

UNIVERSITÀ DEGLI STUDI DI PADOVA

FACOLTÀ DI SCIENZE STATISTICHE

CORSO DI LAUREA IN SCIENZE STATISTICHE ED ECONOMICHE

TESI DI LAUREA

*Inquinamento e salute:
un approfondimento dell'analisi
a breve termine per la città di
Verona*

Relatore:

Ch.ma Prof. MONICA CHIOGNA

Laureanda: SARA POFTE

Matricola N°: 450261

ANNO ACCADEMICO 2002-2003

A mia nonna Pinuccia

Prefazione

Negli ultimi anni si è andata diffondendo nella società italiana una maggior sensibilità nei confronti dell'ambiente e soprattutto della qualità dell'aria. Questa attenzione è nata dalla sempre maggiore evidenza di una associazione piuttosto netta tra inquinamento ed effetti dannosi apportati alla salute umana, evidenza approfondita sia negli aspetti qualitativi che quantitativi dallo svilupparsi di un'abbondante letteratura.

L'obiettivo di questa tesi è quello di studiare gli effetti dell'inquinamento attraverso i modelli additivi generalizzati e al riguardo si sono utilizzate le due funzioni *gam* e *gam.exact*, implementate attraverso il software statistico S-PLUS e si sono posti a confronto i risultati ottenuti.

Il Capitolo 1 introduce al tema dell'inquinamento e al problema della stima dei coefficienti che misurano l'intensità dell'effetto di ogni inquinante preso in considerazione sulla ricoverabilità ospedaliera e, quindi, sulla salute. In tale Capitolo vengono presentate diverse tipologie di studi, tra loro alternative, sottolineando di ognuna gli aspetti salienti. Per concludere vengono introdotti i temi del confondimento, come problema per l'efficienza delle stime dei coefficienti di interesse, e degli effetti ritardati che ogni inquinante può avere sull'organismo umano.

Nel Capitolo 2 vengono introdotti i modelli additivi e i modelli additivi generalizzati e viene proposto come esempio il modello additivo generalizzato nella forma adottata nella metanalisi americana, soffermandosi sul significato e sull'interpretazione delle variabili che lo compongono. Successivamente

vengono affrontati alcuni problemi nella stima dei modelli additivi generalizzati tramite il software statistico S-PLUS, tra cui quelli relativi al criterio di convergenza e alla sottostima degli errori standard. Da ultimo viene riportato il calcolo della matrice di varianze e covarianze del vettore dei coefficienti d'interesse.

Nel Capitolo 3 viene riportata un'analisi descrittiva dei dati utilizzati nel presente progetto per evidenziarne andamenti particolari e ciclicità e valutare le correlazioni esistenti tra le variabili in gioco. L'analisi è stata effettuata attraverso il commento di grafici e tabelle.

Nel Capitolo 4 sono descritte le variabili introdotte nel modello adottato nell'analisi e riportati i risultati emersi dal confronto tra le funzioni *gam* e *gam.exact*. Questo confronto viene fatto anche al variare dei gradi di libertà delle funzioni di lisciamiento introdotte nel modello. Infine viene espresso un giudizio conclusivo sui risultati ottenuti.

I grafici e le elaborazioni presentati in questo progetto sono stati realizzati mediante il programma statistico S-PLUS.

Indice

<i>Prefazione</i>	<i>iii</i>
1 Introduzione	1
1.1 Tipologie di studi	3
1.1.1 Studi di serie temporali	4
1.1.2 Studi case-crossover	5
1.1.3 Studi di panel	5
1.1.4 Studi di coorte	6
1.2 L'approccio basato sulle serie storiche	6
1.2.1 Il problema del confondimento	8
1.2.2 Effetti ritardati	9
2 La modellazione semi-parametrica negli studi di serie tem- porali	11
2.1 I modelli additivi e i modelli additivi generalizzati	13
2.2 I modelli additivi generalizzati nello studio degli effetti dell'in- quinamento sulla salute umana	15
2.3 Alcuni problemi nella stima dei modelli additivi generalizzati tramite il software S-PLUS	17
2.3.1 Il problema della convergenza	19
2.3.2 Il problema della sottostima degli errori standard	20
2.4 Ulteriori elementi di attenzione	22

3	Analisi descrittive e grafiche	25
3.1	Analisi grafiche	28
3.1.1	Gli inquinanti	28
3.1.2	Le variabili confondenti	37
3.1.3	Coefficienti di correlazione	40
3.1.4	La variabile risposta	42
4	Analisi e risultati nel confronto tra le funzioni <i>gam</i> e <i>gam.exact</i>	47
4.1	Stima del modello mediante la funzione <i>gam</i>	47
4.1.1	Costruzione del <i>core model</i>	48
4.1.2	Introduzione dell'inquinante	53
4.1.3	Analisi dell'eventuale autocorrelazione (solo per lag 0-3)	54
4.1.4	Risultati	56
4.2	Confronto tra le funzioni <i>gam</i> e <i>gam.exact</i>	59
4.3	La scelta dei gradi di libertà	63
4.4	Conclusioni	68
A	Calcolo della variazione percentuale	71
B	Stime dei coefficienti e degli errori standard ottenute applicando la funzione <i>gam</i>	73
C	Stime dei coefficienti e degli errori standard ottenute applicando la funzione <i>gam.exact</i>	77
D	Una sintesi del confronto tra le funzioni <i>gam</i> e <i>gam.exact</i>	81
E	Variazioni percentuali relative all'effetto dell'inquinante	87
	Elenco delle figure	94
	Elenco delle tabelle	97
	Bibliografia	98

Capitolo 1

Introduzione

L'associazione tra inquinamento atmosferico ed effetti dannosi per la salute umana, espressa in termini di mortalità e morbosità, è stata messa in luce in occasione di vari episodi di grave inquinamento ambientale che colpirono Stati Uniti ed Europa nel corso del ventesimo secolo (Logan and Glasg, 1953; Ciocco and Thompsom, 1961).

Parlare di una natura causale di tale associazione è prematuro dal momento che essa ancora non è chiara in tutti i suoi termini. Alcuni agenti inquinanti, la cui presenza nell'aria può produrre degli effetti nocivi sulla salute dell'uomo, si possono considerare da tempo individuati, ma la natura e la portata di tali effetti sono ancora largamente da definire con adeguati studi ed indagini tecniche ed epidemiologiche, specie per quanto riguarda gli inquinamenti non episodici e loro conseguenze a lungo termine.

Appare comunque chiaro che gli anziani e i soggetti il cui stato di salute è compromesso costituiscano un bersaglio privilegiato dell'inquinamento; mentre sull'esistenza o meno di una soglia al di sotto della quale l'inquinamento è privo di effetti i risultati attualmente disponibili sono contrastanti.

Negli anni più recenti si è sviluppata un'abbondante letteratura riguardante gli effetti dell'inquinamento sulla salute umana, grazie alla disponibilità sempre maggiore di statistiche correnti di mortalità e morbosità, nonché di strumentazione e procedure di analisi sempre più sofisticate. Gli strumenti

e i disegni metodologici utili per l'analisi dei dati si differenziano in relazione alla durata dell'esposizione ai fattori inquinanti, che può essere di breve, medio o lungo termine, e al tipo di effetto che hanno sulla salute, che può essere acuto o cronico (National Research Council, 1998).

Gli effetti acuti sono generalmente fugaci e dovuti soprattutto ad esposizioni variabili nel tempo, mentre quelli cronici sono persistenti e più probabilmente dovuti all'effetto di una esposizione prolungata e cumulata nel tempo. Le conseguenze a lungo termine rappresentano l'aspetto più grave e controverso nel problema della conoscenza dei rapporti tra inquinamento atmosferico e salute umana. Infatti è assai difficile discriminare, tra i numerosi e variabili fattori causali, la parte che compete all'inquinamento atmosferico.

Nell'analisi di tali dati, particolari attenzioni vanno poste sia nel controllo di fattori confondenti quali l'età e la stagionalità, che si differenziano a seconda della scelta del disegno, sia nella correlazione tra le covariate, che non può essere esclusa completamente dal modello. Per correlazione tra covariate si intende non solo la correlazione tra inquinanti, ma anche tra inquinanti e variabili confondenti, quali temperatura e umidità, o tra variabili confondenti stesse o tra altre covariate presenti nel modello, poiché ognuna di queste variabili interagisce con le altre in un dato periodo e in una data zona.

Nel paragrafo seguente vengono passati in rassegna alcuni tipi di studi che analizzano l'associazione tra inquinamento atmosferico e salute umana. Nel seguito del lavoro ci si è focalizzati invece sugli studi di *serie temporali*, analizzando una particolare classe di modelli, ossia i modelli additivi generalizzati.

1.1 Tipologie di studi

La maggior parte degli studi riguardanti l'effetto dell'inquinamento atmosferico sulla salute sono riconducibili ad una delle seguenti tipologie:

- studi di *serie temporali*;
- studi *case - crossover*;
- studi di *panel*;
- studi di *coorte*.

La scelta del tipo di disegno dipende dagli obiettivi della ricerca e dalla disponibilità di dati: non esiste un disegno ottimale per tutte le applicazioni.

Gli studi di *serie storiche temporali*, *case-crossover* e di *panel* sono più adatti per lo studio dell'esposizione di breve termine all'inquinamento atmosferico. Inoltre i primi consentono un'analisi approfondita degli effetti acuti, basandosi sull'associazione tra la variazione nel tempo dell'esposizione agli agenti inquinanti e dei risultati ottenuti.

Gli studi di *coorte* permettono invece uno studio combinato di medio-lungo termine degli effetti acuti e cronici, cumulando i risultati lungo il periodo di tempo considerato. Così, sebbene l'obiettivo degli studi di *coorte* sia la stima degli effetti cronici, questi possono non essere distinti da quelli acuti: è pressoché impossibile, infatti, identificare le cause di morte di un individuo nell'effetto dell'inquinamento di pochi giorni prima o di quello cumulato nel tempo dal momento che i due effetti interagiscono tra loro nel tempo (Vedal, 1996; Dockery et al., 1993; Pope et al., 1995; Krewski et al., 2000). Si può cercare di separare i due tipi di effetti includendo nei modelli covariate relative a caratteristiche individuali o a fattori ambientali che varino nello spazio e siano dipendenti nel tempo. Il fatto di avere la residenza vicino ad uno stabilimento industriale, ad esempio, può favorire lo sviluppo di malattie cardiache o respiratorie e a lungo andare portare alla morte. Lo stesso vale per i soggetti fumatori o comunque sottoposti a fumo passivo, per i quali la

probabilità di morte è senza dubbio più alta rispetto a quella di un individuo dalle più sane abitudini comportamentali.

Nel seguito verranno brevemente passati in rassegna i singoli tipi di studi.

1.1.1 Studi di serie temporali

Gli studi di *serie temporali* pongono in relazione serie storiche di livelli di esposizione all'inquinamento atmosferico a cui determinati individui sono sottoposti con serie storiche del numero giornaliero di decessi o di diversi tipi di malattie. Questi studi hanno come obiettivo la ricerca di una eventuale associazione tra le variazioni di breve termine nell'indicatore dello stato di salute e le corrispondenti variazioni di breve termine nel livello dell'inquinante di interesse.

Al fine di studiare l'effetto provocato dall'esposizione a particolari agenti inquinanti, tenendo conto di possibili fattori confondenti che potrebbero compromettere le stime, vengono usati modelli lineari generalizzati (GLM¹) (McCullagh and Nelder, 1989) oppure modelli additivi generalizzati (GAM²) (Hastie and Tibshirani, 1990). La seconda classe di modelli è la più usata poiché può modellare con facilità componenti non lineari quali quelle relative alla stagionalità o ad altre variabili temporali ed inoltre è una classe di modelli più flessibile rispetto ad altre completamente parametriche.

Dato che la maggior parte dei dati provengono da fonti pubbliche, gli studi di *serie temporali* sono piuttosto sviluppati e l'evidenza è a favore di una associazione statisticamente significativa tra l'effetto dell'inquinamento e la salute.

¹Generalized Linear Models

²Generalized Additives Models

1.1.2 Studi case-crossover

Gli studi *case-crossover* vengono impiegati per studiare l'esistenza di una associazione tra un'esposizione di breve termine a fattori inquinanti e la presenza di un possibile fattore di rischio.

Il disegno *case-crossover* è simile al disegno *caso-controllo* (Breslan and Day, 1980; Schlesselman, 1994) in cui si pongono a confronto due gruppi distinti di persone: uno affetto da una certa patologia (*caso*) ed uno paragonabile a questo per alcune variabili, ma non colpito dalla malattia (*controllo*). La differenza rispetto ad uno studio *caso-controllo* sta nel fatto che nello studio *case-crossover* si confrontano i dati relativi all'esposizione che precede un determinato evento con quelli relativi ad una serie di tempi presi come riferimento; nello studio *caso-controllo* è invece del tutto assente l'idea temporale del fenomeno dal momento che è previsto semplicemente un confronto tra due individui.

Questo disegno permette di ridurre l'influenza di fattori confondenti legati a caratteristiche proprie degli individui, come l'età e il sesso, grazie al fatto che vengono studiati sempre gli stessi individui, ed è particolarmente adatto per determinare l'associazione tra il rischio di essere colpiti da una certa patologia e l'esposizione di breve termine a fattori inquinanti.

1.1.3 Studi di panel

Gli studi di *panel* si focalizzano su un gruppo di N individui costante nel tempo, che viene analizzato allo scopo di evidenziare eventuali cambiamenti nei risultati ottenuti ad intervalli temporali da 1 a T per i singoli individui e per il gruppo complessivamente. Seguendo nel tempo un gruppo fisso di individui si pone l'attenzione su indicatori del loro stato di salute, primo fra tutti il giorno della morte. Di questi individui vengono studiati atteggiamenti e abitudini comportamentali, quali la condizione di fumatori o meno, l'esposizione a livelli di inquinamento più o meno persistenti e altri fattori di rischio.

Questo disegno è particolarmente usato per studiare gli effetti di breve termine dell'inquinamento atmosferico sulla salute umana. Tuttavia, generalmente, l'interpretazione delle stime è poco chiara, anche per il fatto che spesso gli individui non vengono studiati negli stessi periodi.

1.1.4 Studi di coorte

Gli studi di *coorte* analizzano l'associazione tra l'esposizione di lungo termine all'inquinamento atmosferico e la salute umana. In questo tipo di disegno la variabile di esposizione è l'effetto cumulato nel tempo dell'esposizione all'inquinamento atmosferico.

Sono possibili due tipi di disegno: prospettivo o retrospettivo. Nel primo caso gli individui presi come riferimento per l'analisi compilano un questionario con informazioni riguardanti il sesso, l'età, il peso, l'educazione, la tendenza al fumo e altre informazioni personali, e vengono seguiti nel tempo registrando eventi come la morte o altri eventi che colpiscono la salute umana; nel secondo caso, invece, si fa appello alla memoria dell'intervistato che deve ricordare con precisione gli eventi che riguardano la sua storia passata.

1.2 L'approccio basato sulle serie storiche

La variabile di interesse negli studi di serie storiche dell'effetto a breve termine dell'inquinamento sulla salute umana è quella che conta il numero di morti e/o di ricoveri ospedalieri che avvengono ad una determinata data o in un determinato intervallo temporale. Questa è una variabile di conteggio, ossia limitata a valori interi non negativi, e pertanto il processo che genera i dati è quello di Poisson. Indicato con λ_t il numero atteso di morti in un determinato giorno t , la probabilità che le morti siano y_t è data da:

$$P_{Y_t}(y_t; \lambda_t) = \frac{e^{-\lambda_t} \lambda_t^{y_t}}{y_t!} \quad (1.1)$$

Il processo di Poisson può non risultare stazionario, ovvero la media λ_t

può non essere costante nel tempo, ma variare al variare di p variabili esplicative X_{1t}, \dots, X_{pt} relative a p condizioni osservate il giorno t . In tali studi, infatti, l'unità di osservazione è il giorno e non il singolo individuo: mentre i fattori individuali, come ad esempio la condizione di fumatore o meno o altre abitudini comportamentali, si mantengono generalmente costanti col passare del tempo, altre variabili, come la temperatura e l'umidità, mostrano un andamento variabile in relazione al periodo e alla stagione di riferimento. Così, a differenza dei fattori individuali, i quali non influenzano λ_t , le altre variabili rendono il processo di Poisson non stazionario e di conseguenza è necessario seguire un approccio temporale per studiare l'andamento di questi fenomeni.

Nel modello di Poisson si assume che:

$$\log(E(Y_t)) = \beta_0 + \beta_1 X_{1t} + \dots + \beta_p X_{pt} \quad t = 1, \dots, T \quad (1.2)$$

dove:

- Y_t è il conteggio di morti o di ricoveri giornalieri;
- $E(Y_t) = \lambda_t$ è il valore atteso di Y_t nel giorno t ;
- X_{1t}, \dots, X_{pt} sono le variabili esplicative;
- β_1, \dots, β_p sono i coefficienti di regressione relativi alle variabili esplicative sopraccitate.

Negli ultimi anni la maggior parte degli studi di serie storiche hanno fatto e fanno tuttora uso del modello di Poisson.

Nei paragrafi che seguono vengono trattati due aspetti riguardanti lo studio degli effetti dell'inquinamento: il problema del confondimento e la struttura ad effetti ritardati.

1.2.1 Il problema del confondimento

Nello studio dell'effetto dell'inquinamento sulla salute dell'uomo, come in ogni altro studio di tipo epidemiologico, è importante controllare la presenza o meno di fattori confondenti che possono compromettere le stime.

La maggior parte delle variabili che si riferiscono alla salute, al tempo atmosferico e all'inquinamento presentano una variazione sistematica naturale nel corso dell'anno; così, per evidenziare un'eventuale associazione tra due variabili, è necessario rimuovere l'effetto confondente dovuto a tali andamenti. È necessario separare pertanto nella variazione giornaliera dei conteggi di mortalità o morbosità gli andamenti naturali da ciò che può ritenersi imputabile all'inquinamento e da ciò che può ritenersi un disturbo casuale e che non va tenuto in considerazione nell'adattamento del modello ai dati.

Esistono diversi metodi per affrontare tale problema. Una possibilità, la più usata, è quella di introdurre nel modello funzioni di lisciamiento del tempo, come medie mobili o altri tipi di funzioni, che tengano conto delle fluttuazioni nella mortalità / morbosità non associate all'inquinamento; la variazione residua è usata per stimare l'effetto specifico dell'inquinamento.

Uno dei principali problemi di tale approccio è quello di capire quante e quali funzioni di lisciamiento introdurre nel modello. Infatti, la variabilità nelle serie temporali di mortalità e/o morbosità spiegata dall'inquinamento ha un ordine di grandezza più basso rispetto alla variabilità spiegata dalla stagionalità, dal trend e da altre variabili temporali, essendo il confondimento la causa principale della mortalità. Di conseguenza, la stima dell'effetto dell'inquinamento è sensibile al metodo scelto per tener conto dei possibili fattori confondenti.

Il problema è pertanto quello di capire a quale grado di lisciamiento dei fattori confondenti giungere per modellare in modo corretto i residui. Infatti l'obiettivo non è quello di inserire nel modello, come funzioni di lisciamiento, tutti i pattern presenti nei dati, bensì solo quelli per cui risulta alto il rischio di confondimento che deriva dalla loro omissione.

Altre variabili che possono giocare un ruolo di confondente sono quelle

che si riferiscono a specifici giorni di calendario, quali ad esempio particolari giorni della settimana o giorni di vacanza, avendo i comportamenti delle strutture sanitarie abitudini differenziate nei giorni festivi e nei giorni lavorativi. E' pertanto sensato includere tali variabili nel modello, anche se più pubblicazioni indicano che la loro presenza non ha una notevole influenza sui coefficienti stimati.

Anche termini relativi al tempo atmosferico possono apportare un certo grado di confondimento nella stima della mortalità o morbosità. Le variabili meteorologiche maggiormente utilizzate sono la temperatura e l'umidità. La dipendenza della mortalità e morbosità da tali variabili può risultare non lineare e per modellarla esistono diversi metodi che possono far uso di funzioni di lisciamento (splines, polinomi non lineari o combinazioni di questi metodi) o di variabili indicatrici. È da notare infine che, oltre ad esaminare l'effetto singolo di temperatura e umidità, va studiata anche la loro possibile interazione.

1.2.2 Effetti ritardati

L'effetto delle singole variabili esplicative e di una loro combinazione sulla salute umana può essere immediato ovvero ritardato di alcuni giorni. Spesso, infatti, gli inquinanti e alcune variabili meteorologiche hanno un effetto che si prolunga al di là del giorno stesso.

Per tener conto di ciò, una possibilità è quella di esaminare modelli con più ritardi delle variabili esplicative ma, poiché tali variabili sono serialmente correlate, spesso le stime risultano instabili. Una forzatura è quella di inserire nel modello una media mobile su più giorni, che aiuta ad identificare l'esistenza degli effetti anche ritardati delle variabili esplicative.

Esistono anche degli approcci che adattano funzioni polinomiali al pattern degli effetti ritardati. Ad esempio, una struttura a ritardi distribuiti consentirà l'adattamento di un modello dove può risultare preponderante, per esempio, l'effetto della temperatura di uno e due giorni precedenti il giorno

corrente e del giorno corrente stesso (Francesca Dominici, Lianne Sheppard and Merlise Clyde, 2003).

Capitolo 2

La modellazione semi-parametrica negli studi di serie temporali

Negli ultimi dieci anni sono stati fatti molti progressi nella modellazione statistica delle serie temporali relative all'inquinamento atmosferico e alla salute umana. I modelli standard di regressione usati inizialmente (R.L. Smith, J.M. Davis, J. Sacks, P. Speckman, P. Styer, 2000) sono stati sostituiti quasi completamente da approcci semi-parametrici in cui i termini lineari, corrispondenti agli effetti dei singoli inquinanti, sono i componenti di interesse, mentre i termini di lisciamiento, rappresentati da funzioni di lisciamiento del tempo e della temperatura, sono considerati "rumore" (Speckman, 1988; Hastie and Tibshirani, 1990; Green and Silverman, 1994).

Un approccio molto usato comprende una regressione semi-parametrica di Poisson in cui la variabile risposta è data dal conteggio giornaliero di mortalità / morbosità e sono inclusi nel modello funzioni di lisciamiento del tempo e delle variabili meteorologiche che controllano i fattori confondenti e termini lineari che misurano l'incremento percentuale del tasso di mortalità / morbosità associato ad aumenti nei livelli di inquinamento.

La formulazione generale di un modello di questo tipo è:

$$Y_t \sim \text{Poisson}(\lambda_t) \quad t = 1, \dots, T$$

$$V(Y_t) = \lambda_t \quad (2.1)$$

$$\log(\lambda_t) = \beta_0 + \sum_{j=1}^J \beta_j X_{jt} + \sum_{l=J+1}^L f_l(X_{lt}, df_l) = \eta_t$$

dove:

- i coefficienti β_j misurano l'incremento percentuale nel tasso di mortalità / morbosità per ogni incremento unitario nel livello di inquinamento atmosferico X_{jt} ($j = 1, \dots, J$);
- $f_l(\cdot, df_l)$ rappresentano le funzioni di lisciamento, spesso costruite usando *smoothing splines*, *loess smoothers* o *natural cubic splines* con parametro di lisciamento df_l ($l = J + 1, \dots, L$).

Rispetto ad una specificazione completamente parametrica, questi modelli mostrano una maggior flessibilità nel controllo di fattori confondenti non lineari quali la stagionalità, il trend ed altre variabili temporali, dal momento che includono funzioni non parametriche del tempo e variabili meteorologiche nel modello di regressione.

