

Michele Scagliarini

Lo studio dell'inquinamento atmosferico urbano  
da monossido di carbonio nella città di Bologna  
mediante un modello autoregressivo vettoriale

Serie Ricerche 1997, n.2



stituto di Scienze Statistiche "Paolo Fortunati"  
Università degli studi di Bologna

Lavoro svolto nell'ambito della ricerca "Metodi statistici per il controllo dell'ambiente e la valutazione del rischio ambientale" MURST 40% fondi 1995.

Le elaborazioni dei dati S.A.R.A., forniti dal Comune di Bologna, che compaiono nel lavoro sono eseguite presso il Dipartimento di Scienze Statistiche dell'Università di Bologna sotto la supervisione della Prof.ssa Daniela Cocchi. Sono da intendersi ad uso interno e non vanno confuse con quelle previste dalla legge in materia.

1. Introduzione	pag.5
2. Il modello	pag.6
3. La stima dei parametri	pag.12
3.1 <i>Analisi dei residui</i>	pag.12
3.2 <i>Analisi delle funzioni di autocorrelazione incrociata</i>	pag.18
3.3 <i>Analisi dei risultati</i>	pag.22
3.4 <i>Conclusioni</i>	pag.24
4. Le previsioni	pag.24
4.1 <i>Analisi degli errori di previsione</i>	pag.27
4.2 <i>Considerazioni sulle previsioni</i>	pag.30
Riferimenti Bibliografici	pag.31

Finito di stampare nel mese di Febbraio 1997  
presso le Officine Grafiche TECNOPRINT S.N.C.  
Via del Legatore 3, Bologna.

## 1. Introduzione

La raggiunta consapevolezza dei danni causati dall'inquinamento urbano sulla salute e sull'ambiente ha fatto maturare la coscienza della necessità di una concreta politica di controllo della qualità dell'aria. In quest'ottica si inquadra l'installazione, di reti di rilevazione degli inquinanti al fine di segnalare la presenza di situazioni critiche e di fornire una base di dati per la *previsione* delle condizioni dannose per la salute. Il problema della modellazione dell'inquinamento atmosferico urbano per scopi interpretativi e previsivi è stato affrontato mediante l'impiego di modelli stocastici di tipo autoregressivo-media mobile con ingressi esogeni, perché è possibile introdurre informazioni aggiuntive relative alle condizioni meteorologiche delle zone di interesse. Ciò trova giustificazione nel fatto che esistono dei collegamenti tra i valori di queste ultime ed i livelli delle concentrazioni delle sostanze inquinanti (*Finzi, 1991*). Nel presente studio, dedicato all'analisi dell'inquinamento da *monossido di carbonio* (CO)<sup>1</sup>, si propone l'impiego di una generalizzazione, in ambito multivariato, dei modelli precedenti: i modelli autoregressivi vettoriali con ingressi esogeni (VARX). Le concentrazioni dell'inquinante sono infatti misurate contemporaneamente ed a intervalli regolari da più stazioni di rilevamento ubicate nel territorio urbano. Si dispone così, ad ogni istante, di un vettore di osservazioni che sintetizza un'informazione di tipo "spaziale"; una modellazione che riesca a non perdere questa "fonte" d'informazione può quindi portare ad un miglioramento in termini di capacità esplicativa, pur mantenendo una struttura sufficientemente semplice. L'aspetto che si desidera porre in rilievo è che l'uso di una modellazione di tipo multivariato permette di riconoscere le eventuali dipendenze spazio-temporali tra le rilevazioni delle centraline fornendo risultati che possono essere utilizzati sia per avere delle informazioni sulla struttura del traffico, sia per migliorare la capacità esplicativa e previsiva del modello. In particolare, è specificato un modello autoregressivo vettoriale per le concentrazioni medie giornaliere del monossido di carbonio che, oltre alla componente autoregressiva, comprende: variabili meteorologiche esogene (temperatura, umidità, precipitazione e velocità del vento) e variabili deterministiche per cogliere le eventuali componenti stagionali o cicliche.

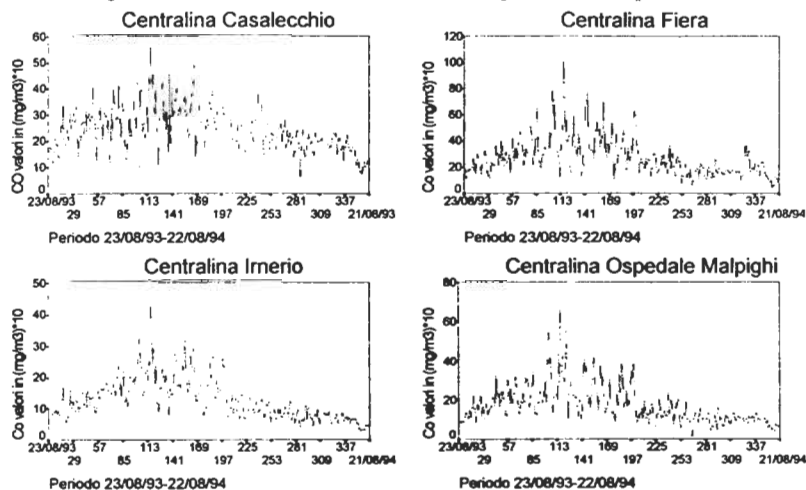
---

<sup>1</sup> Il monossido di carbonio è un inquinante la cui sorgente principale (91,4%) è costituita dagli autoveicoli a motore.

## 2. Il modello

Il periodo campionario di riferimento per la stima del modello copre un anno di calendario: 365 (N = 365) giorni dal 23/8/93 al 22/8/94; le centraline<sup>2</sup> interessate sono: Casalecchio, Fiera, Irnerio e Ospedale Malpighi. Nella fase di stima, per ovviare al problema dell'eteroschedasticità dei dati originali (Fig.1) si sono utilizzate le trasformazioni logaritmiche<sup>3</sup> delle variabili.

Fig. 1 Serie dei valori medi del CO nel periodo campionario



L'espressione generale del modello adottato è la seguente:

$$(2.1) \quad \mathbf{A(L)}\mathbf{y}_t = \Phi \mathbf{u}_t + \Psi \mathbf{s}_t + \Theta \mathbf{m}_t + \mathbf{e}_t$$

con  $E(\mathbf{e}_t) = 0$ ,  $E(\mathbf{e}_t \mathbf{e}_t^T) = \Lambda$ ,  $E(\mathbf{e}_t \mathbf{e}_{t-s}^T) = 0 \quad \forall s \neq 0$ ;

$\mathbf{y}_t$  è un vettore<sup>4</sup> 4x1 che contiene i logaritmi delle concentrazioni medie giornaliere del giorno t;

<sup>2</sup> La rete di monitoraggio S.A.R.A. del comune di Bologna è costituita da nove centraline per le elaborazioni ne sono considerate solo quattro.

<sup>3</sup> La trasformazione logaritmica non crea particolari problemi essendo i dati originali tutti positivi.

<sup>4</sup>  $\mathbf{y}_t = [y_{jt}]$  j=1 indica la centralina CASALECCHIO, j=2 indica la centralina FIERA, j=3 indica la centralina IRNERIO, j=4 indica la centralina OSPEDALE MALPIGHI

$\mathbf{A(L)} = \mathbf{I} - \mathbf{A}_1 \mathbf{L} - \dots - \mathbf{A}_k \mathbf{L}^k$  è il polinomio a coefficienti matriciali nell'operatore ritardo L che definisce la componente autoregressiva;

$\mathbf{m}_t$  è il vettore delle variabili meteo<sup>5</sup> con  $\Theta$  matrice 4x4 dei parametri relativi a questa componente;

$\mathbf{u}_t$  e  $\mathbf{s}_t$  sono i vettori delle variabili deterministiche utilizzate per cogliere la ciclicità annuale e settimanale del fenomeno.

