

UNIVERSITA' DEGLI STUDI DI PADOVA
Facoltà di Scienze Statistiche
Corso di Laurea in Scienze Statistiche ed Economiche

Tesi di Laurea

Anomalie nelle serie temporali del biossido di zolfo
a Sacca Fisola (Venezia)

Relatore: Ch.mo Prof. Carlo Gaetan

Correlatore: Ch.mo Prof. Andrea Pastore

Laureanda: Erica Passadore

ANNO ACCADEMICO 2004 – 2005

Indice

Introduzione	3
Capitolo 1 Inquinamento atmosferico	7
1.1 Introduzione	7
1.2 Il controllo della qualità dell'aria	10
1.3 Inquinamento da ossidi di zolfo	14
1.4 La normativa	17
Capitolo 2 Analisi esplorative	21
2.1 La rete di monitoraggio di Venezia	21
2.2 Analisi dei dati	23
2.3 La variabile SO ₂	26
2.4 Le variabili meteorologiche	29
2.4.1 Temperatura	30
2.4.2 Umidità relativa	33
2.4.3 Vento	34
2.5 Correlazioni	39
Capitolo 3 Identificazione dei punti di cambiamento in una serie temporale	41
3.1 Introduzione	41
3.2 Modelli a funzione di trasferimento	43
3.3 Intervention analysis	45
3.4 Iterative Stepwise Gauss Newton algorithm (ISGN)	48
3.5 Algoritmi genetici	51
3.6 Procedura di Chen e Liu	54
Capitolo 4 Applicazione ai dati	57
4.1 Introduzione	57
4.2 Stima di un modello a funzione di trasferimento	58
4.3 Punti di cambiamento	65
4.4 La serie annuale	74
4.5 Confronti	77
Capitolo 5 Conclusioni	79

Appendice 1	Ulteriori risultati	81
Appendice 2	Tabella dei punti di cambio	87
Appendice 3	Principali funzioni del software R utilizzate	91
Bibliografia		95

Introduzione

La conservazione della qualità dell'aria che respiriamo è diventata materia di legge dal 1988 quando, per la prima volta, il DPR 203/88 definisce l'inquinamento atmosferico come: "ogni modificazione della normale composizione o stato fisico dell'aria atmosferica, dovuta alla presenza nella stessa di una o più sostanze in quantità e con caratteristiche tali da alterare le normali condizioni ambientali e di salubrità dell'aria; da costituire pericolo ovvero pregiudizio diretto o indiretto per la salute dell'uomo; da compromettere le attività ricreative e gli altri usi legittimi dell'ambiente; da alterare le risorse biologiche e gli ecosistemi ed i beni materiali pubblici e privati".

Da allora sono stati compiuti passi importanti quali: l'introduzione di valori limite e valori guida della qualità dell'aria; l'introduzione di livelli di attenzione e di allarme, relativi alle componenti inquinanti dell'aria; la realizzazione e il consolidamento di piani di risanamento nelle zone definite più a rischio. Tutti questi progetti sono stati periodicamente rivisti e aggiornati in base anche alle normative europee.

Uno degli inquinanti atmosferici che la letteratura internazionale considera più pericoloso è l' SO_2 . Il biossido di zolfo si sprigiona dalla combustione di tutte le sostanze contenenti zolfo e produce effetti dannosi alla salute dell'uomo, specie alle fasce sensibili della popolazione. Il suo accumulo e i suoi effetti sulle piante e sul suolo possono essere evidenti anche dopo venti anni. Il vasto e

complesso argomento dell'inquinamento atmosferico è introdotto nel capitolo 1 di questa tesi.

L' SO_2 è proprio uno dei inquinanti tossici sorvegliati nel comune di Venezia, dove, dal 1999, opera una rete di monitoraggio dell'inquinamento atmosferico gestita dal Dipartimento Provinciale di Venezia dell'Agenzia regionale per la Prevenzione e Protezione Ambientale del Veneto (ARPAV).

Grazie ai dati delle concentrazioni dell' SO_2 nell'aria, concessi dalla rete dell'ARPAV, si è potuto prendere in esame il problema del monitoraggio "off line" delle serie temporali.