Per meglio comprendere i metodi attualmente più sviluppati ed usati nello studio dell'inquinamento, si ritiene opportuno introdurre dapprima i modelli additivi per poi passare alla descrizione dei modelli additivi generalizzati.

2.1 I modelli additivi e i modelli additivi generalizzati

Sia Y il vettore casuale delle risposte (Y_1, \dots, Y_T) , α un termine costante che rappresenta l'intercetta del modello, f_j delle arbitrarie funzioni di liscia-mento, X_j ($j = 1, \dots, p$) p variabili esplicative osservate ed ϵ il vettore dei termini d'errore.

La formulazione di un modello additivo (AM) è del tipo:

$$Y = \alpha + \sum_{j=1}^p f_j(X_j) + \epsilon \quad (2.2)$$

dove, come nei modelli di regressione lineare, si assume che:

1. gli errori ϵ siano indipendenti dalle variabili esplicative X_j ;
2. $E(\epsilon) = 0$;
3. $V(\epsilon) = \sigma^2 I$;
4. $f_j(X_j)$ siano funzioni lisce di X_j .

Per stimare le funzioni f_j possono essere usati *smoother* lineari come *medie mobili*, *kernel smoothers*, *smoothing spline* e così via (A. Buja, T. Hastie, R. Tibshirani, 1989).

Si assume, inoltre, che $E\{f_j(X_j)\} = 0$ in modo tale che tutte le costanti siano comprese in α e l'equazione (2.2) risulti identificabile.

Tale struttura ricorda un modello lineare classico, ma coinvolge funzioni non lineari per modellare l'intervento delle variabili esplicative.

I modelli additivi generalizzati (GAM) estendono i modelli additivi nello stesso modo in cui i modelli lineari generalizzati (GLM) estendono quelli di regressione lineare e permettono di modellare gli effetti non lineari facendo

uso di funzioni di lisciamiento non parametriche. In un modello lineare generalizzato la variabile risposta Y ha funzione di densità appartenente alla famiglia esponenziale:

$$P_Y(y; \theta, \phi) = \exp \left\{ \frac{y(\theta) - b(\theta)}{a(\phi)} + c(y, \phi) \right\} \quad (2.3)$$

dove θ e ϕ sono parametri scalari ignoti e la media di Y , μ , è legata all'insieme delle variabili esplicative X_1, \dots, X_p dalla relazione:

$$g(\mu) = \alpha + \sum_{j=1}^p \beta_j X_j \quad (2.4)$$

dove $g(\cdot)$, detta funzione legame, è una funzione invertibile. Tale funzione può essere costituita dal logaritmo, dall'identità, dal reciproco o da altre funzioni.

Analogamente, un modello additivo generalizzato è definito dalla relazione:

$$g(\mu) = \alpha + \sum_{j=1}^p f_j(X_j) \quad (2.5)$$

che assomiglia all'equazione (2.4) ma coinvolge stimatori non parametrici al posto dei coefficienti di regressione β . I termini f_j rappresentano delle funzioni arbitrarie diverse a seconda della covariata che modellano.

Le funzioni non parametriche $f(\cdot, \cdot)$ possono essere gestite mediante una serie di interpolazioni locali su funzioni lineari (*loess*) o polinomiali (di solito polinomi di terzo grado come nelle *spline cubiche*).

Il principio logico su cui sono costruite è la segmentazione dei dati per diversi valori della variabile sulle ascisse seguita da un'interpolazione locale. Tanto maggiore è la segmentazione tanto più la curva interpolante segue i dati osservati e risulta poco lisciata, mascherando così l'effetto dell'inquinamento. Il parametro di lisciamiento governa appunto questo aspetto.

2.2 I modelli additivi generalizzati nello studio degli effetti dell'inquinamento sulla salute umana

Negli studi di serie temporali che analizzano l'associazione tra inquinamento atmosferico e salute, il modello additivo generalizzato è il più largamente usato perché tiene conto di aggiustamenti non parametrici per gli effetti di confondimento non lineari quali stagionalità e altri trend temporali (Schwartz, 1994b; Kelsale et al., 1997; Samet et al., 2000; Katsouyanni et al., 1997; Moolgawkar, 2002).

Nel caso specifico dello studio dell'inquinamento i modelli GAM possono ad esempio assumere la seguente forma semplificata:

$$E(Y_t) = \exp\{\beta_0 + \beta X_{t-l} + f(\text{tempo}_t, df_1) + f(\text{temperatura}_t, df_2) + \gamma \times DOW_t\} \\ t = 1, \dots, T \quad (2.6)$$

dove:

- Y_t è la variabile di conteggio del numero giornaliero di morti o di ricoveri registrati il giorno t ;
- X_t è la variabile che misura il livello giornaliero di esposizione all'inquinamento nello stesso giorno;
- temperatura_t misura il livello medio giornaliero di temperatura;
- DOW_t indica il giorno della settimana;
- l indica il ritardo dell'esposizione generalmente compreso tra zero e sette giorni;

- $f(\cdot, df_i)$ ($i = 1, 2$) è una funzione di lisciamiento, costruita usando *smoothing splines*, *loess smoother* o *natural cubic splines*, con parametro di lisciamiento df_i , che indica il numero di gradi di libertà nelle funzioni di lisciamiento;
- β è il parametro di interesse che descrive la variazione nel logaritmo di Y_t per ogni unità di X_{t-l} ed indica pertanto il rischio relativo nella mortalità media o nella ricoverabilità media associata ad un aumento di 1 unità nel livello di inquinamento atmosferico al ritardo l .

L'analisi può essere condotta sulla totalità degli eventi considerati separatamente per specifiche classi di età o aggiustando per età.

La concentrazione della sostanza inquinante, il determinante in studio, viene comunemente inserita come termine lineare, anche se questa ipotesi è piuttosto forte dal momento che ad oggi non si è ancora chiarito nè se esistano o meno soglie al di sotto delle quali l'inquinamento è privo di effetti nè come sia l'andamento della curva dose-risposta, che rappresenta il modo in cui l'organismo reagisce a dosi variabili di inquinamento.

L'effetto della concentrazione può essere sia studiato in giorni singoli, sia diluito su più giorni. Nel primo caso si studiano gli effetti di esposizioni avvenute nello stesso giorno (*lag0*) o, per esempio, cinque giorni prima (*lag5*); nel secondo si calcola un'esposizione media.

La funzione di lisciamiento $f(\text{tempo}_t, df_1)$ ha il compito di smorzare eventuali fluttuazioni della mortalità o ricoverabilità nel tempo, in modo tale che per stimare β vengano usate solo variazioni di più breve termine (Samet et al., 1995; Kelsall et al., 1997). Solitamente è ragionevole una scelta di df_1 pari ad al più sette gradi di libertà per anno, in modo che siano incluse nella stima di β piccole variazioni di più di due mesi. Questa scelta permette di limitare gli effetti di confondimento dovuti a trend di lungo periodo e alla stagionalità.

Similmente, la funzione di lisciamiento $f(\text{temperatura}_t, df_2)$ provvede a controllare la variazione della mortalità o morbosità dovuta all'effetto confondente della temperatura e di altre variabili meteorologiche quali l'umidità,

tenendo conto ad esempio della temperatura registrata nei medesimi giorni di un anno, della media delle temperature registrate nei tre giorni che precedono l'evento morte e così via.

Nel seguito vengono proposti alcuni problemi relativi all'utilizzo dei modelli additivi generalizzati all'interno del programma statistico S-PLUS.

2.3 Alcuni problemi nella stima dei modelli additivi generalizzati tramite il software S-PLUS

Dal 1996 diverse pubblicazioni riguardanti gli effetti delle fluttuazioni giornaliere dell'inquinamento atmosferico sulla mortalità hanno riportato risultati ottenuti attraverso l'applicazione dei modelli additivi generalizzati.

La maggior parte di questi modelli faceva, e fa ancor oggi, uso del programma statistico S-PLUS, che fornisce per la stima dei modelli additivi generalizzati la procedura *gam*.

Come la maggior parte delle procedure, *gam* dipende da diversi parametri, quattro dei quali controllano la convergenza dell'algoritmo di stima.

Recenti rapporti hanno, comunque, messo in discussione l'adeguatezza delle conclusioni raggiunte negli studi basati sull'utilizzo di tale funzione principalmente perché l'uso della funzione *gam* nel software S-PLUS richiede una certa cautela nella stima del tasso relativo di mortalità / morbosità quando:

1. i coefficienti di regressione stimati sono piuttosto piccoli;
2. sono presenti fattori confondenti modellati usando almeno due funzioni di lisciamiento non parametriche.

In questi due casi, infatti, i parametri di convergenza di default della funzione *gam* non assicurano la convergenza della procedura di stima iterativa ed

inoltre si possono ottenere stime distorte dei coefficienti di regressione e degli errori standard.

Nel maggio del 2002, durante l'analisi dei dati relativi a al NMMAPS¹, uno studio di serie temporali condotto in alcune tra le più grandi città degli Stati Uniti, tra di loro diverse per quanto riguarda il tipo ed il livello di inquinamento presente, Francesca Dominici ed altri colleghi (2002) hanno mostrato infatti che, relativamente all'utilizzo dei modelli additivi generalizzati applicati a dati di serie temporali, i parametri di default della funzione *gam* in S-PLUS potevano far sì che il processo iterativo di stima dei coefficienti che misurano l'effetto dell'inquinamento terminasse prima della convergenza dell'algoritmo.

Questa non è un'anomalia del software, bensì dei valori di default, che non risultano essere adatti a risolvere ogni tipo di problema.

Nello stesso periodo alcuni investigatori canadesi, in studi autonomi, hanno identificato nella procedura *gam* una sorta di “scorciatoia” nel calcolo degli errori standard dei coefficienti β che, in certe circostanze, porta alla sottostima degli stessi (Ramsay et al., 2003).

La sottostima della varianza è un problema ben più grave della sovrastima poiché porta a conclusioni più precarie riguardo la significatività statistica dell'associazione tra inquinamento atmosferico e salute: la sottostima della varianza è sinonimo di intervalli di confidenza più stretti e, pertanto, è facile cogliere effetti significativi di un inquinante laddove, in realtà, non esistono. La funzione *gam* fa uso infatti di uno stimatore degli errori standard valido solo quando non esiste correlazione tra i termini relativi all'inquinamento e le variabili di regressione. In presenza di correlazione gli errori standard relativi agli effetti dell'inquinamento possono sottostimare quelli veri. Ramsay e colleghi (2003) hanno notato che questo aspetto è particolarmente negativo in presenza di valori bassi delle stime e nei casi in cui le serie temporali degli inquinanti correlino con le funzioni di lisciamiento.

Dopo che Dominici, Ramsay ed altri colleghi hanno messo in allerta la

¹National Morbidity, Mortality and Air Pollution Study

comunità scientifica, alcuni studiosi di analisi temporali che utilizzano modelli additivi generalizzati hanno dato inizio a nuove indagini relativamente alle conseguenze derivanti dalla modifica dei criteri di convergenza utilizzati dalla funzione *gam* e alla stima degli errori standard dei coefficienti di interesse.

2.3.1 Il problema della convergenza

Per valutare la convergenza dell'algoritmo, nel software S-PLUS esiste un criterio assoluto. Indicata con $\hat{\beta}^{(i)}$ la stima del coefficiente di interesse all'iterazione i -esima, la procedura sviluppata in S-PLUS prevede il raggiungimento del massimo quando:

$$\left| \hat{\beta}^{(k+1)} - \hat{\beta}^{(k)} \right| < 0.001 \quad (2.7)$$

dove $\hat{\beta}^{(k)}$ e $\hat{\beta}^{(k+1)}$ indicano rispettivamente le stime dei coefficienti che misurano l'intensità dell'effetto di un inquinante all'iterazione k e $(k+1)$ -esima. Tuttavia, essendo fisso il valore 0.001 e assumendo l'effetto dell'inquinante valori assoluti molto piccoli, si rischia che il criterio assoluto non permetta di esplorare piccole variazioni dei coefficienti.

Per evitare tale problema, la soluzione ideale sarebbe quella di relativizzare il tutto, vale a dire di utilizzare un criterio del tipo:

$$\left| \frac{\hat{\beta}^{(k+1)} - \hat{\beta}^{(k)}}{\hat{\beta}^{(k)}} \right| < 0.001 \quad (2.8)$$

Non offrendo la funzione *gam* la possibilità di modificare il criterio di convergenza, Dominici e colleghi (2002) suggeriscono valori di convergenza più stringenti.

Utilizzando tali criteri, essi riscontrano un sostanziale cambiamento nelle stime dei coefficienti rianalizzando i dati relativi al NMMAPS: l'uso di criteri di convergenza più stringenti porta a stime inferiori rispetto all'uso dei criteri di convergenza di default. Altri studi, tuttavia, sembrano essere insensibili al cambiamento dei criteri di convergenza e ad oggi non sono ancora chiari i

motivi di tali discrepanze.

2.3.2 Il problema della sottostima degli errori standard

La procedura di stima dei modelli additivi generalizzati nella funzione *gam* di S-PLUS è data da una combinazione tra l'algoritmo di *local scoring* e quello di *backfitting* (Friedman and Stuetzle; 1981; Buja et al., 1989; Hastie and Tibshirani, 1990). Il primo rappresenta una generalizzazione della procedura di Fisher per trovare la stima di massima verosimiglianza nei modelli GLM (Nelder and Wedderburn, 1972; McCullagh and Nelder, 1989) ed è utilizzato per stimare i termini parametrici del modello; il secondo è appropriato per adattare modelli additivi e riguarda i termini non parametrici.

Sia $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_J)$ lo stimatore della componente parametrica del modello da studiare.

Un'espressione esplicita per la matrice di varianze e covarianze di $\hat{\beta}$ può essere ottenuta come segue:

$$\begin{aligned}\hat{\beta} &= Hz \\ H &= \{\mathbf{X}^t W (I - \mathbf{S})\} \mathbf{X}^{-1} \mathbf{X}^t W (I - \mathbf{S}) \\ V(\hat{\beta}) &= H^t W^{-1} H\end{aligned}\tag{2.9}$$

dove:

- \mathbf{X} è la matrice $T \times J$ con colonne $\mathbf{X}_j = [X_{j1}, \dots, X_{jT}]^t$;
- \mathbf{z} è il risultato dell'ultima iterazione dell'algoritmo IRLS definito da:

$$z_t = \hat{\eta} + \frac{y_t - \hat{\lambda}_t}{\hat{\lambda}_t};\tag{2.10}$$

- W è una matrice diagonale di pesi e $W^{-1} = \widehat{cov}(\mathbf{z})$;

- \mathbf{S} è la matrice di lisciamento $T \times T$ usata nella stima della parte non parametrica del modello.

Dal momento che il calcolo della matrice \mathbf{S} risulta particolarmente oneroso, la versione corrente della funzione *gam* in S-PLUS approssima la varianza di $\hat{\beta}$ assumendo che le componenti di lisciamento del modello siano lineari (Hastie and Tibshirani, 1990; Chambers and Hastie, 1992). Nello studio delle serie temporali relative all'effetto dell'inquinamento sulla salute, tale assunzione di linearità è inadeguata e ciò può portare ad una sottostima degli errori standard dal momento che dal modello è stato escluso un numero più o meno grande di termini non lineari. Si è inoltre dimostrato che il grado di distorsione negli errori standard è proporzionale alla dimensione degli errori standard stessi.

Tuttavia, se \mathbf{S} è una matrice simmetrica, H può essere ricalcolata come segue:

$$H = \{\mathbf{X}^t (W\mathbf{X} - W\mathbf{S}\mathbf{X})\}^{-1} (W\mathbf{X} - W\mathbf{S}\mathbf{X})^t \quad (2.11)$$

Si noti che la simmetria, in questo caso, si intende rispetto alla matrice dei pesi W e ciò implica che:

$$W\mathbf{S} = \mathbf{S}^t W \quad (2.12)$$

La componente onerosa nella stima della varianza di $\hat{\beta}$ è data dal calcolo della matrice $\mathbf{S}\mathbf{X}$ di dimensione $T \times J$, avente come colonna j -esima il vettore \mathbf{X}_j risultante dall'adattamento del modello additivo:

$$Y_t = \sum_{l=1}^L f_l(X_{tl}, df_l) \quad (2.13)$$

Dominici e colleghi (2002) dimostrano che, per ogni matrice simmetrica \mathbf{S} , il calcolo della varianza di $\hat{\beta}$ richiede l'adattamento di un ulteriore modello additivo generalizzato avente \mathbf{X}_t come variabile dipendente e funzioni di

lisciamento come variabili indipendenti. Questo procedimento è stato adottato nello sviluppo della funzione *gam.exact* e fornisce i risultati dopo aver compiuto i seguenti due passi:

1. si adatta il modello semi-parametrico (2.1) mediante l'utilizzo della funzione *gam* e si estraggono i pesi \mathbf{w} e i gradi di libertà df_i^{*2} usati nell'algoritmo di *backfitting*. I pesi \mathbf{w} sono gli elementi che stanno sulla diagonale della matrice W ;
2. si liscia ogni colonna della matrice \mathbf{X} rispetto a $\sum_{l=J+1}^L f_l(X_{lt}, df_l^*)$ usando la funzione *gam* con pesi \mathbf{w} . Le colonne della matrice \mathbf{SX} rappresentano i corrispondenti valori adattati.

Questi autori sostengono che tale funzione porti ad una correzione nella sottostima della matrice di varianze e covarianze delle stime degli effetti dell'inquinamento, tenendo in considerazione la correlazione tra i predittori non lineari; inoltre la funzione *gam.exact* non coinvolge matrici di grandi dimensioni e risulta semplice da implementare.

Attraverso l'uso della funzione *gam.exact* si dovrebbe ottenere una stima precisa della varianza rispetto a quella che si ottiene utilizzando la funzione *gam*. La linea di sviluppo seguita in questo progetto prevede l'implementazione il confronto delle due funzioni *gam* e *gam.exact*.

2.4 Ulteriori elementi di attenzione

La convergenza dell'algoritmo di *backfitting* e la sottostima degli errori standard dipendono in modo critico dal numero di gradi di libertà dei termini di lisciamento del modello.

Negli studi di serie temporali relativi all'inquinamento e alla salute, un maggior controllo nei confronti dei fattori confondenti conduce ad un numero maggiore di gradi di libertà, il che provoca un aumento della correlazione

²I gradi di libertà df_i^* sono leggermente diversi da quelli richiesti nella chiamata alla funzione *gam* a causa del cambiamento dei pesi nell'algoritmo IRLS.

interna alla matrice funzione delle variabili esplicative e rende inadeguata l'approssimazione lineare dei termini di lisciamiento del modello.

Così, l'uso di diverse funzioni di lisciamiento comporta una riduzione della distorsione nelle stime, ma allo stesso tempo un rallentamento della convergenza e una sottostima degli errori standard.

La scelta del numero di gradi di libertà è un aspetto critico nell'analisi delle serie temporali relative all'inquinamento e alla salute umana. La soluzione ideale sarebbe quella di scegliere un numero di gradi di libertà abbastanza grande da rimuovere completamente l'effetto di confondimento, ossia da eliminare la distorsione dello stimatore $\hat{\beta}$, ma non troppo grande per non rischiare di eliminare tutta la variazione nei dati e così inflazionare la varianza dei coefficienti di interesse.

La scelta del numero di gradi di libertà si basa solitamente sull'osservazione delle serie temporali di mortalità (Kellsall et al., 1997; Dominici et al., 2002) oppure su criteri di ottimalità, basati sul criterio di informazione di Akaike, oppure ancora sull'ispezione visiva della funzione di autocorrelazione parziale dei residui, definita come la somma dei valori assoluti delle funzioni di autocorrelazione dei residui (Tuloumi et al., 1997; Burnet et al., 2001).

Selezionando i gradi di libertà che minimizzano AIC³ o PACF⁴ si giunge, rispetto a quanto non avvenga ad esempio selezionandoli attraverso il criterio BIC⁵, ad un modello più ricco, che elimina la distorsione, ma può anche inflazionare la varianza dei coefficienti di interesse.

D'altra parte, scegliendo i gradi di libertà che minimizzano l'MSE⁶, si ottiene un modello più parsimonioso; tuttavia in questo caso i residui potrebbero risultare autocorrelati.

In conclusione, quando l'obiettivo dell'inferenza è la stima dei termini lineari del modello semi-parametrico, l'ispezione visiva dell'MSE e il confronto con AIC e PACF costituiscono un buon strumento diagnostico per la scelta

³Asymptotic Information Criterion (Akaike, 1974)

⁴Partial AutoCorrelation Function

⁵Bayesian Information Criterion (Schwartz, 1978)

⁶Mean Square Error

del numero di gradi di libertà.

Capitolo 3

Analisi descrittive e grafiche

In questo capitolo si riportano alcune analisi descrittive e grafiche per dare una visione abbastanza generale delle variabili prese in considerazione e del loro andamento.

Il dataset utilizzato per l'analisi, che contiene dati relativi alla città di Verona, è costituito da una matrice di 1461 righe e 22 colonne, di cui si riporta una parte a titolo di esempio in Tabella 3.1.

row,	GG,MM,AA,	WDAY,	DAY,	HOL,	TEMP24,	HUM,	CO8H,	NO2,	O3,	TSP,	SO2,	INFLU,	RES,	RES014,	RES1564,	RES6574,	RES75,	O3NEW,	stagione	
1,	1,7,1995,	7,	1,1,	27.29445179,	52.66654,	1.4786549,	47.25212,	186.6139137,	56.79690,	5.5366337,	0,	2,0,0,1,	1,186.6139137,	calda						
2,	2,7,1995,	1,	2,2,	25.97407182,	71.01066,	1.2569096,	39.61336,	187.2541137,	63.51241,	4.0715509,	0,	4,0,0,0,	4,187.2541137,	calda						
3,	3,7,1995,	2,	3,1,	24.63495104,	70.74581,	1.1825793,	43.02449,	151.4906904,	55.04171,	3.7327176,	0,	9,0,4,2,	3,151.4906904,	calda						
4,	4,7,1995,	3,	4,1,	22.91627677,	65.92646,	1.1791897,	41.94411,	126.5750259,	35.87199,	4.0918492,	0,	8,1,3,1,	3,121.0891614,	calda						
5,	5,7,1995,	4,	5,1,	21.99607809,	80.00287,	1.5313849,	49.43704,	113.0193673,	44.70899,	4.1548734,	0,	5,1,2,1,	1,108.1000505,	calda						
6,	6,7,1995,	5,	6,1,	20.95565597,	86.14700,	1.4662859,	44.23556,	105.3411869,	39.12291,	4.0470008,	0,	4,1,1,1,	1,102.5409417,	calda						
7,	7,7,1995,	6,	7,1,	23.98678127,	67.88762,	1.4755206,	51.59159,	139.6746826,	44.51057,	4.8863682,	0,	5,0,0,0,	5,133.1758066,	calda						
8,	8,7,1995,	7,	8,1,	26.65335649,	62.61811,	1.4966302,	50.71763,	145.7429455,	51.21591,	5.2431908,	0,	7,2,1,2,	2,145.4053311,	calda						
9,	9,7,1995,	1,	9,2,	26.03292819,	63.51705,	1.2039238,	37.97736,	162.3513410,	46.40822,	4.9425117,	0,	1,0,0,0,	1,144.0144584,	calda						
10,	10,7,1995,	2,	10,1,	26.81639331,	65.07515,	1.3864565,	51.27171,	151.2677614,	55.05698,	6.9361607,	0,	9,0,1,4,	4,166.5306182,	calda						
.....																				
.....																				
.....																				
1452,	21,6,1999,	2,	1452,1,	19.31479708,	92.00477,	1.2034119,	55.88795,	59.5502002,	18.96113,	2.3293261,	1,	10,2,3,0,	5,	59.5502002,	calda					
1453,	22,6,1999,	3,	1453,1,	21.52841008,	40.14004,	0.6417204,	37.33290,	99.9110820,	19.65811,	1.8248763,	1,	3,1,1,0,	1,	99.9110820,	calda					
1454,	23,6,1999,	4,	1454,1,	21.75717831,	40.55390,	0.7746199,	51.20513,	119.3179364,	19.25620,	3.2228304,	1,	8,3,4,0,	1,119.3179364,	calda						
1455,	24,6,1999,	5,	1455,1,	22.40072218,	64.53855,	1.3267747,	49.29369,	133.4691479,	24.01811,	7.3790056,	0,	2,0,0,0,	2,133.4691479,	calda						
1456,	25,6,1999,	6,	1456,1,	23.21824177,	63.10411,	1.1401672,	49.46770,	132.1603287,	22.29344,	3.6136685,	0,	3,1,0,0,	2,132.1603287,	calda						
1457,	26,6,1999,	7,	1457,1,	24.03968096,	62.73701,	1.1849138,	47.09355,	148.8582028,	26.00732,	4.2085050,	0,	1,1,0,0,	0,148.8582028,	calda						
1458,	27,6,1999,	1,	1458,2,	24.43981584,	79.35555,	1.1241936,	36.52219,	143.9896052,	28.57142,	10.5777052,	0,	2,0,0,0,	2,143.9896052,	calda						
1459,	28,6,1999,	2,	1459,1,	23.85885286,	77.03148,	1.6847633,	53.71951,	134.2130878,	22.86833,	2.3922601,	0,	8,1,3,2,	2,134.2130878,	calda						
1460,	29,6,1999,	3,	1460,1,	25.06734220,	82.60180,	1.5264126,	54.76936,	125.1715167,	26.34309,	3.0940456,	0,	5,1,2,1,	1,125.1715167,	calda						
1461,	30,6,1999,	4,	1461,1,	26.75628869,	69.60199,	1.6718615,	65.69234,	133.8064930,	27.90496,	2.3108669,	0,	3,0,0,0,	3,133.8064930,	calda						

Tabella 3.1: Dataset utilizzato nell'analisi.