Isolando  $\mathbf{y}_t$  dalle componenti autoregressive la (2.1) diventa

$$(2.2) \quad \mathbf{y}_t = \sum_{i=1}^k \mathbf{A}_i \mathbf{y}_{t-i} + \Phi \mathbf{u}_t + \Psi \mathbf{s}_t + \Theta \mathbf{m}_t + \mathbf{e}_t$$

ed esplicitando l'equazione per il generico j-esimo elemento  $y_{jt}$  di  $\mathbf{y}_t$  si ha:

$$(2.3) \quad y_{jt} = \sum_{i=1}^k a_{ji} y_{t-i} + \phi_j u_t + \psi_j s_t + \theta_j m_t + e_{jt}$$

con  $E(e_{jt}) = 0$ ,  $E(e_{jt}^2) = \sigma^2$ ,  $E(e_{jt} e_{j,t-i}) = 0 \quad \forall i \neq 0$ .

L'equazione (2.3) permette di descrivere come sono state specificate le componenti deterministiche  $\mathbf{u}_t$  e  $\mathbf{s}_t$ . Riguardo alla stagionalità settimanale i valori medi giornalieri del CO hanno un andamento molto simile, legato al ciclo del traffico nella settimana lavorativa. Per modellare questa componente si sono impiegate sei variabili dummy<sup>6</sup> D1<sub>t</sub> (per lunedì), ..., D6<sub>t</sub> (per sabato) che valutano l'effetto differenziale di ogni giorno non festivo rispetto alla domenica: si è scelto di considerare l'effetto in questione fisso sull'intero periodo campionario in quanto si ritiene che le differenze, in termini di inquinamento, tra i giorni feriali e la domenica si possano considerare costanti nei vari mesi dell'anno.  $\mathbf{s}_t$  è quindi un vettore 6x1 formato dalle variabili D1<sub>t</sub>, ..., D6<sub>t</sub> definite come segue: D<sub>i</sub> = 1 se t = i + 7\*n dove n = 1, 2, ...; 0 altrimenti (i = 1, ..., 6). Per cogliere il ciclo annuale si sono utilizzate delle funzioni deterministiche del tempo polinomiali a tratti (spline functions). Le quattro serie presentano infatti un profilo pressoché simile e nel periodo campionario considerato sono evidenti tre situazioni chiaramente distinte: una prima fase caratterizzata da un trend positivo; una

<sup>5</sup>  $\mathbf{m}_t = [m_{it}]$  i=1 indica la temperatura media giornaliera, i=2 indica l'umidità relativa, i=3 indica la precipitazione totale giornaliera, i=4 indica la velocità media giornaliera del vento.

<sup>6</sup> Le variabili sono le stesse per tutte e quattro le centraline.

seconda fase dove le concentrazioni medie giornaliere presentano un trend discendente; e una terza situazione dove i valori si mantengono relativamente costanti. In pratica sono stati identificati in modo preciso i seguenti sottoperiodi: 1° dal 23/8/93 (1° osservazione) al 21/12/93 ( $n_1$ -esima osservazione); 2° dal 22/12/93 ( $n_1+1$ -esima osservazione) al 21/03/94 ( $n_2$ -esima osservazione); 3° dal 22/03/94 ( $n_2+1$ -esima osservazione) al 22/08/94 (365° osservazione). In ogni intervallo individuato per cogliere l'andamento delle serie si sono impiegate delle spline quadratiche (Smith, 1979) tali da garantire la continuità della funzione e delle derivate prime nei nodi ( $n_1, n_2$ ) e negli estremi<sup>7</sup> (1, N).

L'espressione che spiega  $y_{jt}$  ( $j=1,2,3,4$ ) diventa:

$$(2.4) \quad y_{jt} = \alpha_{jh} + \beta_{jh}t + \beta_{jh2}t^2 + \psi_j s_t + \theta_j m_t + \sum_{i=1}^k a_{ji} y_{t-i} + e_{jt},$$

dove  $h = 1$  se  $1 \leq t \leq n_1$ ,  $h = 2$  se  $n_1+1 \leq t \leq n_2$ ,  $h = 3$  se  $n_2+1 \leq t \leq N$ . Esplicitando le condizioni di continuità sopra esposte si ha:

$$(2.5) \quad \begin{aligned} \alpha_{j1} + \beta_{j11}n_1 + \beta_{j12}n_1^2 &= \alpha_{j2} + \beta_{j21}n_1 + \beta_{j22}n_1^2 \\ \alpha_{j2} + \beta_{j21}n_2 + \beta_{j22}n_2^2 &= \alpha_{j3} + \beta_{j31}n_2 + \beta_{j32}n_2^2 \\ \alpha_{j1} + \beta_{j11} + \beta_{j12} &= \alpha_{j3} + \beta_{j31}N + \beta_{j32}N^2 \\ \beta_{j11} + 2\beta_{j12}n_1 &= \beta_{j21} + 2\beta_{j22}n_1 \\ \beta_{j21} + 2\beta_{j22}n_2 &= \beta_{j31} + 2\beta_{j32}n_2 \\ \beta_{j11} + 2\beta_{j12} &= \beta_{j31} + 2\beta_{j32}N \end{aligned} \quad \text{per } j=1,2,3,4.$$

Le condizioni (2.5) si traducono in un insieme di vincoli lineari sui parametri  $\phi_j$  della (2.3), ovvero:

$$\begin{bmatrix} 1 & n_1 & n_1^2 & -1 & -n_1 & -n_1^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & n_2 & n_2^2 & -1 & -n_2 & -n_2^2 \\ 1 & 1 & 1 & 0 & 0 & 0 & -1 & -N & -N^2 \\ 0 & 1 & 2n_1 & 0 & -1 & -2n_1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 2n_2 & 0 & -1 & -2n_2 \\ 0 & 1 & 2 & 0 & 0 & 0 & 0 & -1 & 2N \end{bmatrix} [\phi_j^T] = \begin{bmatrix} 0 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{bmatrix}$$

(2.6)

$$\mathbf{R}^T \{\phi_j^T\} = \mathbf{0}$$

sempre per  $j = 1,2,3,4$ .

I vincoli sui parametri si possono esprimere in modo equivalente in forma esplicita:

$$(2.7) \quad \phi_j^T = \mathbf{R}_\perp \delta_j$$

$9 \times 3 \quad 3 \times 1$

dove la matrice  $\mathbf{R}_\perp$  è una base del complemento ortogonale dello spazio lineare delle colonne della matrice  $\mathbf{R}$  e il vettore  $\delta_j$  comprende i parametri non vincolati. Sostituendo i vincoli su  $\phi_j$  nel modello (2.3) si ottiene:

$$(2.8a) \quad y_{jt} = (\delta_j^T \mathbf{R}_\perp^T) \mathbf{u}_t + \psi_j s_t + \theta_j m_t + \sum_{i=1}^k a_{ji} y_{t-i} + e_{jt}$$

$$(2.8b) \quad y_{jt} = \delta_j^T \mathbf{u}_t^* + \psi_j s_t + \theta_j m_t + \sum_{i=1}^k a_{ji} y_{t-i} + e_{jt}$$

Passando alla notazione matriciale del modello (2.8b) si ottiene:

$$(2.9) \quad \mathbf{Y}_j = \mathbf{U}^* \delta_j + \mathbf{S} \psi_j^T + \mathbf{M} \theta_j^T + \mathbf{Z} \mathbf{g}_j + \mathbf{e}_j$$

$N \times 1 \quad N \times 3 \quad N \times 6 \quad N \times 4 \quad N \times 4k \quad N \times 1$

<sup>7</sup> Questo perchè si ritiene che l'inquinamento segua in ciclo annuale stabile: non sono presenti trend annuali

dove  $Y_j^T = [y_{j1}, y_{j2}, \dots, y_{jt}, \dots, y_{jN}]$ ;  $U^{*T} = [u_1^*, u_2^*, \dots, u_t^*, \dots, u_N^*]$ ;

$S^T = [s_1, s_2, \dots, s_t, \dots, s_N]$ ;  $M^T = [m_1, m_2, \dots, m_t, \dots, m_N]$ ;  $g_j^T = [a_{j1}, a_{j2}, \dots, a_{jk}]$

$Z = [L Y_1, L Y_2, L Y_3, L Y_4, L^2 Y_1, L^2 Y_2, \dots, L^k Y_3, L^k Y_4]$ .

