	Background	345,2
38 Chieri	44,55	39,14	49,90	Suburban	Traffic	302,2
39 Chivasso	37,57	41,95	34,20	Suburban	Traffic	183,7
61 Omegna	44,42	33,40	57,50	Urban	Traffic	305,1
16 Ponzzone	42,79	39,68	36,10	Suburban	Background	487,2
53 TO Lingotto	43,42	43,80	93,30	Urban	Background	243,6
65 VC Gastaldi	44,61	38,88	77,90	Urban	Traffic	131,0
17 Verrone	42,91	44,93	25,60	Suburban	Background	279,0
59 Vinovo	51,26	42,06	42,90	Suburban	Background	234,5
<i>CLUSTER 2</i>						
1 A.Terme	55,72	45,83	52,50	Urban	Background	157,8
25 Arona	52,08	44,07	69,40	Urban	Background	209,2
11 AT D'Acquisto	52,68	43,95	66,70	Urban	Background	149,9
14 BI Sturzo	57,13	42,11	85,70	Urban	Background	406,2
21 CN Alpini	58,27	41,12	75,60	Urban	Background	551,6
15 Cossato	52,71	43,55	71,50	Urban	Background	271,7
31 NO Verdi	53,47	49,15	97,50	Urban	Background	154,5
46 Pinerolo	52,14	34,46	78,00	Urban	Traffic	374,9
<i>CLUSTER 3</i>						
12 Butt.d'Asti	63,35	45,08	11,70	Rural	Background	280,4
47 P.Torinese	76,67	39,95	16,40	Rural	Background	612,4
24 Saliceto	58,72	38,99	9,60	Rural	Background	388,5
51 Susa	59,81	41,98	30,40	Suburban	Traffic	494,0

Tabella 14: Gruppi ottenuti dall'applicazione della *functional cluster analysis* 'modificata' all'O₃ nel 2004

Monitoring Stations	Mean	St. Dev.	Quant98	Built %	Zone Type	Station Type	Altitude
<i>CLUSTER 1</i>							
1 A.Terme	55,72	45,83	159,00	52,50	Urban	Background	157,8
18 Alba	49,63	43,62	154,00	59,00	Urban	Background	164,1
33 Alpignano	46,27	37,43	139,00	55,40	Urban	Background	326,0
35 Borgaro	45,87	40,38	146,00	35,80	Suburban	Background	257,5
63 Borgosesia	49,30	38,81	146,00	53,40	Urban	Background	345,2
38 Chieri	44,55	39,14	136,00	49,90	Suburban	Traffic	302,2
39 Chivasso	37,57	41,95	142,00	34,20	Suburban	Traffic	183,7
61 Omegna	44,42	33,40	125,40	57,50	Urban	Traffic	305,1
16 Ponzzone	42,79	39,68	144,00	36,10	Suburban	Background	487,2
59 Vinovo	51,26	42,06	159,00	42,90	Suburban	Background	234,5
<i>CLUSTER 2</i>							
2 AL N.Orti	42,46	41,87	150,00	76,20	Urban	Traffic	91,7
11 AT D'Acquisto	52,68	43,95	159,00	66,70	Urban	Background	149,9
25 Arona	52,08	44,07	163,00	69,40	Urban	Background	209,2
14 BI Sturzo	57,13	42,11	151,00	85,70	Urban	Background	406,2
21 CN Alpini	58,27	41,12	155,48	75,60	Urban	Background	551,6
15 Cossato	52,71	43,55	160,58	71,50	Urban	Background	271,7
31 NO Verdi	53,47	49,15	167,00	97,50	Urban	Background	154,5
46 Pinerolo	52,14	34,46	130,80	78,00	Urban	Traffic	374,9
53 TO Lingotto	43,42	43,80	153,00	93,30	Urban	Background	243,6
65 VC Gataldi	44,61	38,88	138,00	77,90	Urban	Traffic	131,0
<i>CLUSTER 3</i>							
12 Butt.d'Asti	63,35	45,08	166,00	11,70	Rural	Background	280,4
47 P.Torinese	76,67	39,95	162,86	16,40	Rural	Background	612,4
24 Saliceto	58,72	38,99	141,00	9,60	Rural	Background	388,5
51 Susa	59,81	41,98	156,00	30,40	Suburban	Traffic	494,0
17 Verrone	42,91	44,93	151,00	25,60	Suburban	Background	279,0

Tabella 15: Gruppi ottenuti dall'applicazione della *cluster analysis* alla media annuale, deviazione standard, al quantile di ordine 0.98 e ad alcune variabili esogene relative all'O₃ nel 2004