In ogni colonna sono riportate le diverse variabili di cui si tiene conto nell'analisi. Queste sono misurate giornalmente durante l'intero periodo di studio, vale a dire i quattro anni che vanno dal 01 luglio 1995 al 30 giugno 1999. Tra le variabili sono presenti:

1. diversi tipi di inquinanti, i cui valori rappresentano l'oggetto di diretto interesse nell'analisi;
2. fattori confondenti quali i giorni della settimana, i giorni di vacanza, la temperatura, il livello di umidità e le epidemie di influenza;
3. il numero di ricoveri dovuti a cause respiratorie (RESnn) suddivisi in quattro classi di età (0-14, 15-64, 65-74, 75 e più) che costituisce la variabile risposta o dipendente dell'analisi.

Gli inquinanti che vengono presi in considerazione, con relative unità di misura, sono:

- SO_2 - anidride solforosa ($\mu g/m^3$);
- NO_2 - biossido di azoto ($\mu g/m^3$);
- CO - monossido di carbonio (mg/m^3);
- TSP - polveri sospese totali ($\mu g/m^3$);
- O_3 - ozono ($\mu g/m^3$).

Tra questi il TSP è un inquinante complesso sui cui effetti c'è stato e c'è tutt'oggi dibattito, trattandosi non di un'entità fisica ben definita, ma di una miscela di sostanze organiche ed inorganiche in forma solida, che possono penetrare più o meno in profondità nell'apparato respiratorio a seconda della dimensione delle particelle stesse.

La Tabella 3.2 mostra alcune statistiche descrittive.

	SO_2 $\mu g/m^3$	NO_2 $\mu g/m^3$	CO mg/m^3	TSP $\mu g/m^3$	O_3 $\mu g/m^3$	Temp. $^{\circ}C$	Umidità %
media	6.64	57.79	2.54	45.64	73.85	14.86	76.8
d.s. ^a	4.55	18.46	1.29	20.59	51.5	7.73	17.94
min	0.61	16.48	0.64	14.8	0.0	-5.64	11.65
1st Qu ^b	3.36	45.48	1.53	31.93	25.99	8.273	64.42
mediana	5.34	54.56	2.2	39.26	71.38	14.54	78.42
3rd Qu ^c	8.65	67.56	3.28	54.14	112.9	21.28	92.7
95°	15.68	92.59	4.89	87.65	161.53	26.75	99.99
max	33.83	161.3	10.17	153.0	225.6	31.71	100.0

^adeviazione standard

^b1°quartile

^c3°quartile

Tabella 3.2: Statistiche descrittive della concentrazione giornaliera degli inquinanti e delle variabili meteorologiche.

3.1 Analisi grafiche

3.1.1 Gli inquinanti

Come prima analisi vengono costruiti istogrammi, boxplot e serie storiche giornaliere che evidenziano l'andamento della concentrazione dei diversi inquinanti, mentre per rilevare l'eventuale presenza di ciclicità mensili e stagionali nei dati sono riportati i boxplot relativi ad ogni stagione e le serie storiche mensili.

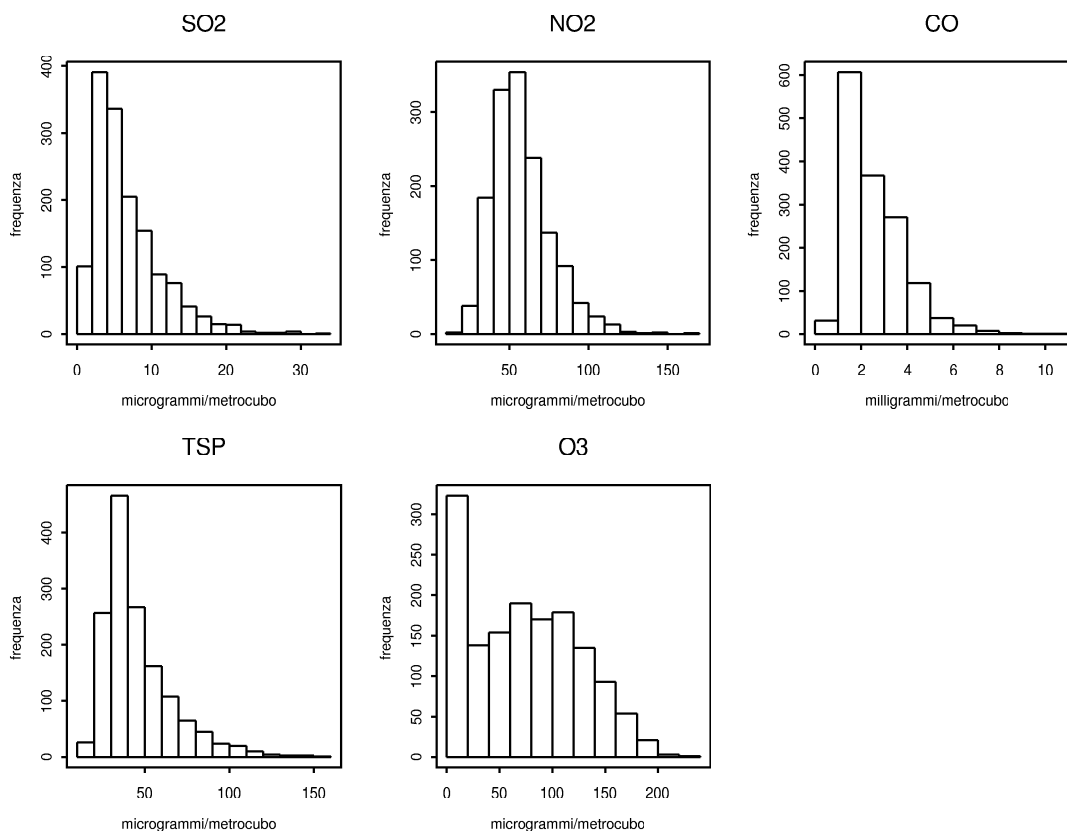


Figura 3.1: Istogrammi della concentrazione degli inquinanti.

Dall'osservazione degli istogrammi in Figura 3.1 risulta evidente che la distribuzione della concentrazione di tutti gli inquinanti è asimmetrica verso sinistra, con la coda destra più pesante. Anche dai boxplot in Figura 3.2 si possono trarre analoghe conclusioni e inoltre emergono ancor più chiaramente i numerosi valori anomali che caratterizzano tutti gli inquinanti eccezion fatta per l' O_3 . Questi valori anomali riguardano esclusivamente la zona oltre il baffo superiore del grafico, ossia oltre il valore calcolato nel modo seguente:

$$(3^{\circ} \text{quartile}) + (1.5 \times (\text{scarto interquartile})) \quad (3.1)$$

dove: scarto interquartile = 3° quartile - 1° quartile.

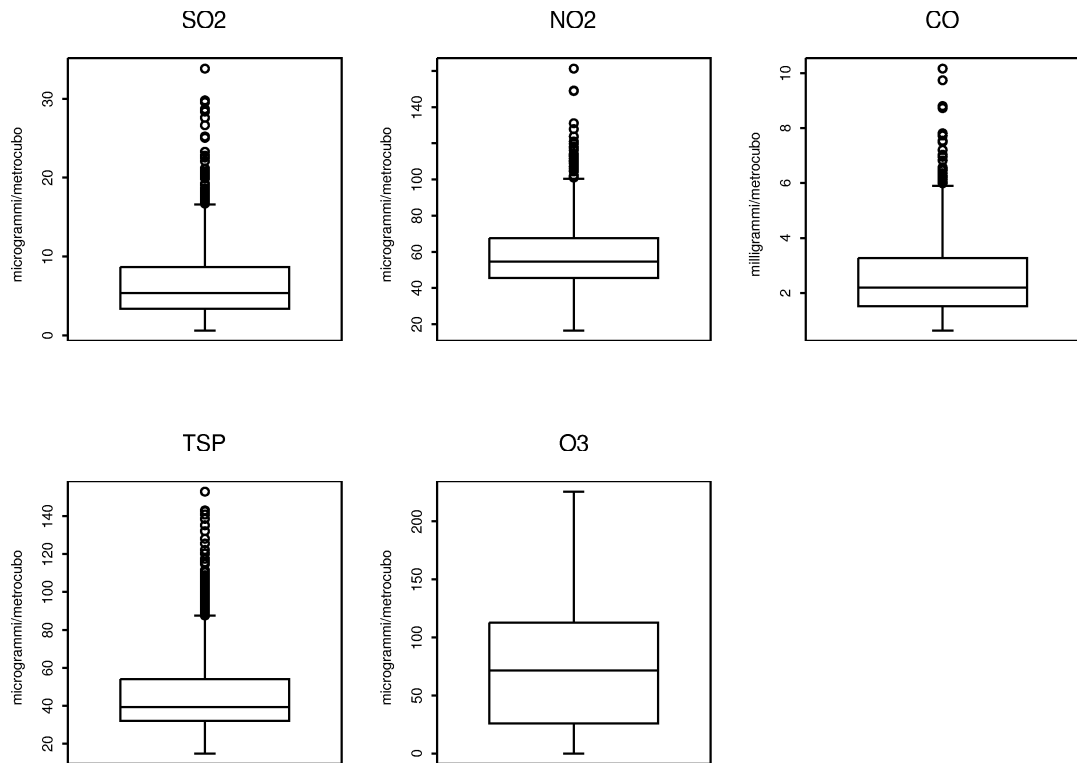


Figura 3.2: Boxplot della concentrazione degli inquinanti.

Le figure 3.3 e 3.4 rappresentano le serie storiche giornaliere della concentrazione di SO_2 , NO_2 , CO , TPS e O_3 .

Dai grafici emerge un andamento ciclico stagionale per tutti gli inquinanti; per l' SO_2 , l' NO_2 , il TPS e il CO si evidenziano valori più elevati nella stagione invernale e più bassi durante il periodo estivo, mentre accade il contrario per l' O_3 , il quale presenta dei picchi nei mesi più caldi.

Si noti poi che, mentre l'anidride solforosa presenta un picco nel mese di gennaio del 1996 e in generale il trend pare leggermente in calo nell'ultimo anno di studio, il biossido di azoto mostra un picco verso il mese di marzo del 1998 e la concentrazione media giornaliera è lievemente in ascesa.

Il monossido di carbonio presenta dei valori particolarmente elevati negli ultimi tre anni di studio, anche se in generale il trend pare mantenersi costante nei quattro anni considerati.

L'andamento delle polveri sospese totali si presenta invece piuttosto irregolare, a differenza dell'ozono, che invece presenta un andamento ciclico decisamente regolare, sebbene contrapposto a quello degli altri inquinanti.

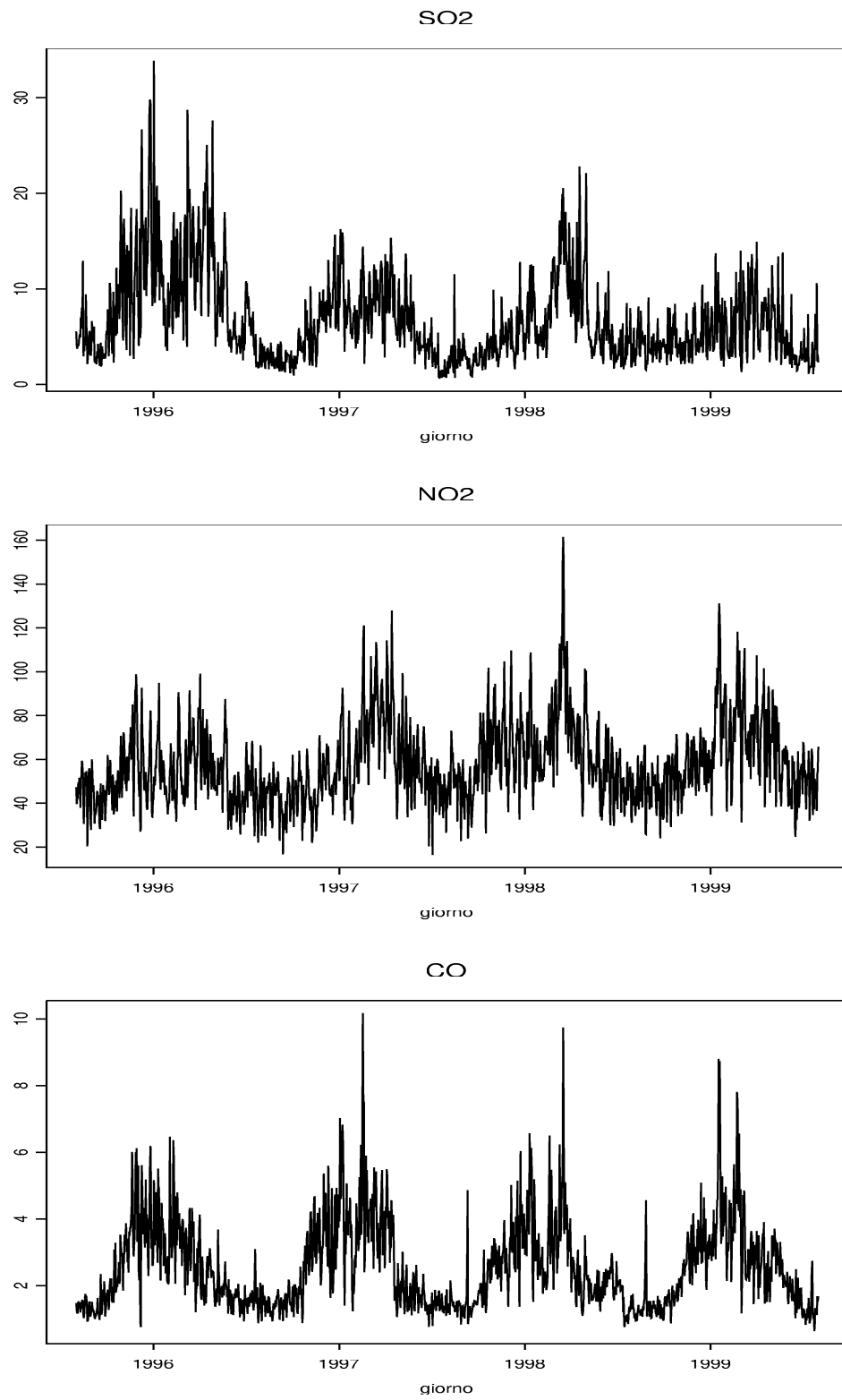


Figura 3.3: Serie storiche giornaliere della concentrazione media rispettivamente di anidride solforosa ($\mu\text{g}/\text{m}^3$), biossido di azoto ($\mu\text{g}/\text{m}^3$) e monossido di carbonio (mg/m^3).

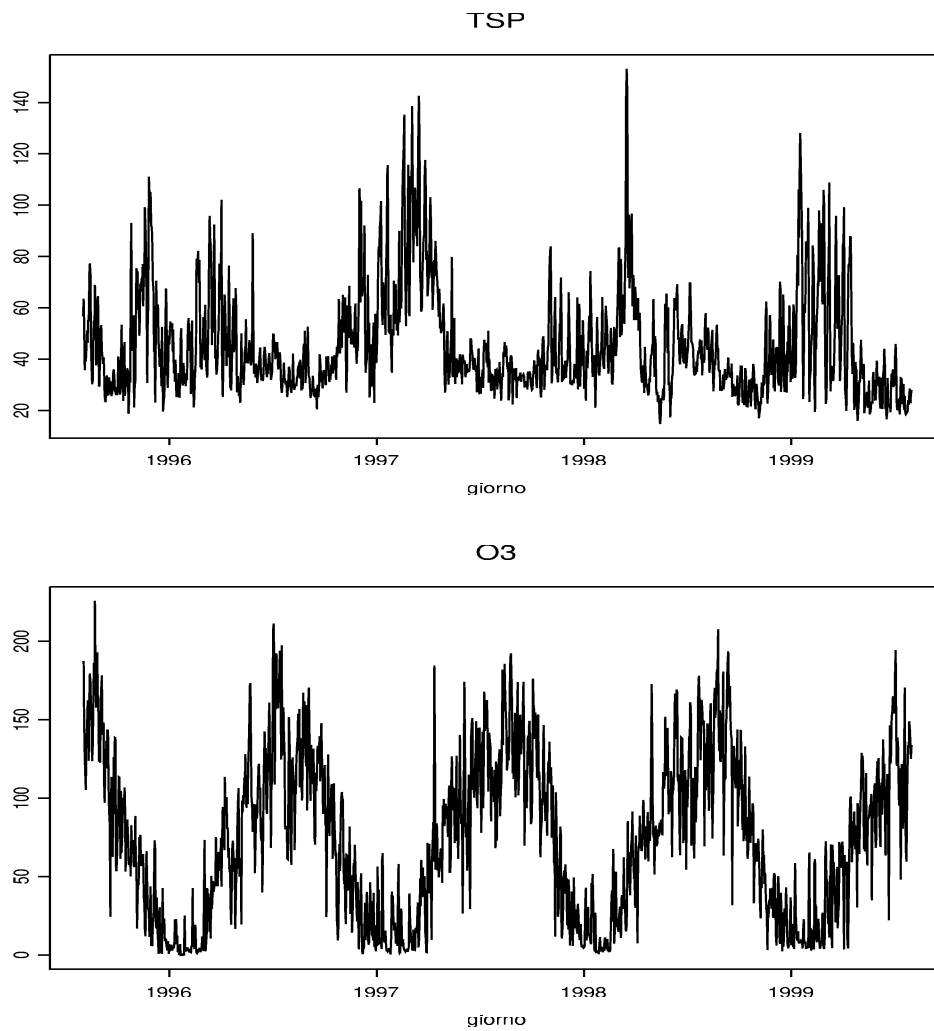


Figura 3.4: Serie storiche giornaliere della concentrazione media rispettivamente di polveri sospese totali ($\mu\text{g}/\text{m}^3$) e ozono ($\mu\text{g}/\text{m}^3$).

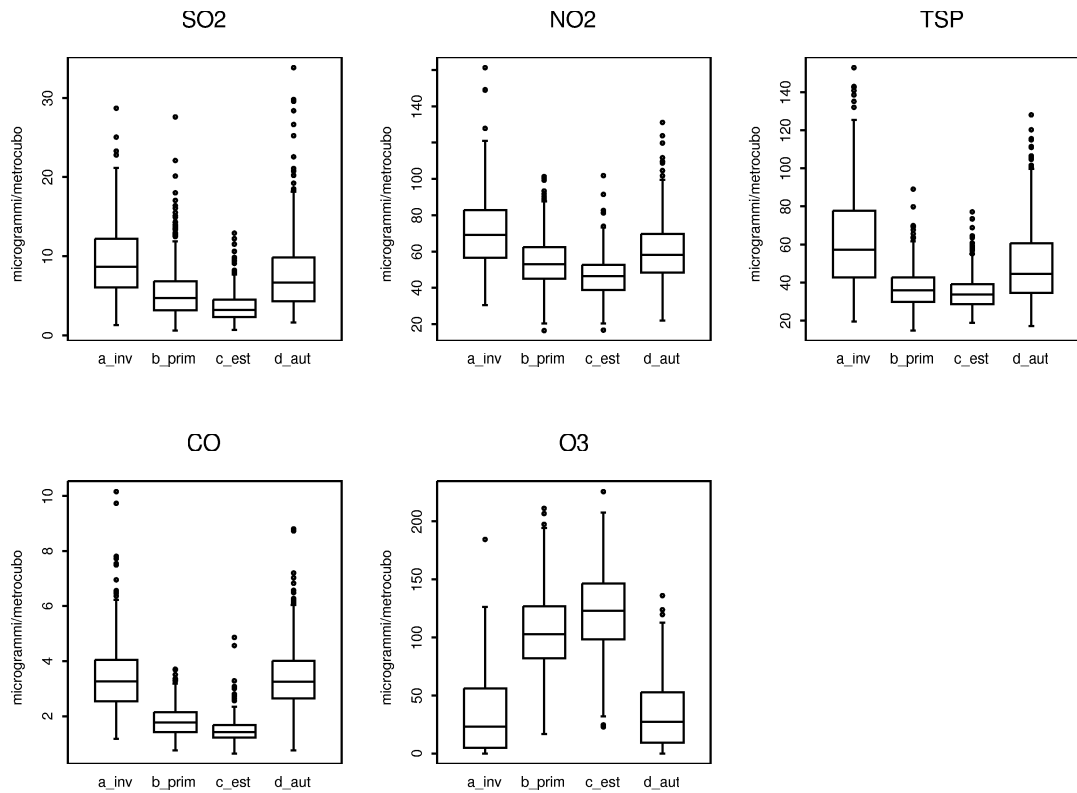


Figura 3.5: Boxplot stagionali della concentrazione media giornaliera degli inquinanti.

L'andamento della concentrazione degli inquinanti che emerge dall'osservazione delle serie storiche giornaliere è confermato dai boxplot in Figura 3.5, che sono stati disegnati suddividendo i valori per stagione. Per tutti gli inquinanti si notano valori più bassi in estate e più alti in inverno, ad eccezione fatta per l'ozono che ha un comportamento decisamente opposto rispetto agli altri inquinanti.

Dalle serie storiche delle concentrazioni medie mensili dei vari inquinanti nei quattro anni di studio emergono analoghe caratteristiche (Figure 3.6 e 3.7).