Il modello statistico complessivo consiste quindi in 4 equazioni lineari ( $j=1,2,3,4$ )

$$(2.10) \quad Y_j = \begin{bmatrix} U^* S & M & Z \end{bmatrix} \begin{bmatrix} \delta_j \\ \psi_j^T \\ \theta_j^T \\ g_j^T \end{bmatrix} + e_j = X \beta_j + e_j$$

dove  $E(e_j) = 0$ ,  $E(e_j e_m^T) = \sigma_{jm} I$ , mentre in generale non vale la condizione

$E(X^T e_j) = 0$  in quanto tra le variabili esplicative compaiono le dipendenti ritardate<sup>8</sup>. Il modello generale può essere rappresentato in un'unica espressione matriciale

$$(2.11a) \quad \begin{bmatrix} Y_1 \\ Y_2 \\ Y_3 \\ Y_4 \end{bmatrix} = \begin{bmatrix} X & 0 & 0 & 0 \\ 0 & X & 0 & 0 \\ 0 & 0 & X & 0 \\ 0 & 0 & 0 & X \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

ovvero

$$(2.11b) \quad Y = X^* \beta^* + e$$

dove  $E(e | X^*) = 0$ ,  $E(e e^T | X) = \Lambda \otimes I_N$ , con  $\Lambda = [\sigma_{jm}]$ .

<sup>8</sup> Tuttavia se il processo risulta stazionario, utilizzando il principio dei minimi quadrati si ottengono stimatori consistenti ed asintoticamente normali ed inoltre se i residui risultano non correlati e normali le stime ottenute coincidono con quelle di massima verosimiglianza

Il modello (2.11) rappresenta pertanto un sistema di equazioni *apparentemente non collegate* (SUR)<sup>9</sup>, dove tuttavia i termini di errore sono simultaneamente correlati. Tale modello introdotto per la prima volta da Zellner (1962) appartiene alla classe dei modelli lineari generalizzati. Per la stima dei parametri incogniti  $\beta^*$  si può quindi utilizzare lo stimatore SUR che appartiene alla classe degli stimatori generalizzati dei minimi quadrati: premoltiplicando entrambi i membri

della (2.11b) per  $(\Lambda \otimes I_N)^{-\frac{1}{2}}$  si ottiene

$$(2.12) \quad (\Lambda \otimes I_N)^{-\frac{1}{2}} Y = (\Lambda \otimes I_N)^{-\frac{1}{2}} X^* \beta^* + (\Lambda \otimes I_N)^{-\frac{1}{2}} e$$

e il principio dei minimi quadrati conduce allo stimatore

$$(2.13) \quad b_{SUR} = [X^{*T} (\Lambda^{-1} \otimes I_N) X^*]^{-1} X^{*T} (\Lambda^{-1} \otimes I_N) Y.$$

Tale espressione, essendo le variabili esplicative identiche fra equazioni, coincide con lo stimatore OLS calcolato sulle singole equazioni:

$$(2.14) \quad b_{SUR} = \begin{bmatrix} b_{OLS \ 1} \\ b_{OLS \ 2} \\ b_{OLS \ 3} \\ b_{OLS \ 4} \end{bmatrix} = \begin{bmatrix} [X^T X]^{-1} X^T Y_1 \\ [X^T X]^{-1} X^T Y_2 \\ [X^T X]^{-1} X^T Y_3 \\ [X^T X]^{-1} X^T Y_4 \end{bmatrix}$$

Essendo la (2.11b) una formulazione equivalente del modello (2.1) è possibile stimare i coefficienti del VARX.

<sup>9</sup> Seemingly Unrelated Regression

### 3. La stima dei parametri

Il primo problema da affrontare per la stima dei parametri del modello riguarda l'identificazione dell'ordine della componente autoregressiva. A tale scopo si è utilizzata inizialmente una specificazione del modello, limitata alla parte deterministica ed esogena, con l'obiettivo di esaminare gli andamenti delle funzioni di autocorrelazione dei residui ottenuti. Dall'analisi delle funzioni di autocorrelazione globale e parziale stimate sui residui, la componente autoregressiva del modello risulta essere di ordine uno. Il modello diventa quindi:

$$(3.1) \quad y_t = \Phi u_t + \Psi s_t + \Theta m_t + A_1 y_{t-1} + e_t$$

Per la stima dei coefficienti si è utilizzato il metodo dei minimi quadrati applicato alle singole equazioni (2.9). Tale procedimento, come s'è detto, riconduce all'uso dello stimatore  $b_{sur}$  nel modello (2.11): i risultati del procedimento di stima<sup>1</sup> e dei tests inferenziali sui coefficienti sono riportati nelle tabelle 1÷4.

#### 3.1. Analisi dei residui

Una volta stimati i parametri incogniti del modello, è necessario esaminare i residui,  $e_t = y_t - y_t^*$ ; l'attenzione è rivolta al controllo dell'ipotesi di normalità e alla struttura di autocorrelazione. Nella Tab.5 sono riportati i risultati del test di Kolmogorov-Smirnov ( $k_c$ ) per il controllo dell'ipotesi di normalità. Nella Fig.1 sono invece illustrati i grafici delle stime delle autocorrelazioni globali e parziali: lo scopo dell'analisi è controllare se i relativi correlogrammi si mantengono

all'interno delle bande di confidenza  $\pm \left( \frac{2}{\sqrt{n}} \right)$ . Riguardo alla autocorrelazione, i

risultati indicano che i residui si possono ritenere non autocorrelati; per quanto concerne l'ipotesi di normalità, si segnala che i residui relativi alla centralina Casalecchio non risultano conformi (al livello di significatività 0.05) alla normale. Tuttavia un allontanamento dalla normalità non compromette i risultati in quanto la distribuzione in oggetto non presenta una evidente asimmetria: si può quindi concludere che i residui rispettano le assunzioni richieste.

<sup>1</sup> La condizione di stazionarietà, affinché le stime ottenute godano delle proprietà richieste è che le radici dell'equazione  $|Iz - A_1|$  siano in modulo minori di uno. La verifica ha portato ai seguenti risultati:  $|z_1| = 0.643$ ;  $|z_2| = 0.358$ ;  $|z_3| = 0.358$ ;  $|z_4| = 0.224$ .

Tab. 1 Risultati della stima dei coefficienti del modello (Casalecchio)

Casi validi: 365	Variabile dipendente: Y <sub>CASALECCHIO</sub> (j=1)
Devianza totale: 46.150	Gradi di libertà: 348
R <sup>2</sup> : 0.644	R <sup>2</sup> : 0.627
Devianza residua: 16.439	Durbin-Watson: 2.065552
F(17,348): 36.997	Prob(F <sub>t</sub> > F): 0.000

Var.	Coeff.	Errore Standard	Valore di $t_c$	Prob( $t >  t_c $ )	Stima Standardiz.	Corr. con la variab. dip.
s <sub>1</sub>	1.864532	0.499446	3.733204	0.000	0.041254	0.354811
s <sub>2</sub>	-1.036085	0.250947	-4.128694	0.000	-0.071286	0.734197
s <sub>3</sub>	1.300735	0.177327	7.335222	0.000	0.463586	0.978572
D <sub>1</sub>	0.396137	0.044932	8.816453	0.000	0.047988	0.376898
D <sub>2</sub>	0.307130	0.043756	7.019160	0.000	0.036853	0.378764
D <sub>3</sub>	0.319080	0.043547	7.327214	0.000	0.038287	0.383989
D <sub>4</sub>	0.319976	0.043857	7.295845	0.000	0.038394	0.382780
D <sub>5</sub>	0.336832	0.043551	7.734166	0.000	0.040417	0.384277
D <sub>6</sub>	0.251742	0.043246	5.821129	0.000	0.030207	0.377239
m <sub>1</sub>	0.002085	0.003299	0.631943	0.528	0.010822	0.827977
m <sub>2</sub>	-0.000246	0.001121	-0.219120	0.827	-0.005925	0.984999
m <sub>3</sub>	0.000949	0.001848	0.513417	0.608	0.002177	0.249824
m <sub>4</sub>	-0.049591	0.008932	-5.552305	0.000	-0.044730	0.861522
Y <sub>CAS-I</sub>	0.300267	0.058136	5.164944	0.000	0.300286	0.995427
Y <sub>FFIE t-1</sub>	0.023259	0.044590	0.521610	0.602	0.023738	0.991922
Y <sub>IRN t-1</sub>	0.314388	0.066060	4.759108	0.000	0.246548	0.988148
Y <sub>OSPI-I</sub>	-0.046245	0.049805	-0.928528	0.354	-0.040584	0.987824

Tab. 2 Risultati della stima dei coefficienti del modello (Fiera)

Casi validi: 365	Variabile dipendente: Y <sub>FIERA</sub> (j=2)
Devianza totale: 89.465	Gradi di libertà: 348
R <sup>2</sup> : 0.663	R <sup>2</sup> : 0.648
Devianza residua: 30.127	Durbin-Watson: 2.0030708
F(17,348): 40.319	Prob(F> F): 0.000

Var.	Coeff.	Errore Standard	Valore di t <sub>c</sub>	Prob(t >  t <sub>d</sub>  )	Stima Standardiz.	Corr. con la variab. dip.
s <sub>1</sub>	2.805611	0.676127	4.149531	0.000	0.060820	0.391711
s <sub>2</sub>	-0.937833	0.339721	-2.760595	0.006	-0.063220	0.723322
s <sub>3</sub>	1.160473	0.240058	4.834138	0.000	0.405224	0.969516
D <sub>1</sub>	0.340838	0.060826	5.603465	0.000	0.040453	0.371017
D <sub>2</sub>	0.322970	0.059235	5.452356	0.000	0.037969	0.382746
D <sub>3</sub>	0.260779	0.058952	4.423562	0.000	0.030658	0.382244
D <sub>4</sub>	0.265512	0.059372	4.471992	0.000	0.031214	0.382215
D <sub>5</sub>	0.262590	0.058958	4.453882	0.000	0.030871	0.381775
D <sub>6</sub>	0.135492	0.058545	2.314338	0.021	0.015929	0.368262
m <sub>1</sub>	-0.004170	0.004466	-0.933728	0.351	-0.021209	0.805013
m <sub>2</sub>	-0.001850	0.001518	-1.218280	0.224	-0.043693	0.981082
m <sub>3</sub>	0.002851	0.002502	1.139245	0.255	0.006408	0.248144
m <sub>4</sub>	-0.037259	0.012091	-3.081434	0.002	-0.032926	0.858087
Y <sub>CASI-t-1</sub>	0.133368	0.078701	1.694609	0.091	0.130677	0.991562
Y <sub>FIE-t-1</sub>	0.432536	0.060364	7.165426	0.000	0.432519	0.993821
Y <sub>IRN-t-1</sub>	0.069415	0.089429	0.776200	0.438	0.053335	0.989146
Y <sub>OSPI-t-1</sub>	0.028924	0.067424	0.428982	0.668	0.024869	0.989636

Tab. 3 Risultati della stima dei coefficienti del modello (Irnerio)

Casi validi: 365	Variabile dipendente: Y <sub>IRNERIO</sub> (j=3)
Devianza totale: 94.486	Gradi di libertà: 348
R <sup>2</sup> : 0.839	R <sup>2</sup> : 0.832
Devianza residua: 15.195	Durbin-Watson: 2.0120224
F(17,348): 106.822	Prob(F> F): 0.000

Var.	Coeff.	Errore Standard	Valore di t <sub>c</sub>	Prob(t >  t <sub>d</sub>  )	Stima Standardiz.	Corr. con la variab. dip.
s <sub>1</sub>	1.571415	0.480175	3.272587	0.001	0.044343	0.438826
s <sub>2</sub>	-0.380165	0.241265	-1.575718	0.116	-0.033359	0.718306
s <sub>3</sub>	0.546447	0.170485	3.205240	0.001	0.248383	0.957453
D <sub>1</sub>	0.318660	0.043198	7.376755	0.000	0.049232	0.364629
D <sub>2</sub>	0.299073	0.042068	7.109307	0.000	0.045768	0.378555
D <sub>3</sub>	0.214269	0.041867	5.117842	0.000	0.032790	0.372704
D <sub>4</sub>	0.252264	0.042165	5.982755	0.000	0.038604	0.372911
D <sub>5</sub>	0.330755	0.041871	7.899436	0.000	0.050616	0.383986
D <sub>6</sub>	0.248777	0.041578	5.983424	0.000	0.038071	0.379195
m <sub>1</sub>	-0.009886	0.003172	-3.116803	0.002	-0.065448	0.765726
m <sub>2</sub>	0.003334	0.001078	3.092085	0.002	0.102519	0.980109
m <sub>3</sub>	-0.001094	0.001777	-0.615850	0.538	-0.003202	0.246205
m <sub>4</sub>	-0.041577	0.008587	-4.841856	0.000	-0.047828	0.842361
Y <sub>CASI-t-1</sub>	0.017245	0.055893	0.308542	0.758	0.021995	0.987719
Y <sub>FIE-t-1</sub>	0.045686	0.042870	1.065693	0.287	0.059468	0.990057
Y <sub>IRN-t-1</sub>	0.535561	0.063511	8.432519	0.000	0.535645	0.993881
Y <sub>OSPI-t-1</sub>	0.037727	0.047883	0.787883	0.431	0.042225	0.990477



Tab. 4 Risultati della stima dei coefficienti del modello (Ospedale Malpighi)

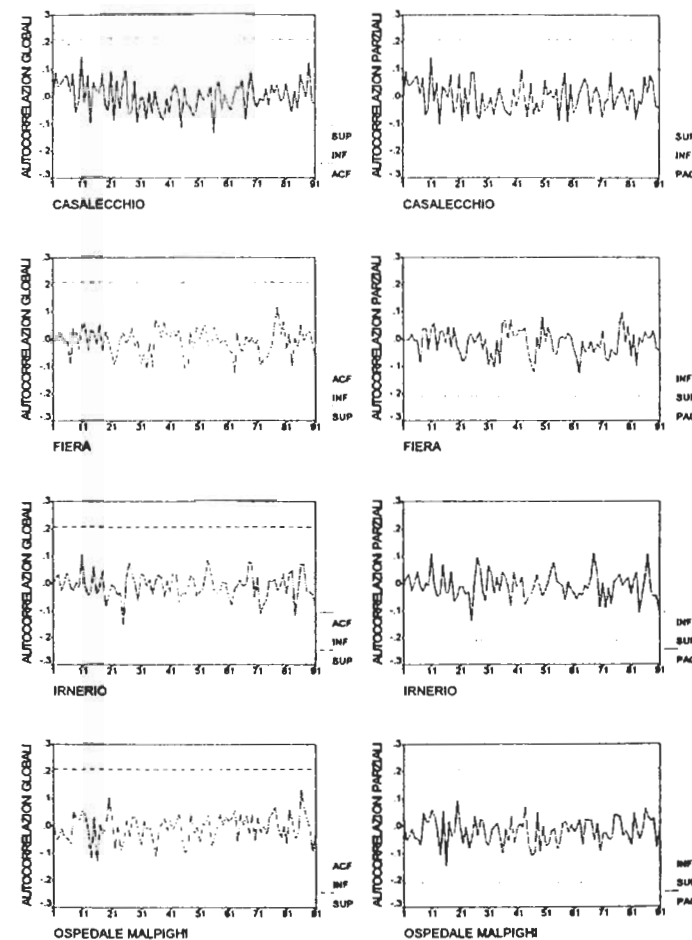
Casi validi: 365	Variabile dipendente: $Y_{OSP. MALPIGHI}$ (j=4)
Devianza totale: 106.282	Gradi di libertà: 348
$R^2$ : 0.731	$R^2$ : 0.719
Devianza residua: 28.604	Durbin-Watson: 1.935032
$F(17,348)$ : 55.596	$Prob(F_p > F)$ : 0.000