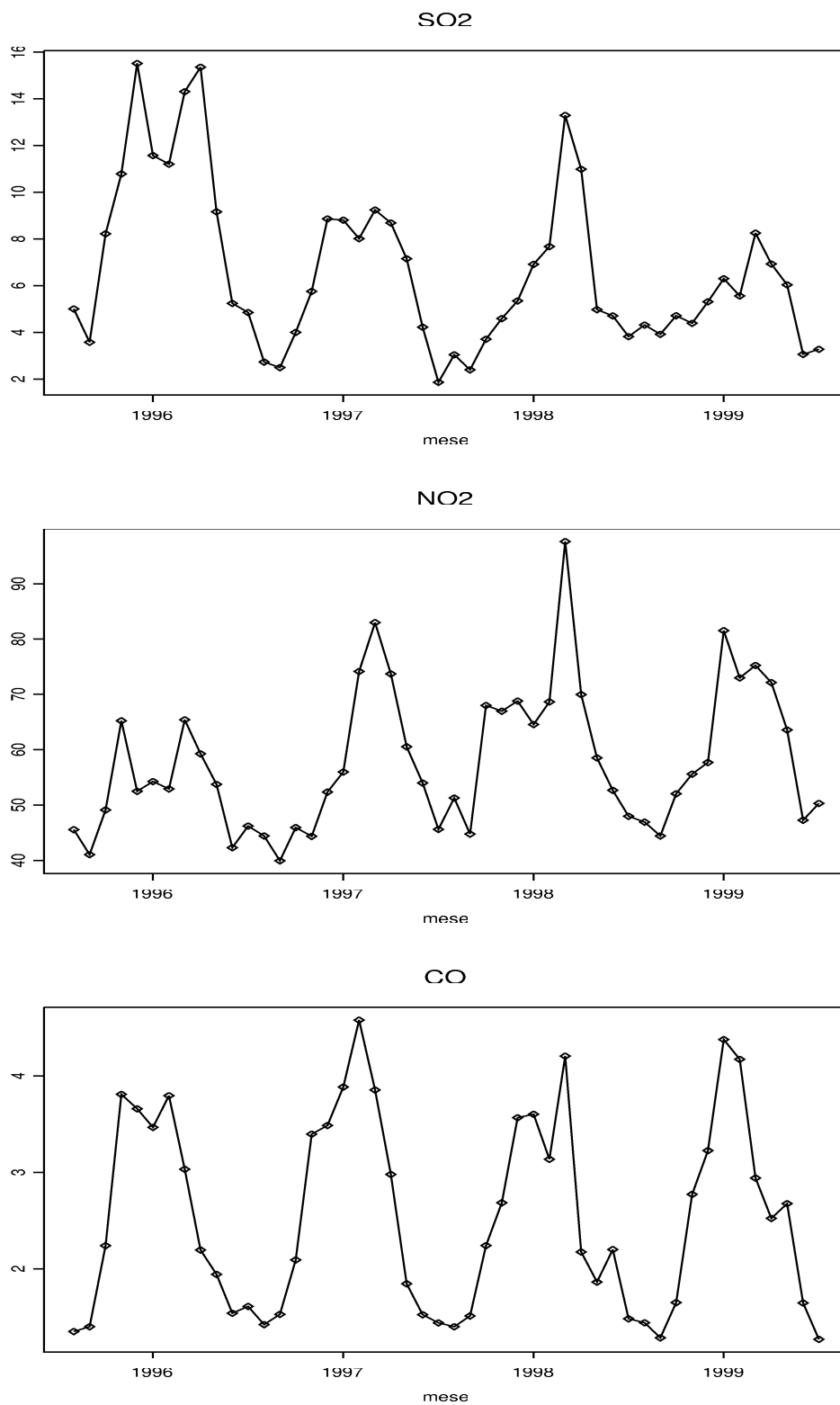


Figura 3.6: Serie storiche della concentrazione media mensile rispettivamente di anidride solforosa ($\mu\text{g}/\text{m}^3$), biossido di azoto ($\mu\text{g}/\text{m}^3$) e monossido di carbonio (mg/m^3).

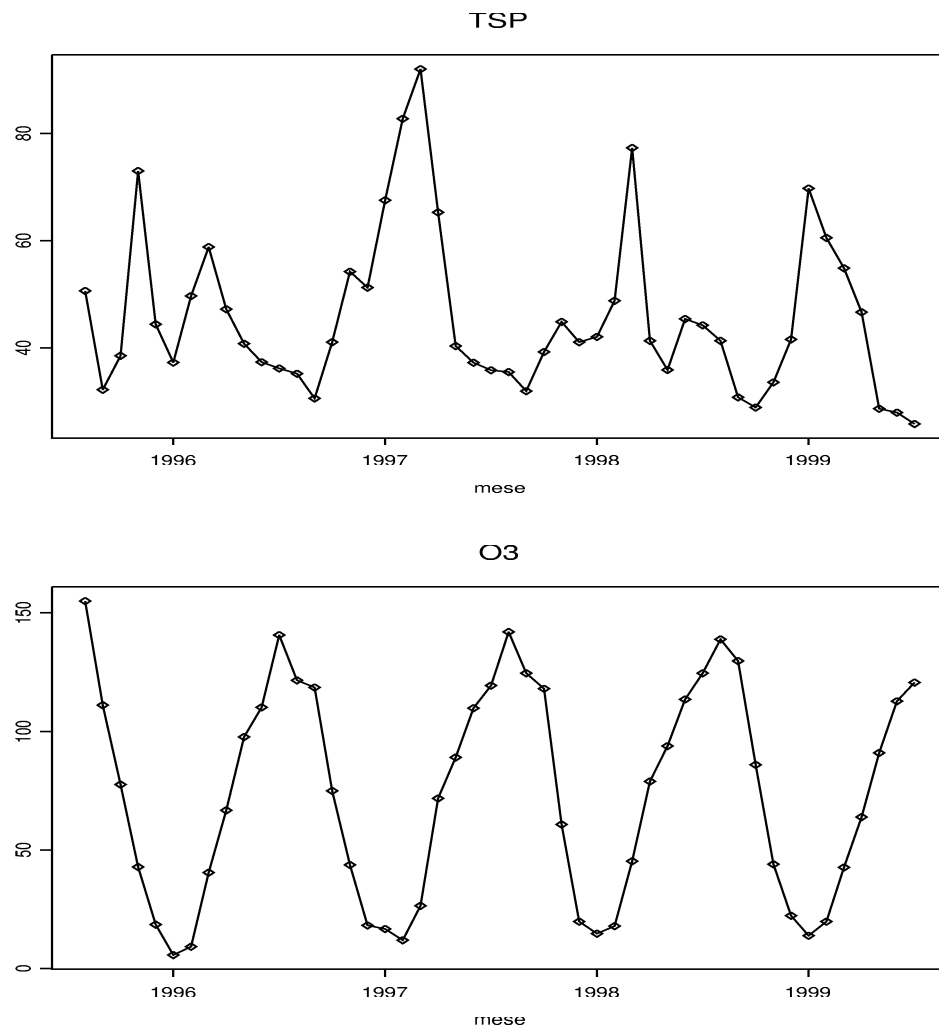


Figura 3.7: Serie storiche della concentrazione media mensile rispettivamente di polveri sospese totali ($\mu\text{g}/\text{m}^3$) e ozono ($\mu\text{g}/\text{m}^3$).

3.1.2 Le variabili confondenti

La temperatura e l'umidità sono considerate le principali variabili confondenti di cui si tiene conto nell'analisi svolta in questo progetto.

Anche per queste variabili meteorologiche sono state condotte delle brevi ma significative analisi grafiche.

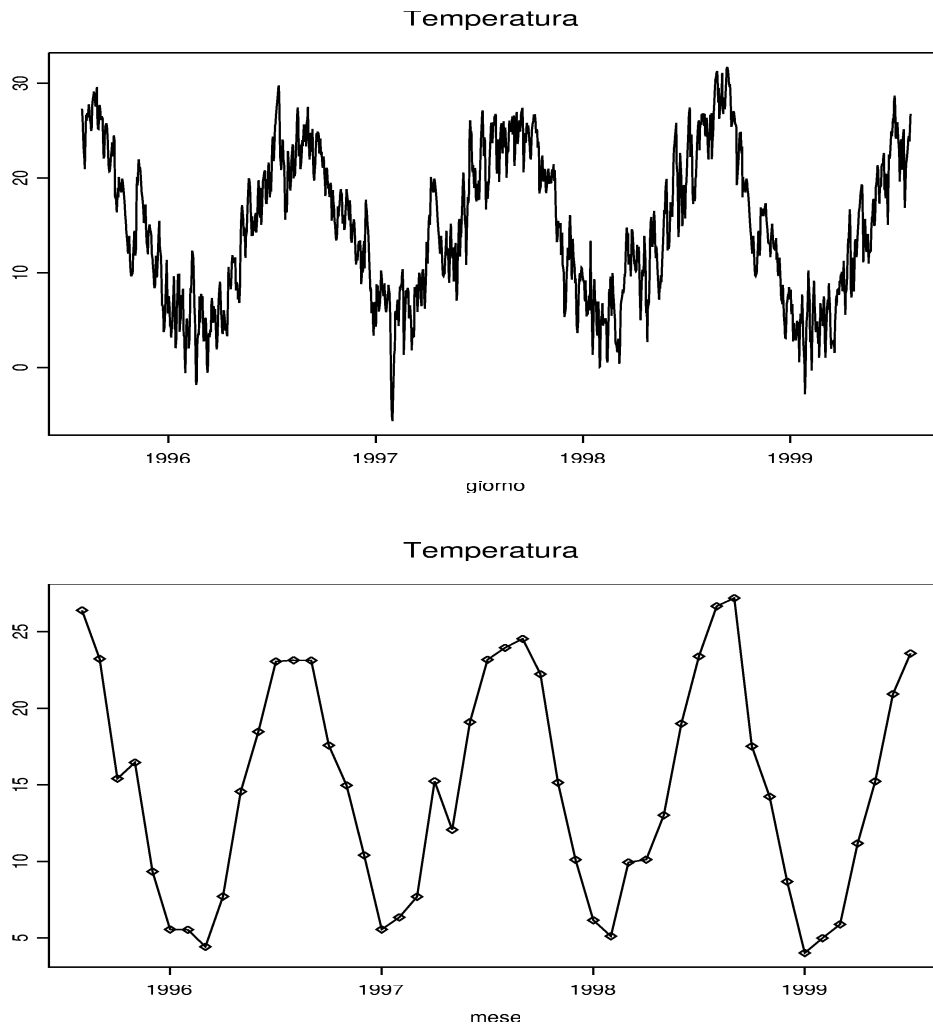


Figura 3.8: Serie storiche della temperatura media giornaliera e mensile ($^{\circ}\text{C}$).

Dai grafici riportati in Figura 3.8 relativi alla temperatura, si può notare come questa variabile presenti un andamento regolare e ciclico nei quattro anni di studio, anche se queste proprietà si potrebbero senza dubbio estendere a qualsiasi altro anno.

I valori massimi della temperatura si riscontrano nei mesi estivi, mentre i minimi nella stagione invernale; si osservi un picco in corrispondenza del mese di febbraio del 1997, che rappresenta un periodo di rigide condizioni climatiche.

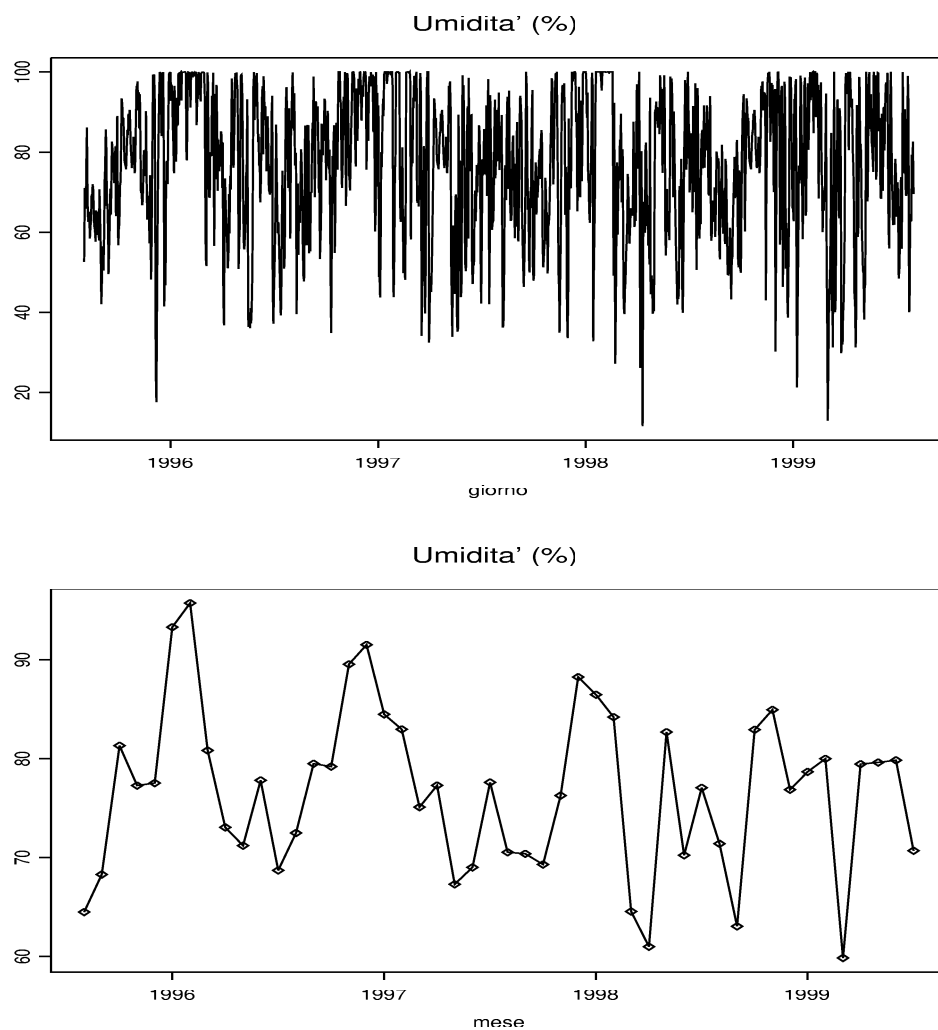


Figura 3.9: Serie storiche dell'umidità media giornaliera e mensile.

I grafici in Figura 3.9 rappresentano le serie storiche dell'umidità media, giornaliera e mensile rispettivamente. Dal confronto di questi due grafici con quelli rappresentati nella Figura 3.8 emerge chiaramente come l'andamento dell'umidità sia più irregolare rispetto a quello della temperatura.

Si osservino, inoltre, i tre picchi di bassa percentuale di umidità nei tre periodi: dicembre 1995, aprile 1998 e marzo 1999.

L'umidità viene misurata in percentuale ed è per questo motivo che il suo andamento è troncato a 100; la maggior parte dei valori si concentra comunque tra il 75 e l'85%.

3.1.3 Coefficienti di correlazione

In natura è altamente improbabile riscontrare la presenza di un solo tipo di inquinante in una determinata zona: solitamente diversi inquinanti interagiscono tra loro e questo rende estremamente difficile valutare singolarmente il loro contributo. Inoltre anche la temperatura e l'umidità influenzano plausibilmente il contributo di ogni inquinante. Al riguardo sono state calcolate le correlazioni tra inquinanti e variabili meteorologiche e riportati in Tabella 3.3.

	SO_2	NO_2	CO	TSP	O_3	Temp.	Umidità
SO_2	1						
NO_2	0.46	1					
CO	0.50	0.66	1				
TSP	0.41	0.66	0.64	1			
O_3	-0.37	-0.31	-0.66	-0.36	1		
Temp.	-0.47	-0.37	-0.56	-0.33	0.83	1	
Umidità	-0.09	-0.11	0.13	0.09	-0.40	-0.23	1

Tabella 3.3: Matrice di correlazione tra inquinanti e variabili meteorologiche.

Dall'ispezione delle correlazioni tra gli inquinanti, si nota che sono abbastanza significative quelle tra il monossido di carbonio e il biossido di azoto, tra le polveri sospese totali e il biossido di azoto, tra le polveri sospese totali e il monossido di carbonio e tra l'ozono e il monossido di carbonio; le altre possono essere considerate non rilevanti dal momento che il coefficiente di correlazione assume un valore inferiore a 0.5.

E' singolare il comportamento dell'ozono che correla negativamente con gli altri inquinanti. Questo particolare comportamento è dovuto al fatto che l'ozono trova origine nell'atmosfera in seguito alla scomposizione da parte delle radiazioni solari delle sostanze contenenti ossigeno, mentre gli altri sono prevalentemente di origine antropogena.

Tutti gli altri inquinanti sono correlati tra loro positivamente.

Per quanto riguarda invece le correlazioni tra gli inquinanti e le variabili meteorologiche, emerge che nessuna di queste risulta particolarmente significativa, tranne quella positiva tra la temperatura e l'ozono e quella negativa tra la temperatura e il monossido di carbonio, che assumono un coefficiente di correlazione rispettivamente pari a 0.83 e -0.56.

Anche dall'osservazione dei segni della correlazione tra l'ozono e le variabili meteorologiche, perlopiù contrari rispetto a quanto avviene per gli altri inquinanti, si può riscontrare il comportamento singolare dell' O_3 .

Si riportano in Figura 3.10 i grafici di dispersione tra coppie di inquinanti e a questi è stata aggiunta la retta stimata dal lisciatore non parametrico *lowess*.

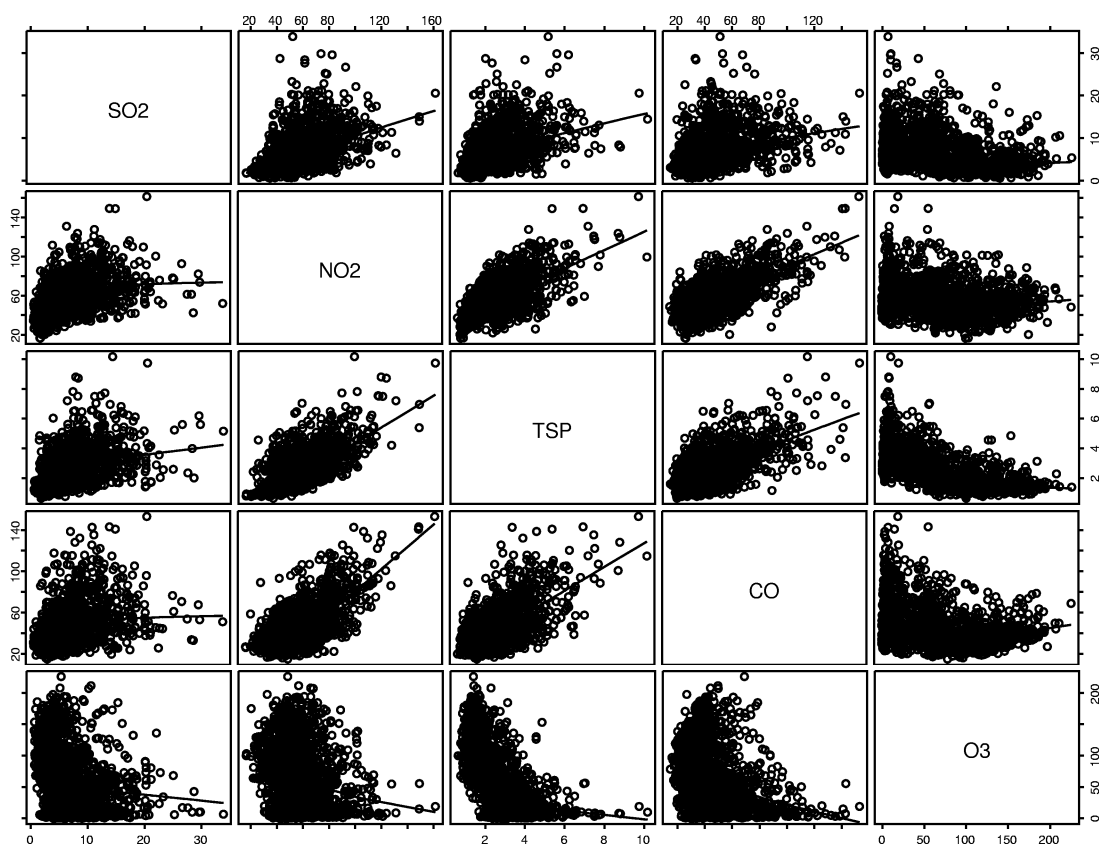


Figura 3.10: Grafici di dispersione relativi ai cinque inquinanti considerati.

3.1.4 La variabile risposta

Nell'analisi condotta in questo progetto la variabile risposta è rappresentata dal numero giornaliero di ricoveri dovuti a cause respiratorie. Anche per tale variabile sono state condotte delle semplici analisi descrittive.

Si riportano in Figura 3.11 gli istogrammi della distribuzione dei ricoveri, separatamente per le tre classi d'età 0-64, 65-74, 75 e più: come si può notare, la distribuzione è decisamente asimmetrica con la coda destra più pesante e questo per tutte e tre le classi d'età.

In Figura 3.12 sono rappresentati i grafici relativi ai ricoveri ospedalieri avvenuti durante il periodo di studio, vale a dire dal 01/07/1995 al 30/06/1999 e relativi alle persone appartenenti alle fasce di età 0-64, 65-74, 75 e più, rispettivamente. In tutti i tre grafici risulta evidente il pattern stagionale che tuttavia è leggermente irregolare da un anno all'altro. Si nota, infatti, che i ricoveri giornalieri sono più frequenti nella stagione invernale rispetto a quella estiva e questo risulta generalmente evidente per le tre classi d'età considerate, indice che la stagione, e quindi la temperatura e l'umidità, hanno influenza sulla salute dell'uomo.

Gli stessi pattern stagionali emergono ancor più chiaramente osservando la Figura 3.13 in cui sono rappresentate le serie storiche mensili dei ricoveri.

Dalla Figura 3.13 si può notare che il numero di ricoveri avvenuti durante il periodo invernale nell'ultimo anno di studio è diminuito rispetto a quello avvenuto nello stesso periodo degli anni precedenti. Questo si nota in modo ancor più evidente nel secondo grafico di Figura 3.13, mentre il numero di ricoveri giornalieri che riguardano le persone più anziane (terzo grafico di Figura 3.13) presenta nel periodo invernale dei picchi più alti nei due anni di studio intermedi e un picco piuttosto basso nell'autunno del 1998.

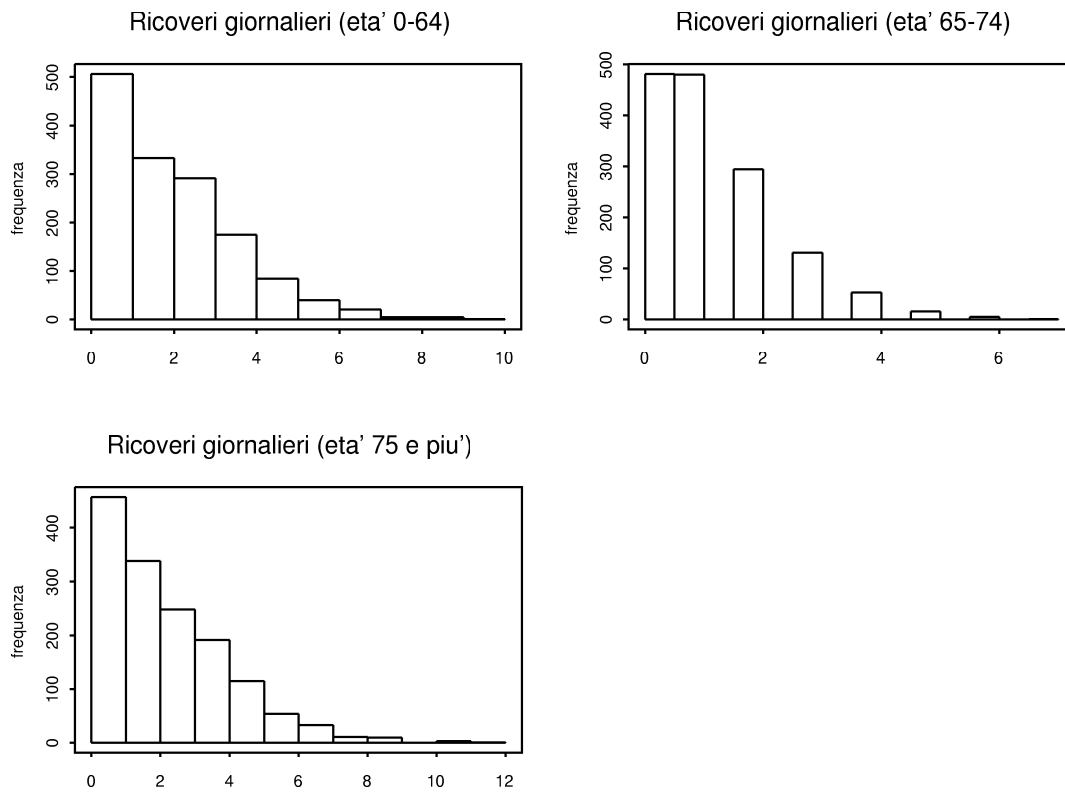


Figura 3.11: Istogrammi della distribuzione dei ricoveri giornalieri.