Var.	Coeff.	Errore Standard	Valore di $t_c$	$Prob(t >  t_c )$	Stima Standardiz.	Corr. con la variab. dip.
$s_1$	2.291045	0.658812	3.477542	0.001	0.057766	0.424673
$s_2$	-0.625960	0.331021	-1.890996	0.059	-0.049079	0.707439
$s_3$	0.646130	0.233910	2.762300	0.006	0.262420	0.958222
$D_1$	0.414360	0.059269	6.991224	0.000	0.057200	0.366956
$D_2$	0.362503	0.057718	6.280587	0.000	0.049567	0.379365
$D_3$	0.315217	0.057442	5.487530	0.000	0.043102	0.381865
$D_4$	0.306390	0.057852	5.296133	0.000	0.041895	0.379692
$D_5$	0.380467	0.057448	6.622836	0.000	0.052024	0.388528
$D_6$	0.176550	0.057045	3.094903	0.002	0.024141	0.366199
$m_1$	-0.002744	0.004352	-0.630633	0.529	-0.016234	0.785567
$m_2$	-0.001784	0.001479	-1.206140	0.229	-0.049025	0.976626
$m_3$	-0.000071	0.002438	-0.029191	0.977	-0.000186	0.236044
$m_4$	-0.040440	0.011782	-3.432427	0.001	-0.041566	0.847180
$Y_{CASI-t}$	0.064840	0.076686	0.845522	0.398	0.073893	0.987721
$Y_{FIE-t}$	0.159656	0.058818	2.714382	0.007	0.185688	0.990853
$Y_{IRN-t}$	0.217028	0.087139	2.490587	0.013	0.193949	0.990123
$Y_{OSP-t}$	0.296712	0.065697	4.516348	0.000	0.296730	0.991189

Tab.5 Risultati del test di normalità sui residui

Variabile	Valore di $k_c$ (Lilliefors)	g.l.	$Prob(k > k_c)$
Residui Casalecchio	.0503	365	.0268
Residui Fiera	.0451	365	.0712
Residui Irnerio	.0453	365	.0692
Residui Osp. Malpighi	.0367	365	>.2000

Fig. 1: autocorrelazioni dei residui



### 3.2 Analisi delle funzioni di autocorrelazione incrociata

L'esame dei residui sembra invitare all'ottimismo; tuttavia, per non cadere nell'errore di una non corretta specificazione del modello è opportuno utilizzare ulteriori strumenti diagnostici per suggerire eventuali ordini aggiuntivi nella componente autoregressiva o l'inserimento delle variabili meteorologiche con alcuni ritardi. Lo strumento in grado di fornire tali indicazioni è la *funzione di autocorrelazione incrociata*, nota nella letteratura anglosassone come CCF (Cross Correlation Function).

La funzione di autocorrelazione incrociata viene impiegata nell'analisi di causalità e consente di evidenziare l'esistenza di dipendenze temporali fra le osservazioni di due diverse serie storiche  $x_t$  e  $y_t$  rivelate dalla presenza di valori significativi di questa funzione in corrispondenza dei relativi lags. In realtà il calcolo delle autocorrelazioni incrociate sulle serie originarie spesso nasconde fenomeni di correlazione spuria; ossia è influenzato dalla presenza di eventuali rapporti di correlazioni interni alle singole variabili. Risulta quindi opportuno calcolare la CCF su serie opportunamente depurate dalle proprie componenti interne. *Pierce e Haugh (1977)* hanno suggerito una possibile soluzione che consiste nel calcolare la funzione sulle serie *sbiancate*, vale a dire sui residui White Noise ottenuti da un opportuno modello applicato alla serie originale. L'analisi delle cross correlation sui residui del modello dovrebbe quindi permettere di individuare le eventuali relazioni causa-effetto tra le variabili e il ritardo temporale con il quale si influenzano. Queste relazioni, se presenti in modo significativo, dovrebbero corrispondere ad ordini della componente autoregressiva del modello non compresi nella forma originale. Nelle Fig.2a÷Fig.2c sono riportati i risultati dei tests per il controllo della significatività della funzione ai vari lags.

Fig. 2a CCF Casalecchio-Fiera (residui) e Casalecchio-Irnerio (residui)

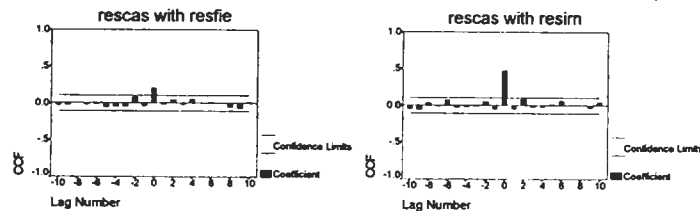


Fig. 2b CCF Casalecchio-Osp. Malp. (residui) e Fiera-Irnerio (residui)

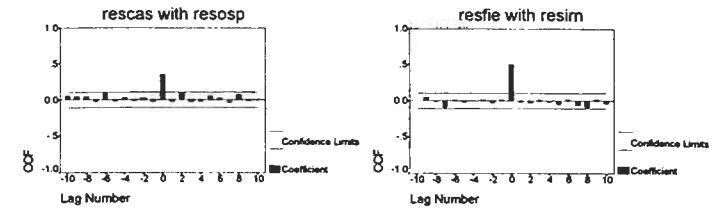
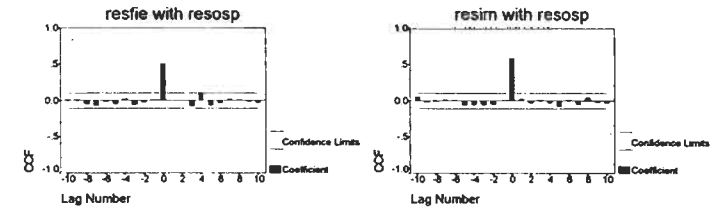


Fig. 2c CCF Fiera-Osp. Malp. (residui) e Irnerio-Osp. Malp. (residui)



I risultati non mostrano la presenza di autocorrelazioni incrociate significative a lags diversi da zero. Questo indica che si possono escludere altri ordini di ritardo nella componente autoregressiva con sufficiente tranquillità. La stessa analisi condotta sulle variabili meteo può essere utile per chiarire se è necessario introdurre le stesse con alcuni ritardi.

### TEMPERATURA

Per l'esame delle funzioni di autocorrelazione incrociata è stato necessario depurare le osservazioni della temperatura dalla struttura di dipendenza interna. In pratica è stato individuato un opportuno modello ARIMA; nel nostro studio si tratta semplicemente di un modello AR(1), in cui i residui tra la serie originale e la serie teorica risultano di tipo *rumore bianco*. La serie così ottenuta (res temp) è stata utilizzata per calcolare la CCF campionaria con i residui del modello (Fig.3a÷Fig.3b).

Fig. 3a CCF Temperatura-Casalecchio, Temperatura-Fiera

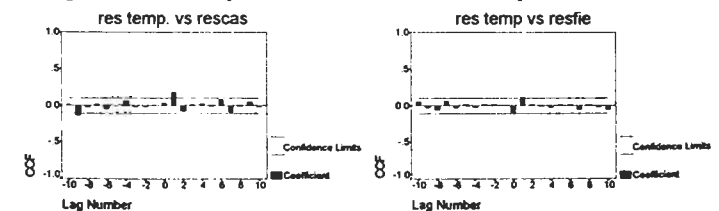
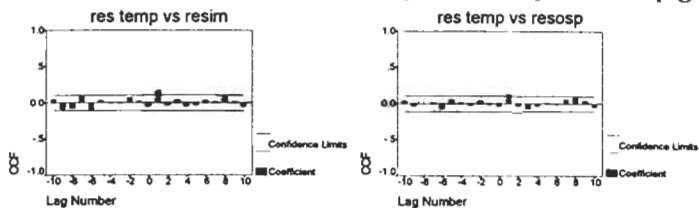


Fig. 3b CCF Temperatura-Irnerio, Temperatura-Ospedale Malpighi



**UMIDITA'**

Sui valori dell'umidità si sono ottenuti dei residui white noise adattando alla serie originale un modello di tipo AR(1). L'analisi della funzione di autocorrelazione incrociata con i residui delle equazioni del modello è visualizzata nelle Fig.4a÷4b

Fig. 4a CCF Umidità-Casalecchio, Umidità-Fiera

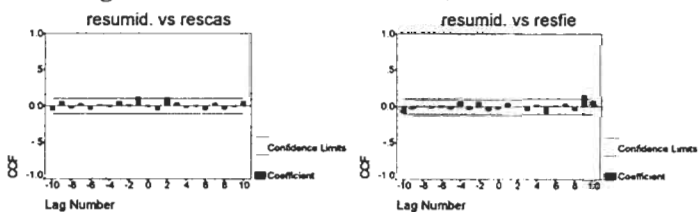
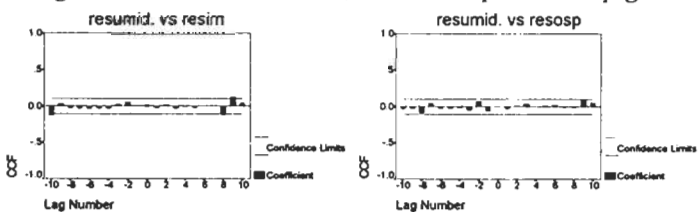