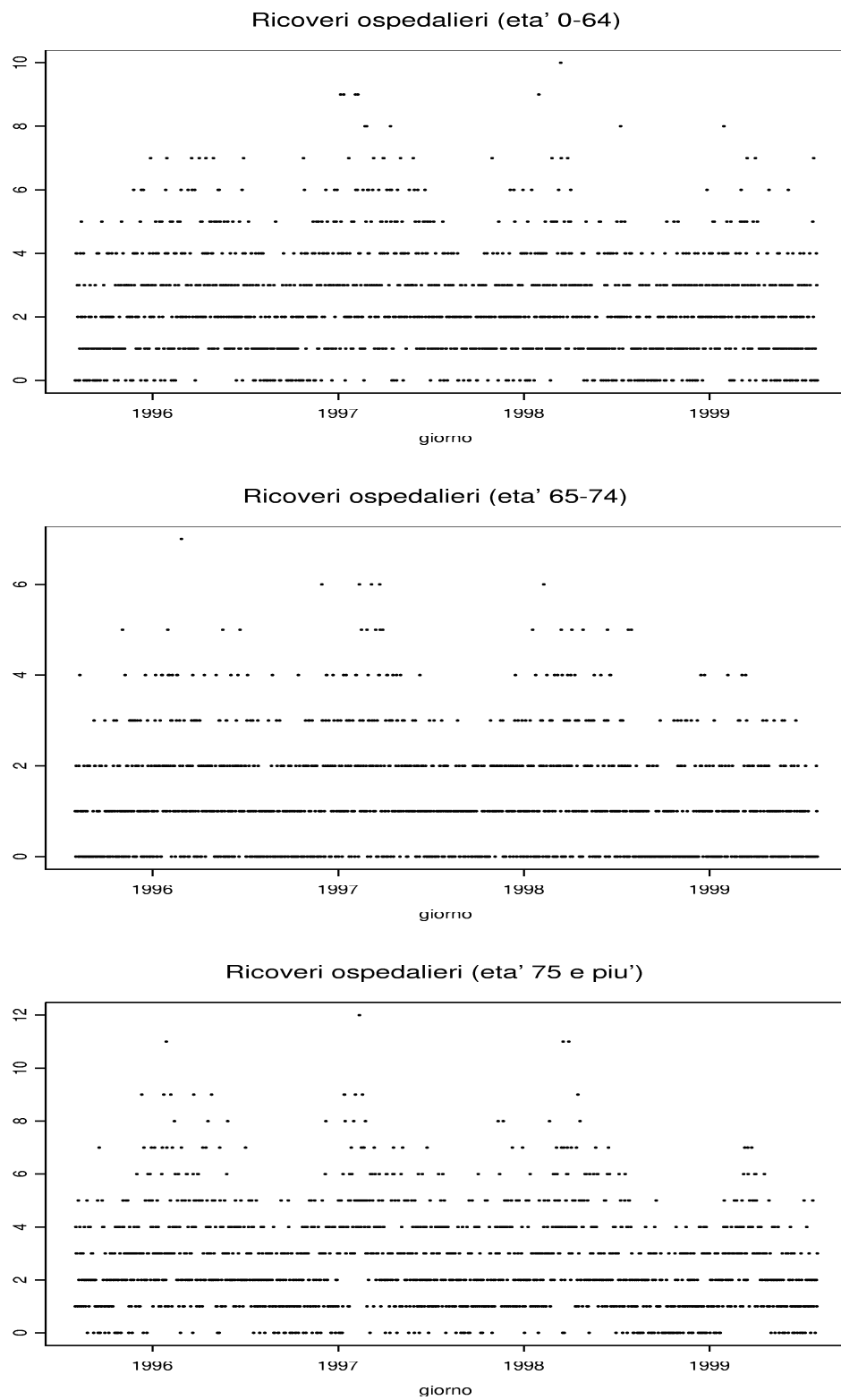


Figura 3.12: Serie storiche dei ricoveri giornalieri rispettivamente per le classe d'età 0-64, 65-74 e 75 e più.

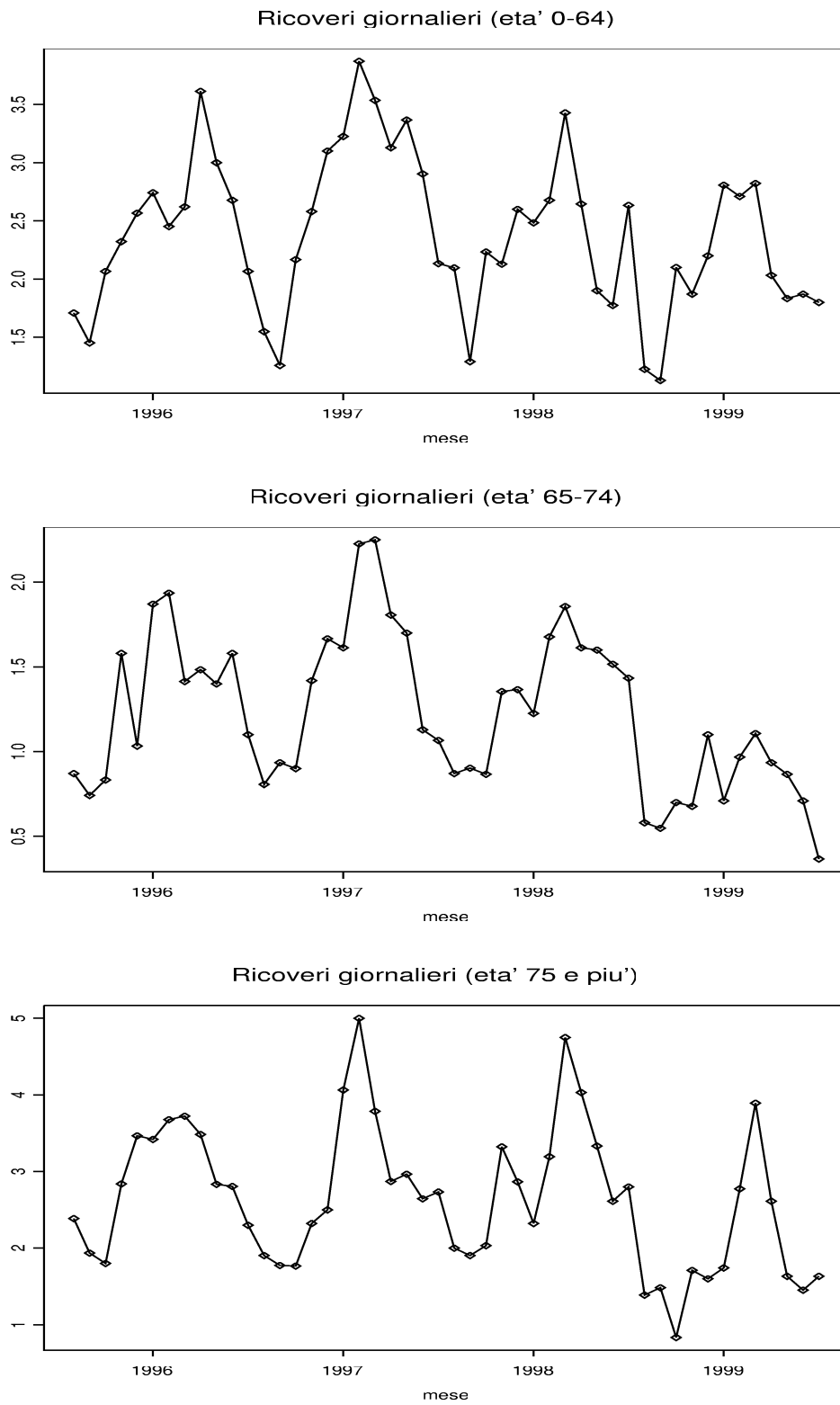


Figura 3.13: Serie storiche mensile dei ricoveri giornalieri rispettivamente per le classe d'età 0-64, 65-74 e 75 e più.

Capitolo 4

Analisi e risultati nel confronto tra le funzioni *gam* e *gam.exact*

In questo capitolo si sono poste a confronto le due funzioni *gam* e *gam.exact*. Il tentativo è quello di vedere se la funzione *gam.exact*, sviluppata come modifica e miglioramento della funzione *gam*, porta, come dichiarato dagli autori (Dominici e colleghi, 2002), ad una correzione della sottostima della matrice di varianze e covarianze dei coefficienti di interesse, che misurano l'intensità dell'effetto di ogni singolo inquinante sulla variabile risposta.

4.1 Stima del modello mediante la funzione *gam*

Dopo aver notato che i valori di default dei criteri di convergenza della funzione *gam* non erano sufficientemente stringenti da garantire la convergenza, Dominici e colleghi (2002) hanno suggerito criteri di convergenza più restrittivi: *maxit* = 1000, *bf.maxit* = 1000; *epsilon* = $10e - 15$, *bf.epsilon* = $10e - 15$. Questi criteri sono stati utilizzati nelle analisi che seguono.

4.1.1 Costruzione del *core model*

I dati a cui si fa riferimento nella seguente analisi sono quelli descritti nel Capitolo 3.

La variabile risposta del modello è rappresentata dal numero dei ricoveri giornalieri dovuti a disagi respiratori e tali eventi si considerano determinazioni di una variabile casuale di Poisson che conta il numero di arrivi in ospedale in un determinato giorno.

Per tener conto della diversa influenza che plausibilmente ha l'età sul numero di ricoveri ospedalieri, questo numero è stato disaggregato per le classi di età 0-64, 65-74, 75 e più. La prima classe in realtà cumula i ricoveri ospedalieri di due classi d'età, 0-14 e 15-64, in considerazione del basso numero di ricoveri riguardanti le persone più giovani.

Tra le variabili esplicative, oltre al singolo inquinante, se ne sono inserite altre che tenessero conto del confondimento temporale, meteorologico ovvero sanitario sul totale dei ricoveri giornalieri.

Un'attenzione particolare è stata data, sempre in fase di analisi, al possibile effetto ritardato che alcune variabili, in special modo temperatura e inquinamento, possono avere sulla salute. I ritardi presi in considerazione, 1, 2 e 3, si riferiscono rispettivamente a uno, due e tre giorni antecedenti quello preso come riferimento, mentre i lag 0-1 e 0-3 sono stati calcolati come media dei valori su più giorni. Per esempio, posta pari a X_{it} la concentrazione dell'inquinante i al giorno t , le concentrazioni ai lag 0-1 e 0-3 sono date rispettivamente da:

$$X_{i01} = \frac{X_{i0} + X_{i1}}{2}$$

e

(4.1)

$$X_{i03} = \frac{X_{i0} + X_{i1} + X_{i2} + X_{i3}}{4}$$

Non si ritiene opportuno andare ulteriormente indietro nel tempo in quanto l'effetto degli inquinanti e di altre variabili non dovrebbe più essere rilevante per lo studio dei ricoveri nel giorno preso in considerazione.

Secondo il protocollo utilizzato nello studio MISA¹, per modellare l'andamento temporale dei ricoveri si sono introdotte nel modello tre *spline cubiche*, uno per ogni classe d'età. I gradi di libertà per Verona sono per ogni classe d'età rispettivamente 5, 4 e 5 per anno. Questi rappresentano un compromesso tra adattamento e parsimonia e vanno inseriti come argomento delle funzioni di liscio, presenti nel modello iniziale come variabili indipendenti.

Il modello statistico, espresso tramite la sintassi di S-PLUS, appare dunque nella forma:

$$\begin{aligned}
 RIC \sim & HOL + WDAY + INFLU + AGE2 + AGE3 + \\
 & + s(DAYAGE1, "20") + \\
 & + s(DAYAGE2, "16") + \\
 & + s(DAYAGE3, "20") \quad (4.2)
 \end{aligned}$$

dove:

- RIC rappresenta la variabile risposta data dai ricoveri ospedalieri avvenuti per disagi respiratori;
- HOL è una variabile indicatrice del giorno di vacanza;
- WDAY è una variabile che tiene conto del giorno della settimana;
- INFLU è la variabile che tiene conto dell'effetto della presenza di epidemie di influenza sui ricoveri giornalieri.

Nel modello si è optato per adottare le funzioni di liscio che tengano conto dell'influenza dell'età; questo perché è plausibile ritenere che la

¹Metanalisi Italiana degli Studi sugli effetti dell'inquinamento Atmosferico

maggior parte delle variabili esplicative contenute nel modello agiscono in modo diverso su soggetti più o meno giovani o anziani.

L'espressione (4.2) rappresenta la formula base da cui ha inizio l'analisi.

Successivamente si verifica la significatività di alcuni termini di interazione tra l'età ed alcune altre variabili di confondimento quali il giorno della settimana, le epidemie di influenza, la temperatura corrente e lo scarto tra la temperatura corrente e la media delle temperature nei tre giorni che precedono il giorno preso come riferimento.

Verona

Ricoveri respiratori

```
interazioni termini nel core
WDAY:AGE    INFLU:AGE    TEMP24:AGE    Tscarto:AGE
"signif"   "non.signif"   "non.signif"   "non.signif"
```

Tabella 4.1: Significatività di alcuni termini nel *core*.

Dalla Tabella 4.1 si nota che solo la prima di queste interazioni risulta significativa ad un livello del 5% e questo significa che c'è un effetto non trascurabile tra l'età ed il giorno della settimana in cui avviene il ricovero. Le altre interazioni, pur non risultando significative dal punto di vista statistico, vengono ugualmente inserite nel modello poiché risultano significative dal punto di vista applicativo.

Nel modello viene anche inserita la variabile umidità come termine lineare, ma non l'interazione con l'età, non prevista dal protocollo. Come per la temperatura, anche la variabile umidità viene inserita a prescindere dalla sua significatività, poiché è una variabile, la cui presenza si deve tenere in considerazione.

Il modello fino ad ora sviluppato è il seguente:

$$\begin{aligned}
RIC \sim & HOL + WDAY + INFLU + AGE2 + AGE3 + \\
& + s(DAYAGE1, "20") + s(DAYAGE2, "16") + \\
& + s(DAYAGE3, "20") + WDAY \times AGE2 + WDAY \times AGE3 + \\
& + INFLU \times AGE2 + INFLU \times AGE3 + TEMP24 + Tscarto + TEMP24 \times AGE2 + \\
& TEMP24 \times AGE3 + Tscarto \times AGE2 + Tscarto \times AGE3 + HUM \quad (4.3)
\end{aligned}$$

dove:

- TEMP24 rappresenta la temperatura corrente espressa in gradi centigradi;
- Tscarto rappresenta la differenza tra la temperatura corrente e quella media dei tre giorni precedenti. Più precisamente, indicata con X_t la temperatura nel giorno t , il valore di $Tscarto$ in quel giorno è esprimibile come:

$$Tscarto_t = \frac{1}{3} (X_{t-1} + X_{t-2} + X_{t-3}) - X_t$$

Si noti che le variabili *INFLU*, *WDAY* e *TEMP24* non sono state introdotte singolarmente nel modello ed è stato considerato l'effetto della loro interazione con ciascuna classe d'età. Queste variabili sono state inserite anche se non significative in modo da depurare per quanto possibile i ricoveri giornalieri dall'effetto di eventuali confondenti.

I residui del modello fino a questo punto specificato, separatamente per le tre classi d'età considerate, mostrano assenza di autocorrelazione trovandosi quasi tutti all'interno delle bande di confidenza (Figura 4.1).

Questo è indice del fatto che il modello è ben specificato, nel senso che riesce a cogliere i principali andamenti temporali e i più rilevanti effetti meteorologici.

Partial Correlation matrix:

lag residui.AGE1		lag residui.AGE2		lag residui.AGE3	
1	0.0255	1	0.0102	1	0.0111
2	0.0067	2	-0.0103	2	0.0038
3	-0.0353	3	-0.0332	3	0.0153
4	0.0023	4	-0.0393	4	-0.0009
5	-0.0387	5	-0.0323	5	-0.0661
6	-0.0408	6	0.0352	6	0.0305
7	0.0268	7	-0.0369	7	-0.0105
8	-0.0648	8	-0.0095	8	0.0448
9	-0.0198	9	0.0132	9	0.0362
10	-0.0344	10	-0.0552	10	-0.0080
11	-0.0320	11	0.0186	11	-0.0156
12	-0.0529	12	-0.0202	12	-0.0348
13	0.0533	13	-0.0185	13	-0.0038
14	0.0032	14	0.0297	14	-0.0361
15	-0.0286	15	-0.0171	15	0.0463
16	-0.0090	16	-0.0221	16	-0.0074
17	-0.0188	17	-0.0088	17	0.0115
18	0.0035	18	-0.0212	18	-0.0736
19	-0.0390	19	0.0473	19	0.0094
20	0.0251	20	-0.0150	20	-0.0361
21	-0.0084	21	0.0252	21	-0.0309
22	-0.0067	22	-0.0016	22	-0.0109
23	-0.0353	23	-0.0727	23	0.0446
24	-0.0022	24	-0.0405	24	-0.0113
25	-0.0027	25	-0.0397	25	-0.0076
26	-0.0013	26	-0.0147	26	-0.0477
27	0.0346	27	-0.0172	27	-0.0147
28	0.0047	28	-0.0438	28	0.0175
29	0.0056	29	-0.0403	29	-0.0151
30	-0.0194	30	-0.0419	30	-0.0248
31	-0.0382	31	0.0047	31	-0.0339

Tabella 4.2: Residui del modello in assenza dell'influenza degli inquinanti.

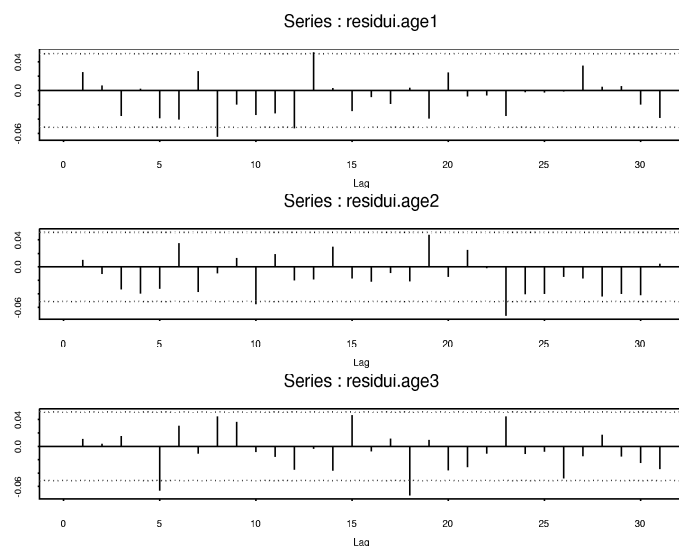


Figura 4.1: Funzioni di autocorrelazione parziale dei residui in assenza dell'influenza degli inquinanti, separatamente per le tre classi d'età.

4.1.2 Introduzione dell'inquinante

Il modello viene a questo punto ampliato attraverso l'introduzione degli inquinanti e valutando l'influenza che ognuno di essi ha sulla variabile risposta ai ritardi 1, 2, 3, 0-1, 0-3. Ogni tipo di inquinante viene studiato indipendentemente dagli altri, poiché è verosimile trovare correlazione tra gli inquinanti: questo rischio di multicollinearità sarebbe difficile da gestire.

In questa fase viene studiata l'interazione con l'età. Dai risultati, che non sono stati riportati, si è notato che per nessun tipo di inquinante esiste interazione con l'età ad un livello di significatività $\alpha = 0.05$. Questa conclusione farebbe supporre che l'inquinamento agisce sui soggetti sotto studio indipendentemente dall'età dei soggetti stessi.

4.1.3 Analisi dell'eventuale autocorrelazione (solo per lag 0-3)

A questo punto si sono calcolati i residui dei modelli che contengono i diversi tipi di inquinanti, separatamente per le tre classi di età. A titolo di esempio sono riportati i risultati ottenuti e i rispettivi grafici per le polveri sospese totali. La scelta di questo tipo di inquinante è dettata dal ruolo preminente che esso gioca negli studi di associazione tra inquinamento e salute. L'effetto dannoso delle polveri sulla ricoverabilità respiratoria è infatti ben documentato in letteratura.

TSP :

Partial Correlation matrix:

laq residui.AGE1		laq residui.AGE2		laq residui.AGE3	
1	0.0252	1	0.0112	1	0.0088
2	0.0053	2	-0.0108	2	0.0010
3	-0.0370	3	-0.0356	3	0.0130
4	-0.0002	4	-0.0408	4	-0.0020
5	-0.0419	5	-0.0356	5	-0.0674
6	-0.0447	6	0.0316	6	0.0294
7	0.0235	7	-0.0385	7	-0.0111
8	-0.0666	8	-0.0117	8	0.0433
9	-0.0206	9	0.0107	9	0.0356
10	-0.0355	10	-0.0562	10	-0.0077
11	-0.0323	11	0.0175	11	-0.0140
12	-0.0537	12	-0.0194	12	-0.0320
13	0.0511	13	-0.0175	13	-0.0031
14	0.0007	14	0.0299	14	-0.0364
15	-0.0319	15	-0.0168	15	0.0448
16	-0.0123	16	-0.0209	16	-0.0104
17	-0.0210	17	-0.0082	17	0.0093
18	0.0011	18	-0.0209	18	-0.0745
19	-0.0411	19	0.0475	19	0.0086
20	0.0236	20	-0.0153	20	-0.0356
21	-0.0119	21	0.0245	21	-0.0320
22	-0.0103	22	-0.0025	22	-0.0134
23	-0.0376	23	-0.0727	23	0.0420
24	-0.0031	24	-0.0397	24	-0.0124
25	-0.0037	25	-0.0396	25	-0.0088
26	-0.0025	26	-0.0161	26	-0.0491
27	0.0346	27	-0.0172	27	-0.0168
28	0.0050	28	-0.0448	28	0.0158
29	0.0051	29	-0.0418	29	-0.0168
30	-0.0204	30	-0.0433	30	-0.0262
31	-0.0393	31	0.0031	31	-0.0360

Tabella 4.3: Residui del modello in presenza di *TSP*.

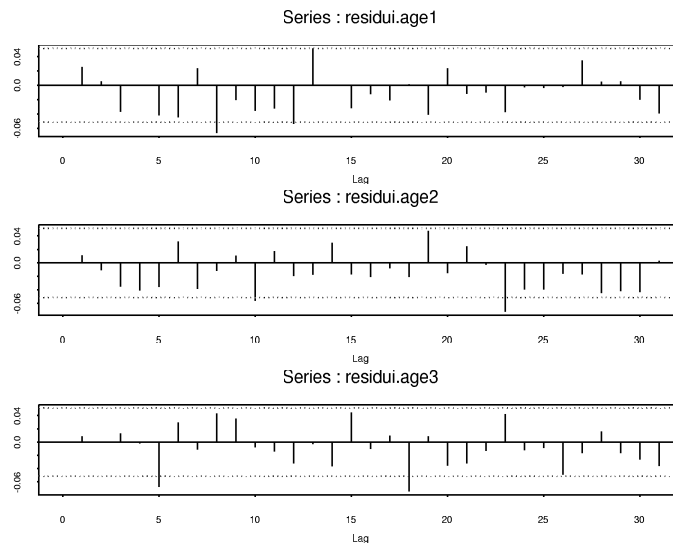


Figura 4.2: Funzioni di autocorrelazione parziale dei residui in presenza di *TSP*.

Dai grafici riportati in Figura 4.2 e da quelli relativi agli altri inquinanti, non riportati per brevità, risulta evidente l'assenza di autocorrelazione nei residui per ciascun tipo di inquinante: quasi tutti i residui si trovano entro le bande di confidenza. Questo sta a significare la buona specificazione del modello, nel senso precedentemente inteso.

4.1.4 Risultati

In questo paragrafo sono riportati e commentati i risultati della città di Verona per ogni tipo di inquinante preso in considerazione.

L'associazione dei dati giornalieri di ricoverabilità ospedaliera con la concentrazione atmosferica degli inquinanti è stata misurata riferendosi alla concentrazione dello stesso giorno (lag 0), in ciascuno dei tre giorni precedenti (lag 1, 2, 3) e con quella media di più giorni (lag 0-1, 0-3). La scelta dei lag 0-1 e 0-3 per i ricoveri ha lo scopo di differenziare, come detto, effetti cumulati immediati rispetto a quelli ritardati.

Si sono ottenute così stime globali sintetiche per il periodo 1995 - 1999. I risultati sono espressi in termini di variazione percentuale nel numero giornaliero di ricoveri associato ad un incremento di $10\mu g/m^3$ nella concentrazione di SO_2 , NO_2 , TSP e O_3 . Per il CO l'incremento è invece di $1mg/m^3$.

In Tabella 4.4 sono riportati le variazioni percentuali e i limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% per ogni inquinante e per ogni ritardo considerato. La scelta di riportare le variazioni percentuali anziché i coefficienti stimati $\hat{\beta}$ è dettata dal fatto che le prime sono numeri più leggibili poiché rappresentano il numero di ricoveri addizionali dovuto all'aumento di un determinato inquinante.

In Appendice A si dà una semplice chiarificazione di cosa si intenda per variazione percentuale.

Il segno positivo del limite inferiore dell'intervallo di confidenza indica che si è in presenza di un effetto statisticamente significativo dell'inquinante al 5%. Nel caso esaminato si può affermare, in generale, che l'effetto dell'inquinante sui ricoveri ospedalieri è significativo ai lag 2, 3 e 0-3, anche se l'effetto maggiore è perlopiù associato al livello di inquinante al ritardo 0-3. Ai ritardi 0, 1 e 0-1 non risulta significativo l'effetto di nessun inquinante.

Singolare è il comportamento dell' O_3 , il quale non risulta avere influenza sui ricoveri a nessun ritardo. Questo può, tuttavia, essere dovuto al fatto che l'ozono è presente nell'atmosfera prevalentemente nel periodo estivo, quando si verifica una deplezione della popolazione residente e quindi la numerosità dei ricoveri è in calo.

In Appendice B sono riportate le stime per tutti i lag dei coefficienti di interesse e dei rispettivi errori standard per tutti gli inquinanti e per le interazioni tra ogni inquinante e l'età, relativamente al periodo di studio (01 luglio 1995 - 30 giugno 1999).

inquinante	LAG 0			LAG 1			LAG 2		
	%	limite inf.	limite sup.	%	limite inf.	limite sup.	%	limite inf.	limite sup.
<i>SO</i> ₂	2.5	-3.12	8.4	0.60	-4.74	6.2	5.0	-0.42	10.7
<i>NO</i> ₂	1.1	-0.21	2.5	0.99	-0.29	2.3	2.5	1.23	3.7
<i>CO</i>	1.1	-0.84	3	0.52	-1.3	2.4	3.4	1.6	5.3
<i>TSP</i>	0.83	-0.19	1.9	0.82	-0.2	1.8	1.9	0.89	2.9
<i>O</i> ₃	-0.66	-1.4	0.12	-0.32	-1.1	0.42	0.28	-0.45	1

inquinante	LAG 3			LAG 0-1			LAG 0-3		
	%	limite inf.	limite sup.	%	limite inf.	limite sup.	%	limite inf.	limite sup.
<i>SO</i> ₂	7.9	2.5	13.6	2.0	-4.07	8.4	7.4	0.53	14.7
<i>NO</i> ₂	2.8	1.6	4.1	1.3	-0.13	2.7	3.0	1.46	4.5
<i>CO</i>	2.7	0.83	4.5	1	-1	3.1	3.4	1.2	5.6
<i>TSP</i>	2.4	1.4	3.4	0.96	-0.11	2	2.2	1.1	3.4
<i>O</i> ₃	0.51	-0.2	1.2	-0.65	-1.5	0.19	-0.03	-0.92	0.87

Tabella 4.4: Variazioni percentuali e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95%, stimati mediante la funzione *gam* e relativi all'effetto di ogni inquinante sui ricoveri ospedalieri.

Per concludere si noti che tutti i risultati che riguardano lo studio del modello mediante la funzione *gam* non cambiano usando un livello di significatività del 7%. Ponendo invece $\alpha = 0.10$ diventano significativi i coefficienti che si riferiscono all'interazione tra l'età e *SO*₂ ai ritardi 1 e 0-3; questo, tuttavia, non è di grande rilevanza nell'analisi condotta.

4.2 Confronto tra le funzioni *gam* e *gam.exact*

A questo punto l'analisi precedentemente svolta viene rieseguita attraverso l'uso della funzione *gam.exact* al posto della funzione *gam*, per i motivi ampiamente argomentati nel capitolo 2.

Tutti i passi precedentemente descritti sono stati ripetuti utilizzando la nuova funzione. Ad ogni passo dell'algoritmo è stato effettuato un confronto tra i risultati ottenuti mediante l'utilizzo della funzione *gam* e quelli ottenuti mediante l'utilizzo di *gam.exact* e si è notato che le conclusioni che si possono trarre sono in entrambi i casi le stesse.

In Tabella 4.5 sono riportati la variazione percentuale e i limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% per ogni inquinante e per ogni ritardo considerato.

inquinante	LAG 0			LAG 1			LAG 2		
	%	limite inf.	limite sup.	%	limite inf.	limite sup.	%	limite inf.	limite sup.
<i>SO</i> ₂	5.2	-0.53	11	3.3	-2.15	9.1	7.8	2.3	13.6
<i>NO</i> ₂	1.7	0.33	3	1.6	0.36	2.9	3.1	1.9	4.4
<i>CO</i>	2.3	0.38	4.2	1.9	0.057	3.8	4.6	2.8	6.5
<i>TSP</i>	1.33	0.26	2.3	1.3	0.31	2.4	2.4	1.4	3.4
<i>O</i> ₃	-1.1	-1.8	-0.27	-0.7	-1.4	0.04	-0.16	-0.89	0.57

inquinante	LAG 3			LAG 0-1			LAG 0-3		
	%	limite inf.	limite sup.	%	limite inf.	limite sup.	%	limite inf.	limite sup.
<i>SO</i> ₂	10.7	5.1	16.5	5.5	-0.73	12.1	12.1	5.0	19.6
<i>NO</i> ₂	3.4	2.2	4.7	2.0	0.59	3.4	3.9	2.4	5.4
<i>CO</i>	4	2.2	5.9	2.7	0.61	4.8	5.4	3.1	7.7
<i>TSP</i>	2.8	1.8	3.8	1.1	0.45	2.6	2.9	1.7	4
<i>O</i> ₃	0.033	-0.68	0.75	-1.1	-2	-0.3	-0.78	-1.7	0.12

Tabella 4.5: Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95%, stimati mediante la funzione *gam.exact* e relativi all'effetto di ogni inquinante sui ricoveri ospedalieri. In grassetto sono evidenziate le differenze significative rispetto ai risultati ottenuti con la funzione *gam*: in questo caso si nota un effetto significativo dell'inquinante sulla salute.

In Appendice C sono contenuti i risultati completi da porre a confronto con quelli contenuti nell'Appendice B.

A questo punto si ritiene opportuno puntualizzare che si sono riscontrati dei problemi nell'applicazione della funzione *gam.exact* in quanto il modello non riusciva ad essere stimato e l'output di S-PLUS presentava il seguente errore:

```
modbase<-gam.exact(eval(parse(text=formula1)),
  data=dati2,maxit=1000,bf.maxit=1000,epsilon=10e-15,
  bf.epsilon=10e-15,family=poisson,na.action=na.omit)

Error in dim(data) <- dim: Length of data (99)
  doesn't match product of dimensions (81)
```

Dopo vari tentativi per trovare una soluzione al problema, si è riscontrato un errore nella funzione *gam.exact*: la matrice di varianze e covarianze dei coefficienti $\hat{\beta}$ di interesse non poteva essere calcolata dal momento che non era stata costruita con un numero intero di righe e colonne. Così, nella funzione *gam.exact*, al comando:

```
varbeta <- matrix(varbeta,nrow=sqrt(length(varbeta)),
  ncol=sqrt(length(varbeta)))
```

è stato sostituito il comando esatto:

```
varbeta <- matrix(varbeta,nrow=round(sqrt(length(varbeta))),
  ncol=round(sqrt(length(varbeta))))
```

Come argomentato per la funzione *gam*, anche per la funzione *gam.exact* i risultati si mantengono pressoché analoghi adottando un livello di significatività del 7 o del 10%.

Ottenuti i risultati, si pongono a confronto le stime dei coefficienti β calcolate con entrambe le funzioni, ai vari ritardi. Il primo grafico di Figura 4.3 rappresenta le stime dei coefficienti β relativamente alle polveri sospese totali, inquinante preso in considerazione anche nelle precedenti analisi.

I punti colorati di nero si riferiscono alle stime dei coefficienti relativi all'inquinante per i sei ritardi considerati, mentre gli altri si riferiscono all'interazione tra l'inquinante e l'età.

La linea continua rappresenta la bisettrice del I quadrante.

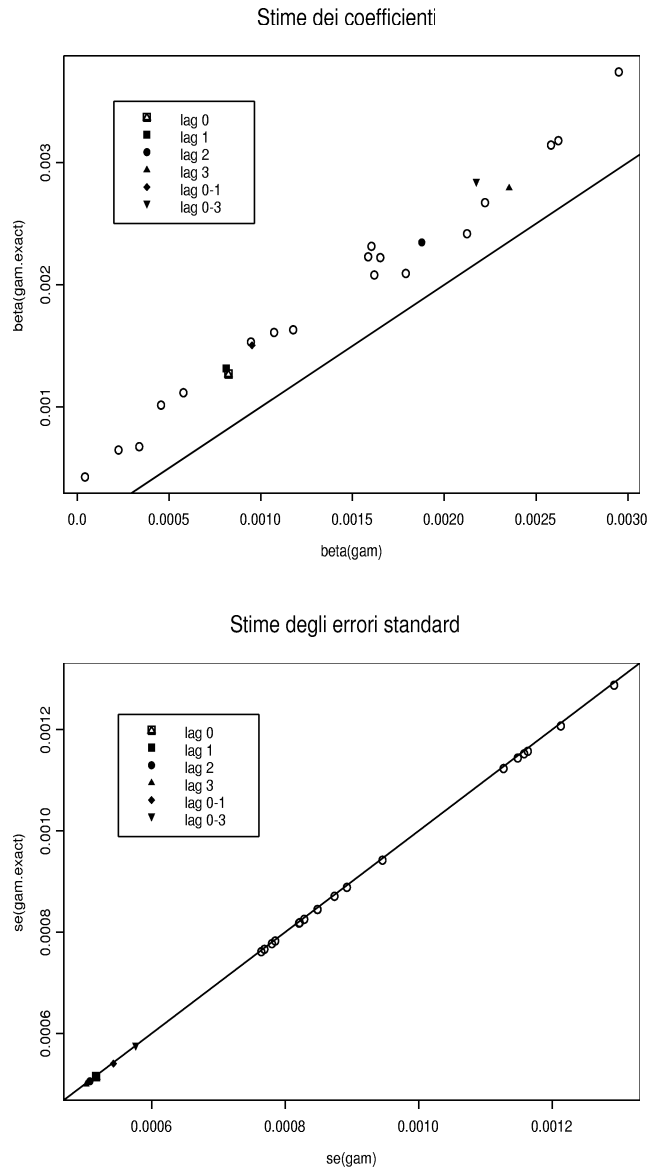


Figura 4.3: Confronto tra le stime dei coefficienti β e degli errori standard (*TSP*).

Si può notare che le stime dei coefficienti calcolati con la funzione *gam.exact* assumono dei valori superiori rispetto a quelle calcolate con la funzione *gam* e questa caratteristica è propria di tutti i ritardi considerati.

Nel secondo grafico di Figura 4.3 si pongono a confronto gli errori standard ottenuti dall'implementazione delle due diverse funzioni, sempre relativi alle polveri sospese totali. I valori degli errori standard risultano pressoché analoghi se calcolati con entrambe le funzioni, dal momento che sono disposti lungo la bisettrice e questo porta alla conclusione che la funzione *gam* non calcola in modo dissimile gli errori standard, rispetto alla *gam.exact*.

Si noti, infine, che gli errori standard dei coefficienti relativi all'inquinante assumono valori decisamente inferiori rispetto a quelli relativi all'interazione tra l'inquinante e l'età. Tuttavia, questo è dovuto probabilmente al fatto che le stime che riguardano l'interazione tra inquinante ed età sono più instabili, poichè calcolate con una più bassa numerosità.

Analoghe conclusioni, sia per quanto riguarda le stime dei coefficienti β sia per quanto riguarda le stime degli errori standard, si possono trarre per gli altri inquinanti ad eccezione dell'ozono, i cui grafici sono stati riportati in Appendice D.

Nel paragrafo che segue si affronta il problema della scelta dei gradi di libertà delle funzioni di lisciamiento del modello.

4.3 La scelta dei gradi di libertà

Uno dei principali interrogativi da porsi nello studio condotto in questo progetto riguarda la scelta del numero di gradi di libertà per i fattori di confondimento, che permettono alle stime degli effetti dell'inquinamento di avere proprietà statistiche ottimali: i gradi di libertà devono essere in numero abbastanza elevato da rimuovere il confondimento, ma non troppo per non eliminare l'effetto dell'inquinamento. Tale scelta rimane ad oggi piuttosto critica e rappresenta un problema ancora irrisolto.

Per vedere se e in che modo cambiano i risultati precedentemente ottenuti scegliendo diversi gradi di aggiustamento per i fattori confondenti, si specifica il modello moltiplicando tutti i gradi di libertà per un parametro α :

$$Y_t \sim \text{Poisson}(\lambda_t)$$

$$V(Y_t) = \lambda_t \quad (4.4)$$

$$\log(\lambda_t) = \beta_0(\alpha) + \sum_{j=1}^J \beta_j(\alpha) X_{jt} + \sum_{l=J+1}^L f_l(X_{lt}, df_l \times \alpha)$$

dove:

- X_{lt} rappresentano i fattori confondenti varianti nel tempo come la stagionalità ed altri trend di lungo periodo;
- df_l rappresentano i gradi di libertà prespecificati;
- $\beta_j(\alpha)$ sono i coefficienti che misurano l'intensità dell'effetto dell'inquinamento sulla salute, il cui valore può cambiare in funzione di un parametro α .

A questo punto si sono scelti venti valori per il parametro α , compresi tra $1/3$ e 2 e si sono ricondotte le analisi svolte nei paragrafi precedenti, con l'utilizzo di entrambe le funzioni *gam* e *gam.exact*.

La scelta di questi valori di α sta a significare che, per la città di Verona, si intendono cogliere andamenti di un diverso numero di mesi. Ad esempio, essendo 5 i gradi di libertà fissati per Verona relativamente alla prima classe d'età considerata, porre $\alpha = 1/3$ significa cogliere andamenti all'incirca semestrali, mentre ponendo $\alpha = 2$ si colgono lisciami mensili. I valori di α intermedi colgono ovviamente andamenti compresi tra sei mesi ed un mese.

A titolo di esempio, nei grafici in Figura 4.4 e nella Tabella 4.6 sono riportati le variazioni percentuali e i limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% per ogni scelta di α , relativamente alle polveri sospese totali e al lag 0-3.

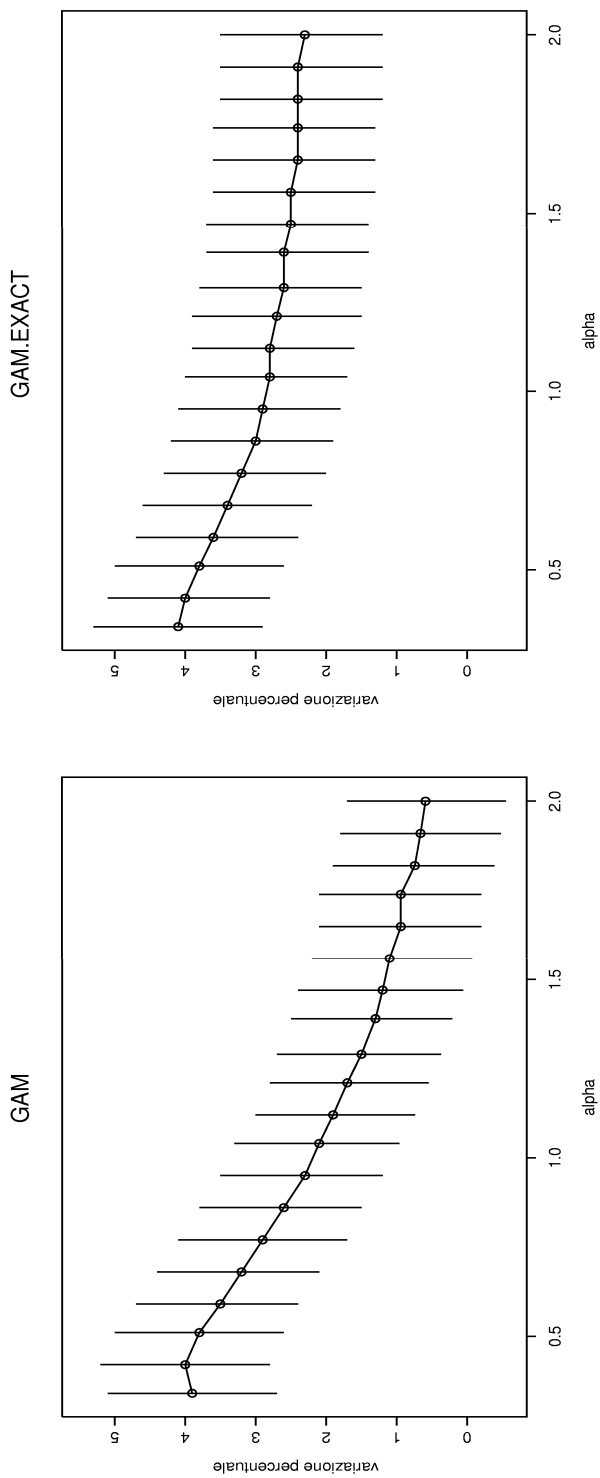


Figura 4.4: Variazioni percentuali ed intervalli di confidenza approssimati al 95% relativi all'effetto delle polveri sospese totali (*TSP*) sulla ricoverabilità, stimate utilizzando rispettivamente la funzione *gam* e *gam.exact*.

α	GAM			GAM.EXACT		
	variazione percentuale	limite inferiore	limite superiore	variazione percentuale	limite inferiore	limite superiore
0.34	3.9	2.7	5.1	4.1	2.9	5.3
0.42	4	2.8	5.2	4	2.8	5.1
0.51	3.8	2.6	5	3.8	2.6	5
0.59	3.5	2.4	4.7	3.6	2.4	4.7
0.68	3.2	2.1	4.4	3.4	2.2	4.6
0.77	2.9	1.7	4.1	3.2	2	4.3
0.86	2.6	1.5	3.8	3	1.9	4.2
0.95	2.3	1.2	3.5	2.9	1.8	4.1
1.04	2.1	0.96	3.3	2.8	1.7	4
1.12	1.9	0.74	3	2.8	1.6	3.9
1.21	1.7	0.55	2.8	2.7	1.5	3.9
1.29	1.5	0.37	2.7	2.6	1.5	3.8
1.39	1.3	0.21	2.5	2.6	1.4	3.7
1.47	1.2	0.058	2.4	2.5	1.4	3.7
1.56	1.1	-0.075	2.2	2.5	1.3	3.6
1.65	0.94	-0.2	2.1	2.4	1.3	3.6
1.74	0.94	-0.2	2.1	2.4	1.3	3.6
1.82	0.74	-0.39	1.9	2.4	1.2	3.5
1.91	0.66	-0.48	1.8	2.4	1.2	3.5
2	0.59	-0.55	1.7	2.3	1.2	3.5

Tabella 4.6: Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto delle polveri sospese totali (*TSP*) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu g/m^3$. In grassetto sono evidenziate le differenze significative tra i risultati ottenuti applicando le due funzioni *gam* e *gam.exact*.

All'aumentare dei gradi di libertà la variazione percentuale diminuisce, e questo accade sia nel caso in cui si utilizzi la funzione *gam* sia quello in cui si faccia uso della funzione *gam.exact*. Questo è dovuto al fatto che, aumentando α , il modello tende a seguire maggiormente l'andamento dei dati e quindi la variabilità residua ad essere più piccola. Il modello risulta pertanto sempre meno rigido ma più sensibile al campione a cui viene applicato.

Oltre a questa osservazione, dall'analisi dei grafici nella Figura 4.4 ne emerge un'altra decisamente più interessante e sorprendente: mentre utilizzando la funzione *gam* la variazione percentuale diminuisce rapidamente all'aumentare dei gradi di libertà, l'uso della funzione *gam.exact* comporta una diminuzione decisamente più lenta. Altrettanto interessante è notare che, a partire da un valore di α pari a 1.56, l'effetto dell'inquinante risulta essere non significativo se si applica la funzione *gam*: l'intervallo di confidenza contiene infatti il valore zero. Utilizzando la funzione *gam.exact*, al contrario, l'effetto delle polveri sospese totali si mantiene significativo per ogni scelta del parametro α .

Una possibile spiegazione del risultato può derivare dalla diversa stima dei coefficienti β ottenuta mediante le due funzioni: tenendo conto delle conclusioni che emergono dalla Figura 4.3 e della definizione di variazione percentuale riportata in Appendice A, può risultare che, a parità di gradi di libertà, i valori delle variazioni percentuali siano maggiori se calcolati con la funzione *gam.exact*. Si noti, tuttavia, che questa caratteristica emerge solamente all'aumentare dei gradi di libertà; per valori bassi del parametro α , infatti, le stime delle variazioni percentuali sono pressoché analoghe se si utilizzano le due funzioni.

In base alle analisi effettuate si può pertanto sospettare che, le funzioni *gam* e *gam.exact* conducano ad interpretazioni differenti riguardo l'effetto di un inquinante, in questo caso delle polveri sospese totali, solo all'aumentare dei gradi di libertà. Questa conclusione non può certamente essere considerata definitiva per almeno due ovvie ragioni: innanzitutto sarebbe opportuno condurre analisi simili in diverse città e per diversi inquinanti ed inoltre la

città di Verona non conta un numero sufficientemente elevato di abitanti, e quindi di ricoveri giornalieri, da poter rendere abbastanza generalizzabili i risultati ottenuti.