Fig. 4b CCF Umidità-Irnerio, Umidità-Ospedale Malpighi



**PRECIPITAZIONE**

Un procedimento analogo ai precedenti è stato seguito per analizzare la funzione di autocorrelazione incrociata della precipitazione. Il modello utilizzato per "sbiancare" la variabile precipitazione è un AR(2). I risultati sono riportati nelle Fig.5a÷5b

Fig. 5a CCF Precipitazione-Casalecchio, Precipitazione-Fiera

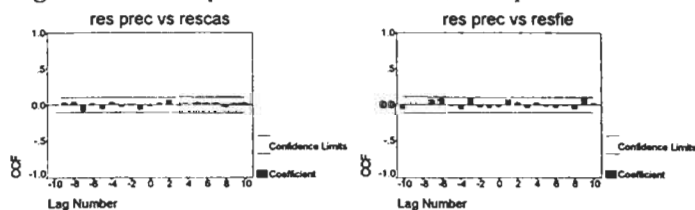
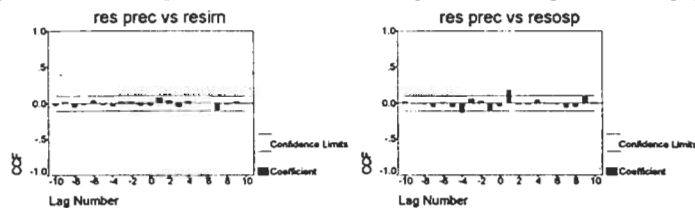


Fig. 5b CCF Precipitazione-Irnerio, Precipitazione-Ospedale Malpighi



**VELOCITA' DEL VENTO**

Per la velocità del vento non è stato necessario identificare alcun modello, in quanto dall'esame delle funzioni di autocorrelazione globale e parziale non sono stati identificati particolari patterns. Le Fig.6a÷6b riportano l'andamento della funzione di autocorrelazione incrociata con i residui del modello.

Fig. 6a CCF Vento-Casalecchio, Vento-Fiera

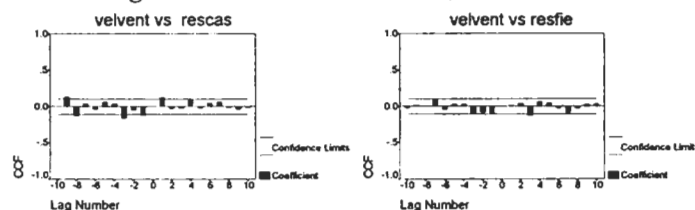
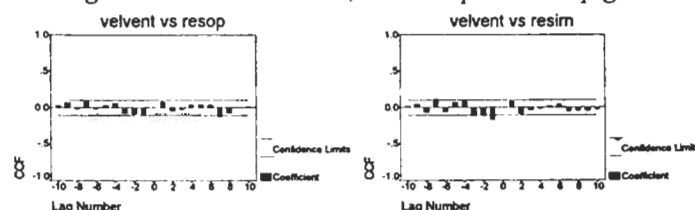


Fig. 6b CCF Vento-Irnerio, Vento-Ospedale Malpighi



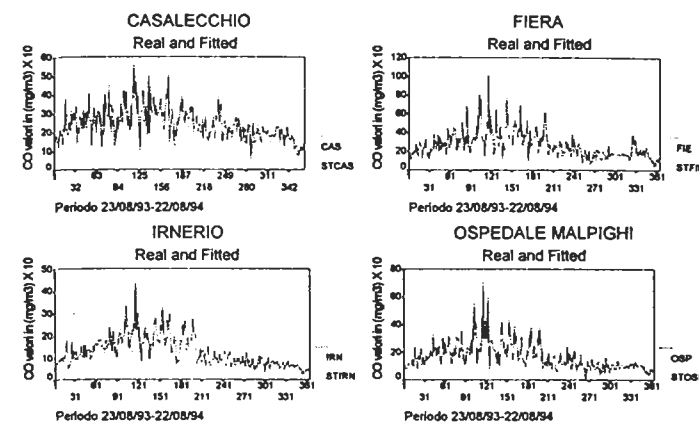
Le indicazioni fornite dall'esame delle funzioni CCF sembrano suggerire l'inserimento della temperatura come variabile esplicativa nel modello con un lag di ritardo per tutte le centraline. Per le altre variabili, invece, non si hanno segnali altrettanto precisi. Tra queste, la velocità del vento potrebbe avere un qualche ruolo esplicativo, se inserita nel modello con qualche lags di ritardo. Visti i risultati interlocutori dell'analisi, si è deciso di procedere nel modo seguente: si inseriscono come esplicative nel modello (3.1), oltre alle variabili già presenti, la temperatura con un lag di ritardo (TEMP\_1) e la velocità del vento con uno e due lags di ritardo (VENT\_1 e VENT\_2); si esaminano quindi i risultati della stima dei rispettivi coefficienti per trarre riscontri oggettivi sul loro contributo esplicativo.

I risultati, non riportati per esteso per esigenze di sintesi, mostrano che l'inserimento delle variabili in questione non comporta un miglioramento della capacità esplicativa del modello. I valori di  $R^2$  non si differenziano in modo sensibile da quelli del modello originale (Tab.1÷Tab.4). L'unico coefficiente significativamente diverso da zero (al livello 0.05) è relativo a VENT\_1 per la centralina Casalecchio. Tuttavia, il contributo di questa variabile è trascurabile e difficilmente interpretabile, in quanto il parametro risulta positivo. Si può quindi concludere che il modello inizialmente specificato, (3.1), sembra essere il migliore. Nel seguito, viene valutata la performance del modello nel spiegare i dati e sono commentati i risultati.

### 3.3 Analisi dei risultati

Osservando i risultati complessivi (Tab.1÷Tab.4), emerge un buon adattamento del modello: infatti, le serie stimate seguono senza eccessivi scostamenti l'andamento dei valori originali (Fig.7) e i valori di  $R^2$  risultano tutti sufficientemente elevati (Range di  $R^2$  : 0.644÷0.838). I parametri corrispondenti alle variabili  $D_1, D_2, \dots, D_6$  risultano, in tutte le equazioni, positivi e significativamente diversi da zero. Si conferma così l'ipotesi che mediamente l'inquinamento da monossido di carbonio risulta, alla domenica, sensibilmente minore rispetto ai giorni feriali, probabilmente a causa del minor numero di veicoli a motore che circolano nel giorno festivo. I coefficienti relativi alle ritardate tra le centraline. A questo proposito, nella Tab.6 è riportato uno schema riassuntivo dove sono componenti autoregressive permettono di analizzare le interdipendenze spaziali indicati i valori dei coefficienti della matrice  $A_1$  e i corrispondenti risultati dei test ( $\text{Prob}(t > |t_c|)$ ).