In appendice E vengono riportati i grafici e le tabelle delle variazioni percentuali e degli intervalli di confidenza approssimati al 95% costruiti con entrambe le funzioni e relativi agli altri inquinanti considerati: anidride solforosa, biossido di azoto, monossido di carbonio ed ozono.

4.4 Conclusioni

Riassumendo, si può affermare che nel caso esaminato della città di Verona, e utilizzando la funzione *gam*, l'effetto dell'inquinamento sui ricoveri giornalieri è perlopiù significativo ai ritardi 2, 3 e 0-3. Introducendo nel modello il singolo inquinante come interazione col termine d'età, si è notato che per nessun tipo di inquinante esiste interazione con l'età ad un livello di significatività pari al 5%.

I risultati non cambiano in modo evidente se si utilizza la funzione *gam.exact*. Confrontando però le stime dei coefficienti delle variabili di inquinamento, calcolate con entrambe le funzioni ai vari ritardi, si nota che tali stime calcolate con la funzione *gam.exact* assumono dei valori superiori rispetto a quelle calcolate con la funzione *gam*; rimangono invece perlopiù analoghe le stime degli errori standard. Questo vale per tutti gli inquinanti escluso l'ozono.

Da ultimo, dal confronto delle variazioni percentuali e degli intervalli di confidenza approssimati al 95% relativi all'effetto di ogni inquinante sui ricoveri giornalieri, e calcolati per diversi valori dei gradi libertà, si possono evidenziare significative differenze tra le due funzioni: si nota una diminuzione più rapida della variazione percentuale, all'aumentare dei gradi di libertà, se si utilizza la funzione *gam*, tale da rendere non significativo l'effetto di ogni inquinante per scelte di α più vicine al valore 2. Il contrario accade per l'ozono.

In conclusione, si ritiene opportuno sottolineare, come aspetto limitativo dei risultati di questa analisi, la bassa numerosità campionaria dovuta sia alla dimensione demografica sia al numero di posti letto ospedaliero presenti nella città di Verona.

Risultati diversi potrebbero ottenersi stimando modelli alternativi oppure considerando altre variabili confondenti o ancora altri tipi di inquinanti.

Appendice A

Calcolo della variazione percentuale

Sia X_t il valore dell'inquinante al tempo t e X_{t+10} il valore dell'inquinante al tempo $t + 10$.

Il numero medio di eventi al tempo t è il seguente:

$$E(Y_t) = \exp \left\{ \beta_0 + \beta X_t + \sum_j f_j(X_t, df_j) \right\} \quad (\text{A.1})$$

mentre al tempo $t + 10$ è:

$$E(Y_{t+10}) = \exp \left\{ \beta_0 + \beta (X_t + 10) + \sum_j f_j(X_t, df_j) \right\} \quad (\text{A.2})$$

La variazione percentuale è definita come:

$$\begin{aligned}
& \left(\frac{E(Y_{t+1}) - E(Y_t)}{E(Y_t)} \right) \times 100 \\
&= \left(\frac{\exp \left\{ \beta_0 + \beta X_{t+10} + \sum_j f_j(X_t, df_j) \right\} - \exp \left\{ \beta_0 + \beta X_t + \sum_j f_j(X_t, df_j) \right\}}{\exp \left\{ \beta_0 + \beta X_t + \sum_j f_j(X_t, df_j) \right\}} \right) \\
&\quad \times 100 \\
&= \left(\frac{\exp \{ \beta X_t \} \exp \{ 10\beta \} - \exp \{ \beta X_t \}}{\exp \{ \beta X_t \}} \right) \times 100 \\
&= (\exp \{ 10\beta \} - 1) \times 100 \tag{A.3}
\end{aligned}$$

L'intervallo di confidenza per la variazione percentuale è dato da:

$$\hat{\beta} \pm 1.96 s.e.(\hat{\beta}) \tag{A.4}$$

dove $s.e.(\hat{\beta})$ indica la stima dell'errore standard di $\hat{\beta}$.

Appendice B

Stime dei coefficienti e degli errori standard ottenute applicando la funzione *gam*

Per ognuno dei lag considerati si riportano di seguito le stime dei coefficienti di interesse e dei rispettivi errori standard relativi agli inquinanti SO_2 , NO_2 , CO , TSP e O_3 .

Si noti che la notazione usata dal programma statistico S-PLUS $so_2 : age_i$ ($i = 1, 2, 3$) indica l'interazione tra l'inquinante e l'età i e che le stime per ogni inquinante non sono al netto degli altri inquinanti.

row.names	coeff.lag0	se.lag0	coeff.lag1	se.lag1	Citta'
so2	0.0024662171	0.0028734052	0.00060053678	0.0027864603	verona
so2:age1	0.0059553164	0.0045544045	0.00062724339	0.0045311886	verona
so2:age2	-0.0004507626	0.0061684463	-0.01134139787	0.0062929812	verona
so2:age3	0.0009161615	0.0042485079	0.00615647071	0.0041326275	verona
no2	0.0011061803	0.0006735931	0.00098535204	0.0006491912	verona
no2:age1	0.0009878498	0.0010796974	-0.00004608989	0.0010605014	verona
no2:age2	0.0014988470	0.0014758116	0.00003568724	0.0014603984	verona
no2:age3	0.0010197862	0.0010115769	0.00234811123	0.0009795149	verona
co	0.0105744566	0.0096919797	0.00513680140	0.0095440481	verona
co:age1	0.0058761862	0.0157887457	-0.01794586730	0.0157136449	verona
co:age2	0.0439970584	0.0212684785	0.01201505118	0.0214715104	verona
co:age3	-0.0022490613	0.0149189483	0.02170887844	0.0144146021	verona
tsp	0.0008240137	0.0005166149	0.00081249994	0.0005160045	verona
tsp:age1	0.0003382739	0.0008484372	0.00004277727	0.0008486338	verona
tsp:age2	0.0011769500	0.0011485751	0.00045696212	0.0011636415	verona
tsp:age3	0.0010726124	0.0007850348	0.00165178478	0.0007800359	verona
o3	-0.0006667695	0.0004022592	-0.00032433840	0.0003801491	verona
o3:age1	-0.0001706929	0.0006162123	0.00003325591	0.0005965753	verona
o3:age2	-0.0020115482	0.0008425830	-0.00095832971	0.0008224972	verona
o3:age3	-0.0004466483	0.0005894713	-0.00033584125	0.0005693569	verona

Tabella B.1: Stime dei coefficienti e degli errori standard ai ritardi 0 e 1.

row.names	coeff.lag2	se.lag2	coeff.lag3	se.lag3	Citta'
so2	0.0048834093	0.0027053662	0.00761069904	0.0026444148	verona
so2:age1	0.0065783784	0.0044179715	0.00934041413	0.0043403465	verona
so2:age2	-0.0033618101	0.0061656048	-0.00740302811	0.0061261320	verona
so2:age3	0.0073049739	0.0040501702	0.01310519379	0.0039456010	verona
no2	0.0024462471	0.0006242245	0.00279759410	0.0006136810	verona
no2:age1	0.0020634370	0.0010191845	0.00367824959	0.0009952162	verona
no2:age2	0.0006236374	0.0014163369	0.00076083808	0.0013956004	verona
no2:age3	0.0036663545	0.0009434427	0.00301816080	0.0009353441	verona
co	0.0333140634	0.0091456203	0.02620765918	0.0091689771	verona
co:age1	0.0109373509	0.0150852733	0.02485390440	0.0149502712	verona
co:age2	0.0351939750	0.0205751049	0.01458415321	0.0206813361	verona
co:age3	0.0516836381	0.0138019351	0.03322636304	0.0139633989	verona
tsp	0.0018773184	0.0005073729	0.00235359577	0.0005016760	verona
tsp:age1	0.0017903872	0.0008286551	0.00212367516	0.0008208716	verona
tsp:age2	0.0005788471	0.0011581834	0.00222169418	0.0011272377	verona
tsp:age3	0.0025810926	0.0007688868	0.00262137313	0.0007642367	verona
o3	0.0002748129	0.0003713022	0.00051362021	0.0003627904	verona
o3:age1	0.0002143570	0.0005919268	0.00065321195	0.0005814760	verona
o3:age2	-0.0012754047	0.0008217855	0.00005190979	0.0008045291	verona
o3:age3	0.0010931115	0.0005638715	0.00061658957	0.0005549303	verona

Tabella B.2: Stime dei coefficienti e degli errori standard ai ritardi 2 e 3.

row.names	coeff.lag01	se.lag01	coeff.lag03	se.lag03	Citta'
so2	0.00197115641	0.0031246458	0.00711302011	0.0033575438	verona
so2:age1	0.00427005265	0.0049970844	0.01025240611	0.0054236511	verona
so2:age2	-0.00738256846	0.0068464290	-0.00930580814	0.0075313152	verona
so2:age3	0.00457416994	0.0046125029	0.01252268469	0.0049817003	verona
no2	0.00125796951	0.0007081663	0.00290823249	0.0007451290	verona
no2:age1	0.00057159831	0.0011470237	0.00270481464	0.0012107280	verona
no2:age2	0.00091557067	0.0015706177	0.00112471169	0.0016737561	verona
no2:age3	0.00202650430	0.0010643596	0.00398558754	0.0011228611	verona
co	0.01012802389	0.0104516289	0.03318271929	0.0110885480	verona
co:age1	-0.00778958787	0.0171248237	0.01113703013	0.0181614873	verona
co:age2	0.03562986787	0.0231943051	0.04397760240	0.0248272984	verona
co:age3	0.01282982901	0.0159272122	0.04681651179	0.0168453320	verona
tsp	0.00095243456	0.0005425876	0.00217453283	0.0005760236	verona
tsp:age1	0.00022523018	0.0008926093	0.00161854166	0.0009459543	verona
tsp:age2	0.00094643248	0.0012131246	0.00160347573	0.0012930323	verona
tsp:age3	0.00158510837	0.0008217877	0.00295051170	0.0008738972	verona
o3	-0.00064991954	0.0004280567	-0.00002884691	0.0004561711	verona
o3:age1	-0.00009512878	0.0006596730	0.00040873652	0.0007135004	verona
o3:age2	-0.00190431601	0.0009039440	-0.00179905091	0.0009809266	verona
o3:age3	-0.00051857916	0.0006301161	0.00050175938	0.0006806510	verona

Tabella B.3: Stime dei coefficienti e degli errori standard ai ritardi 0-1 e 0-3.

Appendice C

Stime dei coefficienti e degli errori standard ottenute applicando la funzione *gam.exact*

Si riportano in tabella le stime dei coefficienti di interesse β e i rispettivi errori standard per tutti i ritardi considerati e per tutti gli inquinanti presi in esame, come fatto nella precedente analisi relativa all'utilizzo della funzione *gam*.

78 Stime dei coefficienti e degli errori standard ottenute applicando la funzione *gam.exact*

row.names	coeff.lag0	se.lag0	coeff.lag1	se.lag1	Citta'
so2	0.0050629857	0.0028559254	0.0032569819	0.0027693961	verona
so2:age1	0.0079119673	0.0045303139	0.0026903197	0.0045063391	verona
so2:age2	0.0030312525	0.0061231822	-0.0072672681	0.0062380075	verona
so2:age3	0.0036287369	0.0042218966	0.0086108505	0.0041110902	verona
no2	0.0016486390	0.0006734029	0.0016350475	0.0006485307	verona
no2:age1	0.0013943796	0.0010783217	0.0004855244	0.0010585169	verona
no2:age2	0.0020755094	0.0014770082	0.0007630053	0.0014601794	verona
no2:age3	0.0016678141	0.0010121209	0.0030481565	0.0009791648	verona
co	0.0226603273	0.0096273816	0.0191250279	0.0094689136	verona
co:age1	0.0155960592	0.0156843534	-0.0041389301	0.0155709987	verona
co:age2	0.0559518465	0.0211564603	0.0274268125	0.0213121298	verona
co:age3	0.0124958463	0.0148084432	0.0351890028	0.0143156407	verona
tsp	0.0012702677	0.0005146466	0.0013160398	0.0005137007	verona
tsp:age1	0.0006761033	0.0008447665	0.0004280212	0.0008446439	verona
tsp:age2	0.0016318397	0.0011438565	0.0010161151	0.0011570522	verona
tsp:age3	0.0016098791	0.0007823758	0.0022220108	0.0007770380	verona
o3	-0.0010625599	0.0004030629	-0.0006977947	0.0003808588	verona
o3:age1	-0.0005118930	0.0006160911	-0.0002900337	0.0005965191	verona
o3:age2	-0.0024640512	0.0008444777	-0.0013985564	0.0008243813	verona
o3:age3	-0.0008761828	0.0005913074	-0.0007267385	0.0005712741	verona

Tabella C.1: Stime dei coefficienti e degli errori standard ai ritardi 0 e 1.

row.names	coeff.lag2	se.lag2	coeff.lag3	se.lag3	Citta'
so2	0.0075211466	0.0026893432	0.01014067665	0.0026288233	verona
so2:age1	0.0084089558	0.0043962535	0.01108422328	0.0043185725	verona
so2:age2	0.0005902145	0.0061159881	-0.00320705439	0.0060710412	verona
so2:age3	0.0099762951	0.0040276969	0.01543669216	0.0039260284	verona
no2	0.0030646758	0.0006234883	0.00339141566	0.0006129484	verona
no2:age1	0.0024968592	0.0010171721	0.00400077837	0.0009933696	verona
no2:age2	0.0013934872	0.0014156707	0.00150149574	0.0013951876	verona
no2:age3	0.0043436140	0.0009428353	0.00375544675	0.0009345115	verona
co	0.0452643450	0.0090832691	0.03933152093	0.0091025751	verona
co:age1	0.0221893501	0.0149677783	0.03456556488	0.0148507796	verona
co:age2	0.0492575478	0.0204342861	0.03068207060	0.0205211905	verona
co:age3	0.0631265492	0.0137213598	0.04771094399	0.0138574174	verona
tsp	0.0023464646	0.0005052447	0.00279261103	0.0004997973	verona
tsp:age1	0.0020935680	0.0008253413	0.00241758816	0.0008177469	verona
tsp:age2	0.0011178897	0.0011522824	0.00267238106	0.0011230144	verona
tsp:age3	0.0031426548	0.0007660052	0.00317877219	0.0007614630	verona
o3	-0.0001627519	0.0003719085	0.00003345153	0.0003632610	verona
o3:age1	-0.0001479636	0.0005918071	0.00023510620	0.0005811945	verona
o3:age2	-0.0017377345	0.0008233724	-0.00050479180	0.0008056722	verona
o3:age3	0.0005833981	0.0005657737	0.00011224456	0.0005565733	verona

Tabella C.2: Stime dei coefficienti e degli errori standard ai ritardi 2 e 3.

row.names	coeff.lag01	se.lag01	coeff.lag03	se.lag03	Citta'
so2	0.0053538767	0.0031043033	0.0113845712	0.0033378437	verona
so2:age1	0.0068136382	0.0049693181	0.0131078438	0.0053978643	verona
so2:age2	-0.0025242650	0.0067869204	-0.0025100972	0.0074655645	verona
so2:age3	0.0079176283	0.0045841152	0.0166244212	0.0049565612	verona
no2	0.0019747750	0.0007071804	0.0037997810	0.0007434197	verona
no2:age1	0.0011337099	0.0011446861	0.0033165366	0.0012075936	verona
no2:age2	0.0016902935	0.0015701657	0.0021814089	0.0016716628	verona
no2:age3	0.0028437591	0.0010634517	0.0050027244	0.0011205029	verona
co	0.0264317824	0.0103715798	0.0525175955	0.0110114706	verona
co:age1	0.0086688450	0.0169671339	0.0292975560	0.0180168978	verona
co:age2	0.0529855828	0.0230315326	0.0644363419	0.0246810660	verona
co:age3	0.0327255127	0.0157846146	0.0669251178	0.0167345862	verona
tsp	0.0015052131	0.0005402736	0.0028347887	0.0005737838	verona
tsp:age1	0.0006486649	0.0008884832	0.0020796410	0.0009421557	verona
tsp:age2	0.0015321017	0.0012071217	0.0023148408	0.0012875684	verona
tsp:age3	0.0022292332	0.0008186408	0.0037406807	0.0008707714	verona
o3	-0.0011406399	0.0004288029	-0.0007804077	0.0004569375	verona
o3:age1	-0.0005287326	0.0006595074	-0.0002686767	0.0007134008	verona
o3:age2	-0.0024465978	0.0009058142	-0.0025227191	0.0009827685	verona
o3:age3	-0.0010494483	0.0006320501	-0.0003680266	0.0006828717	verona

Tabella C.3: Stime dei coefficienti e degli errori standard ai ritardi 0-1 e 0-3.

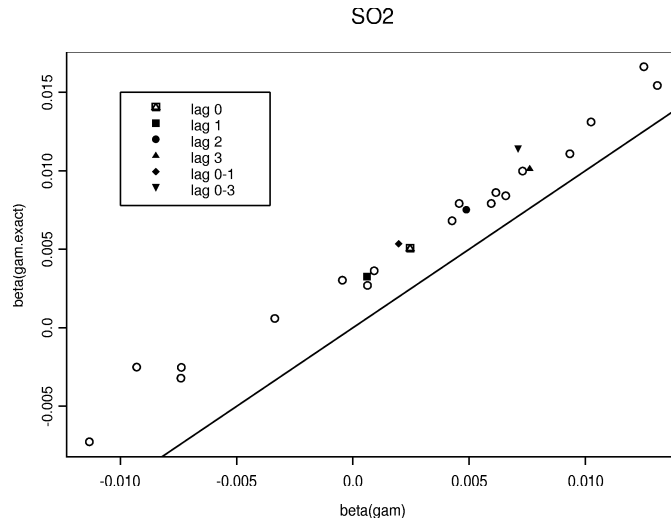
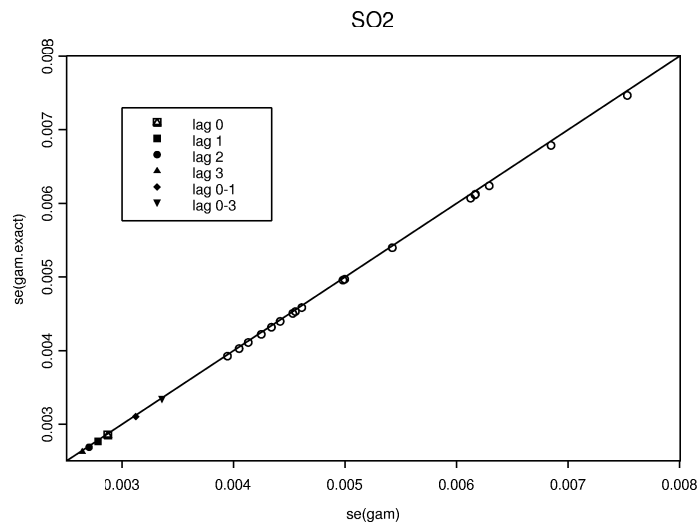
Appendice D

Una sintesi del confronto tra le funzioni *gam* e *gam.exact*

I grafici che seguono rappresentano un confronto tra le stime dei coefficienti β e tra gli errori standard calcolati con le funzioni *gam* e *gam.exact*, per ognuno degli inquinanti presi in considerazione, eccezion fatta per il *TSP*, i cui grafici sono stati riportati nel Paragrafo 4.2.

Le conclusioni che si possono trarre sono comunque analoghe a quelle riportate per le polveri sospese totali: le stime dei coefficienti calcolati con la funzione *gam.exact* sono maggiori di quelle calcolate con la funzione *gam*, mentre i valori degli errori standard rimangono per lo più gli stessi.

Solo l' O_3 , come constatato anche in altre analisi, ha un comportamento diverso: le stime dei coefficienti assumono, infatti, valori maggiori se calcolati con la funzione *gam*. Il comportamento singolare dell'ozono è anche indice della debolezza del disegno dal momento che, assumendo molti dei coefficienti stimati valori negativi, indurrebbe a pensare che l' O_3 non abbia effetti negativi sulla salute umana, ma anzi effetti benefici. Questo è naturalmente assurdo, come confermato anche in letteratura in cui si leggono i numerosi effetti negativi dell'ozono sulla salute umana.

Figura D.1: Confronto tra le stime dei **coefficienti** β (SO_2).Figura D.2: Confronto tra le stime degli **errori standard** (SO_2).

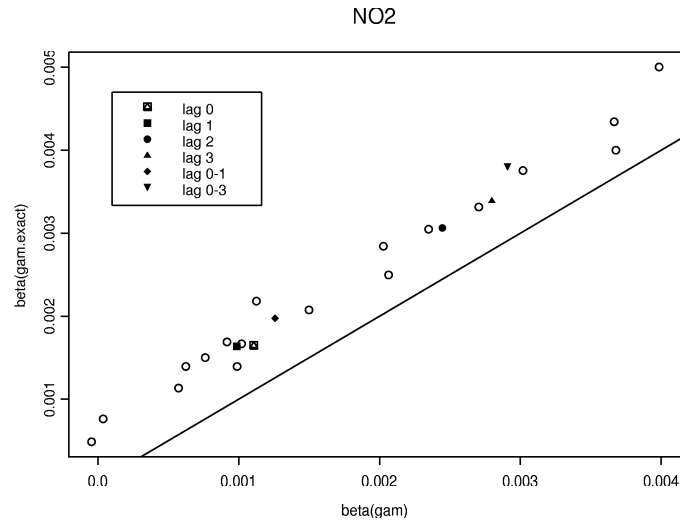


Figura D.3: Confronto tra le stime dei **coefficienti** β (NO_2).

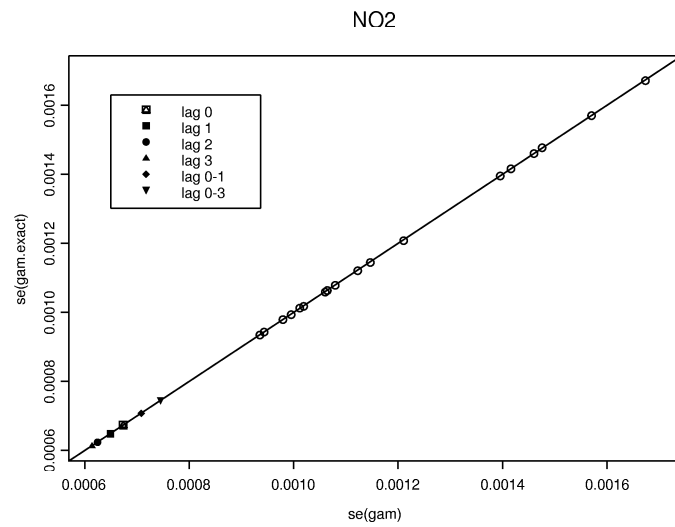
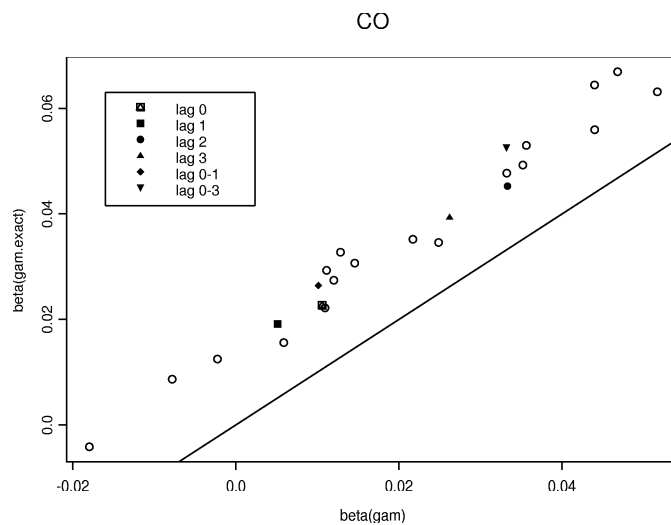
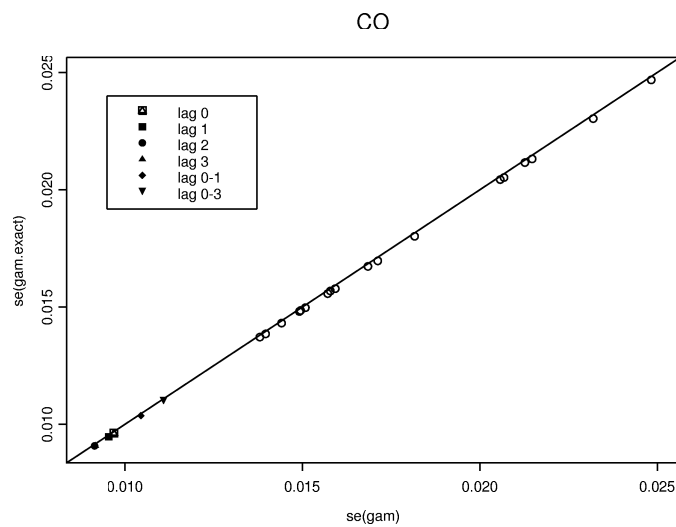


Figura D.4: Confronto tra le stime degli **errori standard** (NO_2).