Fig.7 Confronto tra i dati reali e quelli stimati



Tab. 6 Significatività dei coefficienti della matrice  $A_1$

Variabile rit→	$Y_{CAS\ t-1}$	$Y_{FIE\ t-1}$	$Y_{IRN\ t-1}$	$Y_{OSP\ t-1}$
Variabile dip.↓				
$Y_{CAS\ t}$	0.300267 (0.000)	0.023259 (0.602)	0.314388 (0.000)	-0.046245 (0.354)
$Y_{FIE\ t}$	0.133368 (0.091)	0.432536 (0.000)	0.069415 (0.438)	0.028924 (0.668)
$Y_{IRN\ t}$	0.017245 (0.758)	0.045686 (0.287)	0.535561 (0.000)	0.037727 (0.431)
$Y_{OSP\ t}$	0.064840 (0.398)	0.159656 (0.007)	0.217028 (0.013)	0.296712 (0.000)

Si può notare che sulle variabili  $Y_{FIE\ t}$  e  $Y_{IRN\ t}$  non si manifestano dipendenze spaziali significative.  $Y_{CAS\ t}$  risulta essere invece influenzata da  $Y_{IRN\ t-1}$  e sempre  $Y_{IRN\ t-1}$  manifesta un'influenza su  $Y_{OSP\ t}$  che risulta dipendere anche da  $Y_{FIE\ t-1}$ . I risultati mettono in evidenza diverse caratteristiche importanti sul problema dell'inquinamento da monossido di carbonio. In primo luogo, il fenomeno della persistenza nella bassa atmosfera degli inquinanti: in ogni centralina le concentrazioni medie del monossido di carbonio dipendono fortemente da quelle registrate il giorno prima. Anche le dipendenze spaziali ritardate tra le osservazioni possono essere dovute in parte alla persistenza dell'inquinante. E' probabile tuttavia che queste relazioni statistiche siano da attribuire anche alle interdipendenze presenti tra i flussi di traffico che interessano le aree di rilevazione delle varie centraline. Riguardo alle variabili meteorologiche considerate nel modello come ingressi esogeni, si osserva che la velocità del vento ( $m_4$ ) compare con un coefficiente negativo e significativamente diverso da

zero ( $\alpha = 0.05$ ) in tutte e quattro le centraline. Il vento si caratterizza quindi in modo chiaro come un fattore di pulizia dell'atmosfera. La temperatura ( $m_1$ ) e l'umidità ( $m_2$ ) invece non manifestano altrettanto chiaramente la loro influenza. Agiscono infatti direttamente solo su  $Y_{IRN_t}$  e indirettamente, attraverso  $Y_{IRN_{t-1}}$ , su  $Y_{CAS_t}$  e  $Y_{OSP_t}$ , mentre la pioggia ( $m_3$ ), contrariamente a quanto ci si poteva aspettare, non risulta avere influenza sull'inquinante<sup>2</sup>.

### 3.4 Conclusioni

Il modello specificato ha confermato solo in parte le aspettative sulle relazioni intercorrenti fra le variabili meteorologiche e l'inquinamento. La pioggia risulta essere l'unica variabile priva di influenza sul CO, mentre ci si aspettava che si rivelasse come un fattore di "lavaggio" dell'aria. Tale risultato, ovviamente, può dipendere dal fatto che la relazione pioggia-inquinamento non sia del tipo specificato nel modello, ma può anche essere dovuto ad un altro ordine di motivi sui quali sarebbe interessante fare luce.

Si è confermato, come è ovvio, che la domenica risulta essere il giorno mediamente meno inquinato della settimana mentre il sabato non si distingue, sotto questo aspetto, dagli altri giorni. A questo proposito sarebbe interessante sapere se ciò è dovuto ad un numero di veicoli circolanti di sabato pari a quello degli altri giorni, oppure se il sabato sconta un effetto persistenza dal venerdì. L'impiego di un modello autoregressivo vettoriale ha consentito inoltre di utilizzare i coefficienti della componente autoregressiva per spiegare le relazioni spaziali ritardate tra le centraline, fornendo così interessanti informazioni sulla struttura spazio-temporale dell'inquinamento da monossido di carbonio. Tali informazioni possono essere sfruttate, con le dovute precauzioni, per ottenere indicazioni sulle dinamiche del traffico.

### 4. Le previsioni

L'interesse per i temi legati alla previsione dei livelli degli inquinanti si inserisce in una politica rivolta alla messa in opera di provvedimenti preventivi a salvaguardia della salute dei cittadini. L'attenzione è motivata anche dalla considerazione che la costruzione di modelli previsivi può consentire di misurare e valutare l'efficacia dei diversi strumenti di cui dispongono le amministrazioni cittadine per modificare le condizioni della qualità dell'aria: lo spostamento tra previsioni e realizzazioni in presenza di un provvedimento (per esempio il

<sup>2</sup> Va ricordato che a causa della non disponibilità immediata dei dati non sono state incluse nel modello numerose variabili (come la pressione atmosferica) che avrebbero potuto aumentare la capacità esplicativa.

blocco del traffico) riflette non tanto il fallimento della previsione, quanto la misura dell'efficacia del provvedimento messo in atto. Prima di parlare delle previsioni è opportuno ricordare che il modello utilizzato,

$$(4.1) \quad y_t = \Phi u_t + \Psi s_t + \Theta m_t + A_1 y_{t-1} + e_t,$$

considera esplicative per  $y_t$  le variabili meteorologiche,  $m_t$ , osservate nell'istante  $t$ . La previsione  $\hat{y}_{t+1,(t)}$  fatta al tempo  $t$  implica la conoscenza dei valori delle variabili  $y_t$ ,  $u_{t+1}$ ,  $s_{t+1}$  e  $m_{t+1}$ . Se per  $y_t$ ,  $u_{t+1}$ , e  $s_{t+1}$  non sussistono particolari problemi, in quanto la prima variabile è osservata al tempo  $t$  e le altre due, essendo deterministiche, assumono valori al tempo  $t+1$  noti dall'istante  $t$ , per le variabili meteorologiche è invece necessario conoscere una loro previsione  $\hat{m}_{t+1,(t)}$  fatta al tempo  $t$ . In realtà sarebbe stato più corretto utilizzare nella fase di modellazione e di stima del modello (4.1) non i valori osservati delle variabili meteo,  $m_t$ , ma una previsione  $\hat{m}_{t,(t-1)}$  fatta al tempo  $t-1$ . Purtroppo, per diversi ostacoli "istituzionali", i dati sulle previsioni a un giorno delle variabili meteorologiche non sono stati messi a disposizione; questo ha comportato una limitazione costringendo ad utilizzare per le previsioni, i dati osservati  $m_{t+1}$  al posto di quelli previsti. Naturalmente si è ben consapevoli dei limiti che tale semplificazione comporta; si ritiene tuttavia che i risultati raggiunti possano costituire un punto di riferimento per valutare i limiti e le potenzialità dei modelli autoregressivi vettoriali nella previsione di dati ambientali. Passando alla fase operativa, le previsioni a un giorno sono date dall'espressione

$$(4.2) \quad \hat{y}_{t+1,(t)} = \hat{\Phi} u_{t+1} + \hat{\Psi} s_{t+1} + \hat{\Theta} m_{t+1} + \hat{A}_1 y_t$$

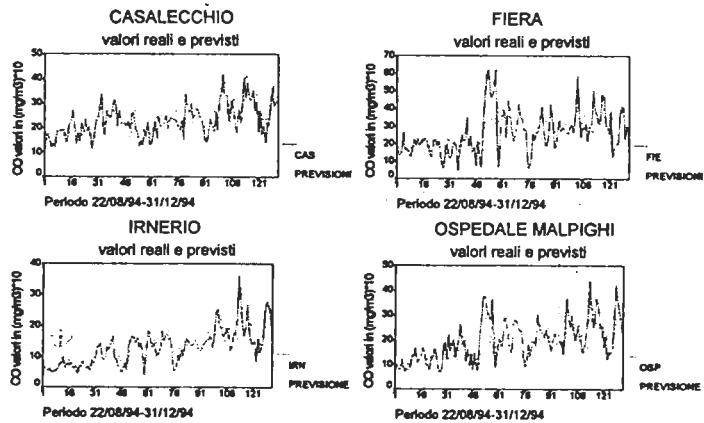
dove  $\hat{\Phi}$ ,  $\hat{\Psi}$ ,  $\hat{\Theta}$ ,  $\hat{A}_1$  sono le stime (ricavate nel capitolo precedente) delle

matrici  $\Phi$ ,  $\Psi$ ,  $\Theta$ ,  $A_1$ .

Per valutare l'abilità previsiva del modello il metodo più semplice, ma non il meno efficace, si fonda sul confronto grafico tra i valori previsti e i valori osservati. Nei grafici di Fig.1 si confrontano le previsioni a un giorno realizzate per il periodo dal 22/08/94 al 31/12/94 con i corrispondenti valori originali. Si può notare, in prima battuta, che l'andamento di fondo dei valori osservati è sostanzialmente lo stesso dei valori previsti; il modello quindi riesce a cogliere

questo aspetto. In secondo luogo non sono presenti differenze macroscopiche tra le previsioni e i valori osservati. A supporto del confronto grafico può essere utile ricorrere alla misura della correlazione vero-previsto, cioè la correlazione tra la previsione,  $\hat{y}_{t+1,(t)}$ , e il valore osservato,  $y_{t+1}$ : tali valori sono riportati nella Tab.1.