Figura D.5: Confronto tra le stime dei **coefficienti** β (*CO*).Figura D.6: Confronto tra le stime degli **errori standard** (*CO*).

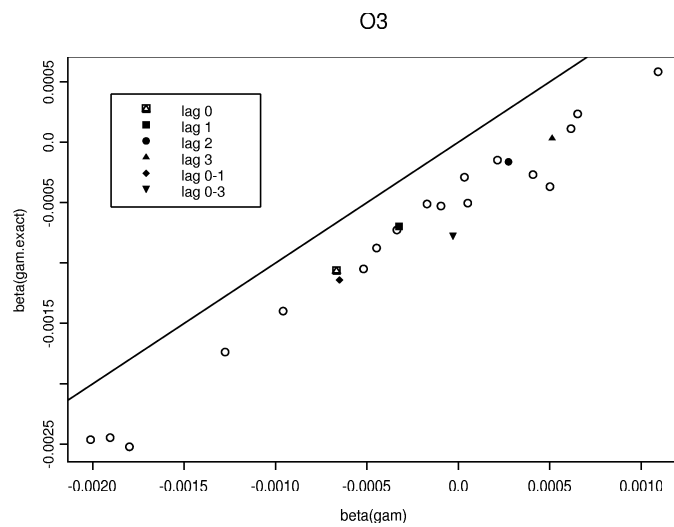


Figura D.7: Confronto tra le stime dei **coefficienti** β (O_3).

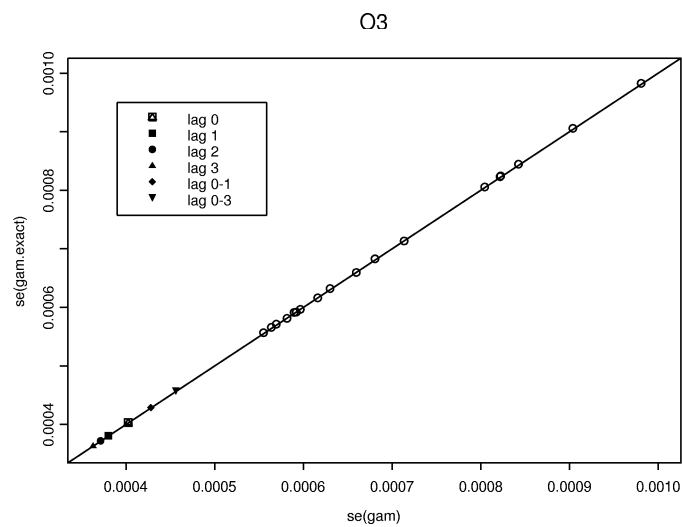


Figura D.8: Confronto tra le stime degli **errori standard** (O_3).

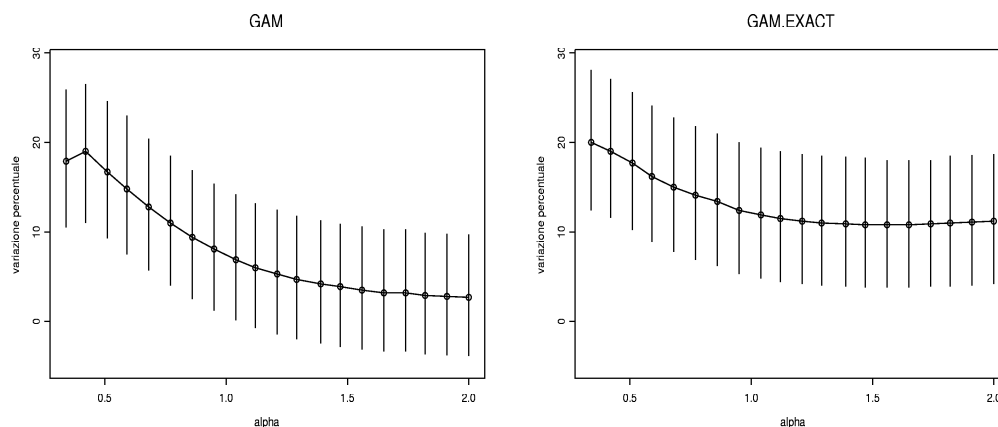
Appendice E

Variazioni percentuali relative all'effetto dell'inquinante

Nei grafici e nelle tabelle presentati in questa Appndice sono riportate le variazioni percentuali e gli intervalli di confidenza approssimati al 95% per diverse scelte di gradi di libertà e per diversi tipi di inquinante.

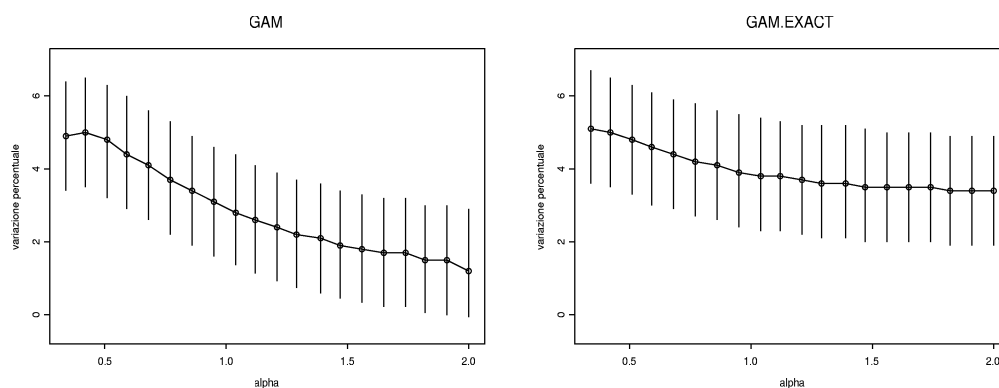
Dai grafici e dalle tabelle riportati si possono trarre le stesse conclusioni a cui si era pervenuti per le poleveri sospese totali, notando così il diverso comportamento, sottolineato evidenziando i valori in grassetto, delle funzioni *gam* e *gam.exact* nel calcolo delle variazioni percentuali all'aumentare dei gradi di libertà.

Si noti il comportamento singolare dell'ozono, come già sottolineato precedentemente.



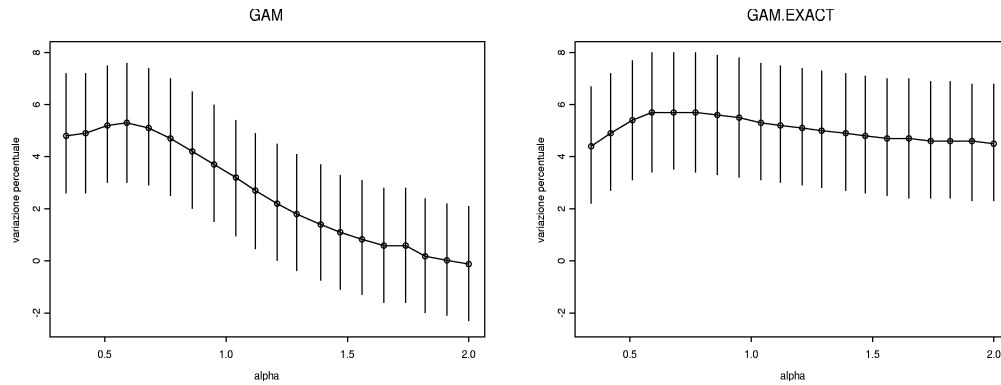
α	GAM			GAM.EXACT		
	variazione percentuale	limite inferiore	limite superiore	variazione percentuale	limite inferiore	limite superiore
0.34	17.9	10.5	25.9	20	12.4	28.1
0.42	19	11	26.5	19	11.6	27.1
0.51	16.7	9.3	24.6	17.7	10.2	25.6
0.59	14.8	7.5	23	16.2	8.9	24.1
0.68	12.8	5.7	20.4	15	7.8	22.8
0.77	11	4	18.5	14.1	6.9	21.8
0.86	9.4	2.5	16.9	13.4	6.2	21
0.95	8.1	1.2	15.4	12.4	5.3	20
1.04	6.9	0.14	14.2	11.9	4.8	19.4
1.12	6.0	-0.73	13.2	11.5	4.4	19
1.21	5.3	-1.44	12.5	11.2	4.2	18.7
1.29	4.7	-2	11.8	11	4	18.5
1.39	4.2	-2.46	11.3	10.9	3.9	18.4
1.47	3.9	-2.83	10.9	10.8	3.8	18.3
1.56	3.5	-3.12	10.6	10.8	3.8	18
1.65	3.2	-3.35	10.3	10.8	3.8	18
1.74	3.2	-3.35	10.3	10.9	3.9	18
1.82	2.9	-3.67	9.9	11	3.9	18.5
1.91	2.8	-3.774	9.8	11.1	4	18.6
2	2.7	-3.847	9.7	11.2	4.2	18.7

Tabella E.1: Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto dell'**anidride solforosa** (SO_2) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu g/m^3$.



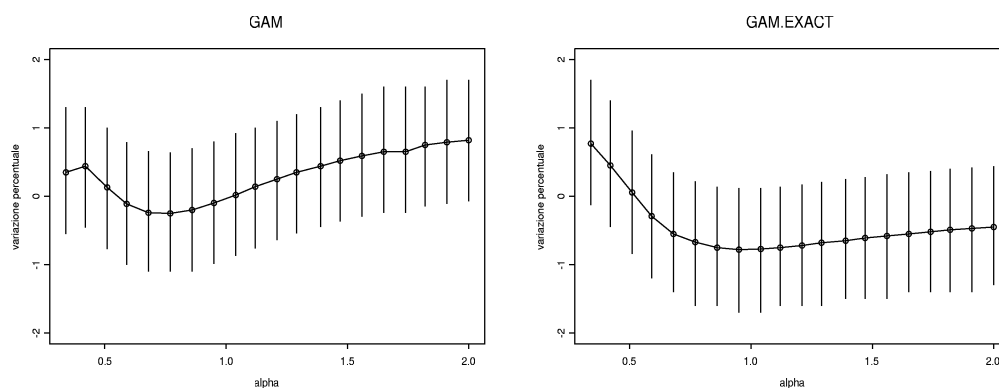
α	GAM			GAM.EXACT		
	variazione percentuale	limite inferiore	limite superiore	variazione percentuale	limite inferiore	limite superiore
0.34	4.9	3.4	6.4	5.1	3.6	6.7
0.42	5	3.5	6.5	5	3.5	6.5
0.51	4.8	3.2	6.3	4.8	3.3	6.3
0.59	4.4	2.9	6	4.6	3	6.1
0.68	4.1	2.6	5.6	4.4	2.9	5.9
0.77	3.7	2.2	5.3	4.2	2.7	5.8
0.86	3.4	1.9	4.9	4.1	2.6	5.6
0.95	3.1	1.6	4.6	3.9	2.4	5.6
1.04	2.8	1.36	4.4	3.8	2.3	5.4
1.12	2.6	1.13	4.1	3.8	2.3	5.3
1.21	2.4	0.92	3.9	3.7	2.2	5.2
1.29	2.2	0.74	3.7	3.6	2.1	5.2
1.39	2.1	0.59	3.6	3.6	2.1	5.1
1.47	1.9	0.45	3.4	3.5	2	5.1
1.56	1.8	0.33	3.3	3.5	2	5
1.65	1.7	0.22	3.2	3.5	2	5
1.74	1.7	0.22	3.2	3.5	2	5
1.82	1.5	0.056	3.0	3.4	1.9	4.9
1.91	1.5	-0.009	3	3.4	1.9	4.9
2	1.2	-0.064	2.9	3.4	1.9	4.9

Tabella E.2: Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto del **biossido di azoto** (NO_2) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu g/m^3$.



α	GAM			GAM.EXACT		
	variazione percentuale	limite inferiore	limite superiore	variazione percentuale	limite inferiore	limite superiore
0.34	4.8	2.6	7.2	4.4	2.2	6.7
0.42	4.9	2.6	7.2	4.9	2.7	7.2
0.51	5.2	3	7.5	5.4	3.1	7.7
0.59	5.2	3	7.6	5.7	3.4	8
0.68	5.1	2.9	7.4	5.7	3.5	8
0.77	4.7	2.5	7	5.7	3.4	8
0.86	4.2	2	6.5	5.6	3.3	7.9
0.95	3.7	1.5	6	5.5	3.2	7.8
1.04	3.2	0.95	5.4	5.3	3.1	7.6
1.12	2.7	0.46	4.9	5.2	3	7.5
1.21	2.2	0.014	4.5	5.1	2.9	7.4
1.29	1.8	-0.39	4.1	5	2.8	7.3
1.39	1.4	-0.75	3.7	4.9	2.7	7.2
1.47	1.1	-1.1	3.3	4.8	2.6	7.1
1.56	0.83	-1.3	3.1	4.7	2.5	7
1.65	0.59	-1.6	2.8	4.7	2.4	7
1.74	0.59	-1.6	2.8	4.6	2.4	6.9
1.82	0.18	-2	2.4	4.6	2.4	6.9
1.91	0.023	-2.1	2.2	4.6	2.3	6.8
2	-0.12	-2.3	2.1	4.5	2.3	6.8

Tabella E.3: Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto del **monossido di carbonio** (CO) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $1mg/m^3$.



α	GAM			GAM.EXACT		
	variazione percentuale	limite inferiore	limite superiore	variazione percentuale	limite inferiore	limite superiore
0.34	0.35	-0.55	1.3	0.77	-0.13	1.7
0.42	0.44	-0.46	1.3	0.45	-0.45	1.4
0.51	0.13	-0.77	1	0.055	-0.84	0.96
0.59	-0.11	-1	0.79	-0.29	-1.2	0.61
0.68	-0.24	-1.1	0.66	-0.55	-1.4	0.35
0.77	-0.25	-1.1	0.64	-0.67	-1.6	0.22
0.86	-0.2	-1.1	0.7	-0.75	-1.6	0.14
0.95	-0.098	-0.99	0.8	-0.78	-1.7	0.12
1.04	-0.018	-0.87	0.92	-0.77	-1.7	0.12
1.12	0.14	-0.76	1	-0.75	-1.6	0.14
1.21	0.25	-0.64	1.1	-0.72	-1.6	0.17
1.29	0.35	-0.54	1.2	-0.68	-1.6	0.21
1.39	0.44	-0.45	1.3	-0.65	-1.5	0.25
1.47	0.52	-0.37	1.4	-0.61	-1.5	0.28
1.56	0.59	-0.3	1.5	-0.58	-1.5	0.32
1.65	0.65	-0.24	1.6	-0.55	-1.4	0.35
1.74	0.65	-0.24	1.6	-0.52	-1.4	0.37
1.82	0.75	-0.15	1.6	-0.49	-1.4	0.4
1.91	0.79	-0.11	1.7	-0.47	-1.4	0.42
2	0.82	-0.075	1.7	-0.45	-1.3	0.44

Tabella E.4: Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto dell'**ozono** (O_3) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu g/m^3$.

Elenco delle figure

3.1	Istogrammi della concentrazione degli inquinanti.	29
3.2	Boxplot della concentrazione degli inquinanti.	30
3.3	Serie storiche giornaliere della concentrazione media rispettivamente di anidride solforosa ($\mu g/m^3$), biossido di azoto ($\mu g/m^3$) e monossido di carbonio (mg/m^3).	32
3.4	Serie storiche giornaliere della concentrazione media rispettivamente di polveri sospese totali ($\mu g/m^3$) e ozono ($\mu g/m^3$).	33
3.5	Boxplot stagionali della concentrazione media giornaliera degli inquinanti.	34
3.6	Serie storiche della concentrazione media mensile rispettivamente di anidride solforosa ($\mu g/m^3$), biossido di azoto ($\mu g/m^3$) e monossido di carbonio (mg/m^3).	35
3.7	Serie storiche della concentrazione media mensile rispettivamente di polveri sospese totali ($\mu g/m^3$) e ozono ($\mu g/m^3$).	36
3.8	Serie storiche della temperatura media giornaliera e mensile ($^{\circ}C$).	37
3.9	Serie storiche dell'umidità media giornaliera e mensile.	38
3.10	Grafici di dispersione relativi ai cinque inquinanti considerati.	41
3.11	Istogrammi della distribuzione dei ricoveri giornalieri.	43
3.12	Serie storiche dei ricoveri giornalieri rispettivamente per le classe d'età 0-64, 65-74 e 75 e più.	44
3.13	Serie storiche mensile dei ricoveri giornalieri rispettivamente per le classe d'età 0-64, 65-74 e 75 e più.	45

4.1	Funzioni di autocorrelazione parziale dei residui in assenza dell'influenza degli inquinanti, separatamente per le tre classi d'età.	53
4.2	Funzioni di autocorrelazione parziale dei residui in presenza di <i>TSP</i>	56
4.3	Confronto tra le stime dei coefficienti β e degli errori standard (<i>TSP</i>).	62
4.4	Variazioni percentuali ed intervalli di confidenza approssimati al 95% relativi all'effetto delle polveri sospese totali (<i>TSP</i>) sulla ricoverabilità, stimate utilizzando rispettivamente la funzione <i>gam</i> e <i>gam.exact</i>	65
D.1	Confronto tra le stime dei coefficienti β (SO_2).	82
D.2	Confronto tra le stime degli errori standard (SO_2).	82
D.3	Confronto tra le stime dei coefficienti β (NO_2).	83
D.4	Confronto tra le stime degli errori standard (NO_2).	83
D.5	Confronto tra le stime dei coefficienti β (CO).	84
D.6	Confronto tra le stime degli errori standard (CO).	84
D.7	Confronto tra le stime dei coefficienti β (O_3).	85
D.8	Confronto tra le stime degli errori standard (O_3).	85

Elenco delle tabelle

3.1	Dataset utilizzato nell'analisi.	26
3.2	Statistiche descrittive della concentrazione giornaliera degli inquinanti e delle variabili meteorologiche.	28
3.3	Matrice di correlazione tra inquinanti e variabili meteorologiche.	40
4.1	Significatività di alcuni termini nel <i>core</i>	50
4.2	Residui del modello in assenza dell'influenza degli inquinanti. .	52
4.3	Residui del modello in presenza di <i>TSP</i>	55
4.4	Variazioni percentuali e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95%, stimati mediante la funzione <i>gam</i> e relativi all'effetto di ogni inquinante sui ricoveri ospedalieri.	58
4.5	Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95%, stimati mediante la funzione <i>gam.exact</i> e relativi all'effetto di ogni inquinante sui ricoveri ospedalieri. In grassetto sono evidenziate le differenze significative rispetto ai risultati ottenuti con la funzione <i>gam</i> : in questo caso si nota un effetto significativo dell'inquinante sulla salute.	60

4.6	Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto delle polveri sospese totali (<i>TSP</i>) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu\text{g}/\text{m}^3$. In grassetto sono evidenziate le differenze significative tra i risultati ottenuti applicando le due funzioni <i>gam</i> e <i>gam.exact.</i> .	66
B.1	Stime dei coefficienti e degli errori standard ai ritardi 0 e 1. . .	74
B.2	Stime dei coefficienti e degli errori standard ai ritardi 2 e 3. . .	74
B.3	Stime dei coefficienti e degli errori standard ai ritardi 0-1 e 0-3.	75
C.1	Stime dei coefficienti e degli errori standard ai ritardi 0 e 1. . .	78
C.2	Stime dei coefficienti e degli errori standard ai ritardi 2 e 3. . .	78
C.3	Stime dei coefficienti e degli errori standard ai ritardi 0-1 e 0-3.	79
E.1	Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto dell' anidride solforosa (SO_2) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu\text{g}/\text{m}^3$. . .	88
E.2	Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto del biossido di azoto (NO_2) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu\text{g}/\text{m}^3$	89
E.3	Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto del monossido di carbonio (CO) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $1\text{mg}/\text{m}^3$. . .	90

E.4	Variazione percentuale e limiti inferiore e superiore degli intervalli di confidenza approssimati al 95% relativi all'effetto del'ozono (O_3) sulla ricoverabilità (lag 0-3). I valori sono espressi in termini di variazioni percentuali associate ad un incremento dell'inquinante pari a $10\mu g/m^3$	91
-----	---	----

Bibliografia

- [1] T.J. Hastie, R.J. Tibshirani. *Generalized Additive Models* (1990). Chapman & Hall/CRC.
- [2] J. Schwartz, C. Spix, G. Touloumi, L. Bachárová, T. Barumamdzadeh, A. le Tertre, T. Piekarksi, A. Ponce de Leon, A. Pönkä, G. Rossi, M. Saez, J.P. Schouten. *Methodological Issues in studies of air pollution and daily counts of deaths or hospital admissions* (Journal of Epidemiology and Community Health, 1996, n° 50, suppl.1).
- [3] Richard L. Smith, Jerry M. Davis, Jerome Sacks, Paul Speckman, Patricia Styer. *Regression models for air pollution and daily mortality: analysis of data from Birmingham, Alabama* (Environmetrics, 2000, n° 11).
- [4] J.M. Samet, F. Dominici, F. Curriero, I. Coursac and S.L. Zeger. *Particulate Air Pollution and Mortality: Findings from 20 U.S. Cities* (New England Journal of Medicine, 343, 24, 1742-757 (2000)).
- [5] *Epidemiologia e prevenzione* (2001 suppl.2). Zadig Editore.
- [6] Timothy O. Ramsay, Richard T. Burnett and Daniel Krewski. *The Effect of Concurvity in Generalized Additive Models Linking Mortality to Ambient Particulate Matter* (2002).
- [7] S.L. Zeger, F. Dominici, A. McDermott, J.M. Samet. *Modeling the Health Effects of Environmental Exposures: Particulate Air Pollution*

- and Mortality in Urban America* (in Monitoring the Health of Populations Statistical Methods for Public Health Surveillance, (Edited by Brookmeyer R. and Stroup D.), Oxford University Press 2002).
- [8] Francesca Dominici, Lianne Sheppard and Merlise Clyde. *Health effects of air pollution: a statistical review* (2003).
- [9] Francesca Dominici, Aidan McDermott and Trevor Hastie. *Issues in semi-parametric regression with applications in time series models for air pollution and mortality* (2003).
- [10] *Revised Analyses of Time-Series Studies of Air Pollution and Health* (Health Effects Institute, maggio 2003).
- [11] M. Daniels, F. Dominici, S.L. Zeger. *Underestimation of Standard Errors in Multi Site Time Series Studies* (2003).
- [12] Francesca Dominici, Aidan McDermott, Scott L. Zeger, Jonathan M. Samet, Trevor Hastie. *Health Effects of Air Pollutions: Statistical Challenges, Findings, and Policy Implications (Part I e II)* (Department of Biostatistics, Bloomberg School of Public Health, Johns Hopkins University, 2003).