Fig.1 Confronto previsioni-dati reali

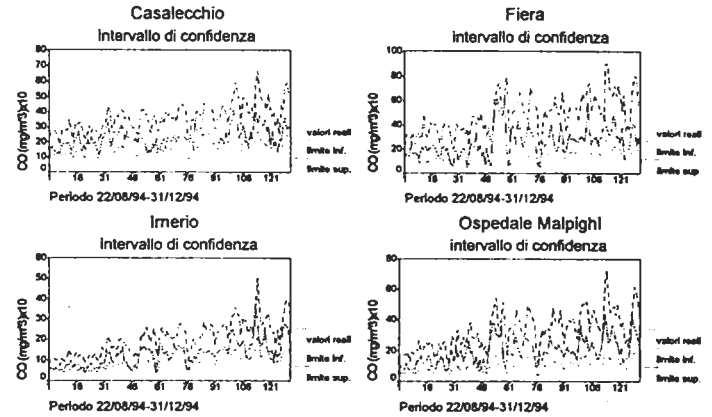


Tab.1 Correlazione Vero-Previsto

Centralina	Casalecchio	Fiera	Imerio	Osp. Malpighi
$\text{corr}(\hat{y}_{t+1,(t)}, y_{t+1})$	0.6818	0.6203	0.8629	0.7790

La correlazione più bassa si registra nella centralina della Fiera: questo può essere dovuto al fatto che il traffico, nella zona in questione, segue un andamento particolare, strettamente legato alle manifestazioni fieristiche. Di conseguenza, la dinamica delle concentrazioni di monossido di carbonio, riflettendo una situazione particolare del traffico, può discostarsi dalla previsione in alcune circostanze. Per le altre centraline invece i valori delle correlazioni risultano piuttosto elevati; tali risultati non fanno altro che confermare le valutazioni emerse precedentemente nel confronto grafico. Il calcolo degli intervalli di confidenza consente di rendere più accurata la valutazione della capacità previsiva del modello: nei grafici rappresentati nelle Fig.2 sono confrontati gli intervalli di confidenza al 95% per le previsioni al lag1,  $\hat{y}_{jt+1,(t)}$ , con le osservazioni,  $y_{jt+1}$ .

Fig.2 Intervallo di confidenza delle previsioni



Dai grafici si osserva che soltanto un numero limitato di punti cade fuori dell'intervallo di previsione, a conferma della discreta abilità previsiva del modello.

4.1 Analisi degli errori di previsione

Per valutare in modo esaustivo la performance previsiva è opportuno analizzare anche gli errori di previsione. Indicando con  $e_{t+1,(t)} = y_{t+1} - \hat{y}_{t+1,(t)}$  l'errore di previsione, si possono definire le seguenti misure di sintesi:

$$\text{MPE} = \frac{100}{m} \sum_{t=1}^m \frac{e_{t+1,(t)}}{y_{t+1}} \text{ (errore medio \%);}$$

$$\text{MSE} = \frac{1}{m} \sum_{t=1}^m e_{t+1,(t)}^2 \text{ (errore quadratico medio);}$$

$$\text{MAE} = \frac{1}{m} \sum_{t=1}^m |e_{t+1,(t)}| \text{ (errore medio assoluto);}$$

$$\text{MAPE} = \frac{100}{m} \sum_{t=1}^m \left| \frac{e_{t+1,(t)}}{y_{t+1}} \right| \text{ (errore medio assoluto \%)}.$$

I valori di tali indici, insieme alle più comuni statistiche descrittive, sono riportati nelle Tab. 2a÷2d. Per completare l'analisi descrittiva, nelle Fig.3a÷3b sono visualizzati gli istogrammi degli errori, in modo da evidenziarne

graficamente e sinteticamente alcune delle caratteristiche emerse dai calcoli precedenti.

*Tab. 2a Casalecchio errori di previsione*

m=132	Media = -0.0483
Min = -0.6963	Max = 0.4692
Varianza = 0.0481	Mediana = -0.0211
Kurtosis = -0.0363	Skewness = -0.2570
MPE = -1.948	MSE = 0.0500
MAE = 0.1770	MAPE = 5.9235

*Tab. 2b Fiera errori di previsione*

m=132	Media = -0.0833
Min = -1.6507	Max = 0.8414
Varianza = 0.1646	Mediana = -0.0778
Kurtosis = 2.2866	Skewness = -0.9448
MPE = -4.9034	MSE = 0.1703
MAE = 0.2957	MAPE = 10.9641

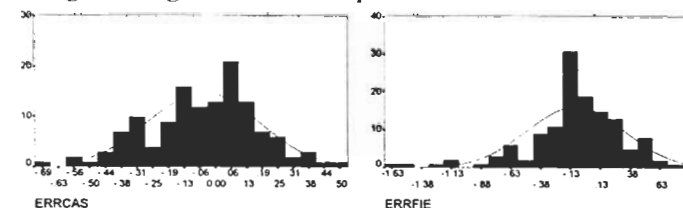
*Tab. 2c Irnerio errori di previsione*

m = 132	Media = -0.0751
Min = -1.1334	Max = 0.4706
Varianza = 0.0582	Mediana = -0.0592
Kurtosis = 2.2958	Skewness = -0.6454
MPE = -4.9034	MSE = 0.0634
MAE = 0.1898	MAPE = 8.6680

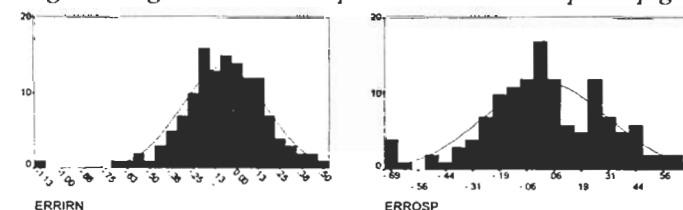
*Tab. 2d Osp. Malpighi errori di previsione*

m=132	Media = 0.0242
Min = -0.6861	Max = 0.6825
Varianza = 0.0809	Mediana = 0.0062
Kurtosis = 0.0051	Skewness = -0.1645
MPE = -0.0007	MSE = 0.0808
MAE = 0.2222	MAPE = 8.0376

*Fig. 3a Istogramma errori di previsione Casalecchio-Fiera*



*Fig. 3b Istogramma errori di previsione Irnerio-Osp. Malpighi*



La legge distributiva degli errori è un altro punto importante per giudicare la qualità delle previsioni: infatti, se le differenze tra valore reale e valore previsto si adeguano alla distribuzione normale, si può ritenere che tali errori siano da imputare solo a cause aleatorie.

Nel prospetto (Tab.3) sono illustrati i risultati del test (test di Kolomogorov-Smirnov ( $k_c$ )) per il controllo della normalità degli errori di previsione.

Inventario N. 16986 D.

Niu X. Frederick J. E. Stein M. L. Tiao G.C. (1992) *Trends in column ozone based on TOMS Data: Dependence on Month, Latitude and Longitude*, Journal of Geophysical Research, 97, 14661-14669.

Niu X. Tiao G.C. (1995) *Modeling satellite ozono data*, Journal of the American Statistical Association, Vol. 90, pp 969-983.

Pierce A., Haugh D. (1977) *Causality in temporal system*, Journal of Econometrics Vol.5, pp 265-293.

Reinsel G.C. Tiao G.C. Wang M.N. Lewis R. Nychka D. (1981) *Statistical Analysis of Stratospheric Ozone Data for the Detection of Trends*, Atmospheric Enviroment. Vol. 15, pp 1569-1577.

Roch R. Pellerin J. (1982) *On long term air quality trends and intervention analysis*, Atmospheric Enviroment. 16, pp 161-169.

Simpson R. W. Layton A. P.(1983) *Forecasting peak ozone levels*, Atmospheric Enviroment 17, pp 1644-16654.

Smith P.L. (1979) *Splines as a useful and convenient statistical tool*, The American Statistician. Vol. 33, No. 2.

Zannetti P. (1990) *Air Pollution Modelling*, Van Nostrand Reinhold, New York.

Zellner R. A. (1962) *An Efficient Method of Estimating Seemingly Unrelated Regression and Tests for Aggregation Bias*, Journal of the American Statistical Association, Vol. 57, pp 348-368.