La ricerca e la comprensione dei punti di cambiamento delle serie temporali di questo inquinante si è svolta a posteriori, cioè senza conoscere la natura dell'anomalia e il momento in cui questa avviene.

La stima di un modello a funzione di trasferimento per la serie del biossido di zolfo e di alcune variabili meteorologiche è stata solo la fase iniziale della procedura che ha portato all'identificazione di *outliers* e cambi di livello. Questi sono stati ricercati con l'utilizzo congiunto degli algoritmi genetici e di un algoritmo iterativo di tipo *Gauss Newton*.

Tutte le tecniche adottate per lo studio dei punti di cambiamento sono presentate nel capitolo 3.

A causa della calibrazione degli strumenti di rilevamento, o perchè le centraline erano spente, molti dati nel *dataset* erano mancanti, quindi si sono previsti i tratti mancanti per utilizzarli in tutte le analisi seguenti.

Con il *dataset* completo di tutti i dati per i due anni considerati, 2001 e 2002, si è potuto analizzare dal punto di vista descrittivo le serie delle concentrazioni orarie di biossido di zolfo, le serie delle

misurazioni orarie di temperatura e umidità relativa, e i dati relativi al vento. I risultati ottenuti sono riportati nel capitolo 2.

L'applicazione delle tecniche per la ricerca dei punti di cambio alle serie dei dati reali a disposizione è affrontata nel capitolo 4. Qui lo studio è stato fatto su serie di trimestri di dati orari, e i punti di cambio risultati sono stati confrontati con quelli ottenuti dall'analisi della serie di concentrazioni medie giornaliere di SO₂.

Nel capitolo 5 trovano spazio le conclusioni sul lavoro svolto. Queste valgono solo per gli anni considerati nelle analisi e non hanno carattere di previsione sulle condizioni dell'aria nel sito di rilevamento di Sacca Fisola.

Inoltre non è stato possibile confrontare i risultati ottenuti con le analisi di dati provenienti da altre stazioni di rilevamento della zona, in quanto i *dataset* corrispondenti non erano adatti ad essere analizzati con le stesse tecniche; alcuni presentavano una quantità molto elevata di dati mancanti per le concentrazioni di SO₂, che sarebbe stato impossibile prevedere, altri non erano provvisti dei rilevamenti sulle variabili meteorologiche.

Nell'appendice 1 si presentano i risultati dell'identificazione dei punti di cambiamento di altre quattro serie di trimestri di dati non riportate nel capitolo 4. L'appendice 2 mostra la tabella con i punti di cambiamento avvenuti negli stessi giorni sia per l'analisi annuale della serie sia per l'analisi dei trimestri. Nell'appendice 3 si spiega brevemente l'utilizzo delle funzioni del software R adoperate per le analisi.

Capitolo 1

Inquinamento atmosferico

1.1 Introduzione

In questo momento dell'era postindustriale la società si sta rendendo conto dei danni causati dall'uomo all'ambiente e di quelli che, ancora e consapevolmente, continua a determinare.

Fenomeni di inquinamento a scala mondiale e a scala locale rappresentano una minaccia per l'esistenza della vita sul pianeta Terra. (Zavatti, Biondi, Mazzali, 1994)

Pur con notevole ritardo rispetto agli altri paesi a sviluppo avanzato, si è dato vita anche in Italia ad un progetto innovativo e pratico nel campo della difesa ambientale. Questo piano è strutturato in una serie di interventi operativi stabiliti con Decreto del Ministero dell'Ambiente di concerto con il Ministero della Sanità.

Uno tra i problemi più significativi presi in considerazione dalla legislatura in materia di difesa ambientale è rappresentato dall'inquinamento atmosferico, che già nel DPR 203/88 è definito, nel suo concetto più ampio, come: "ogni modificazione della normale composizione o stato fisico dell'aria atmosferica, dovuta alla presenza nella stessa di una o più sostanze in quantità e con caratteristiche tali da alterare le normali condizioni ambientali e di salubrità dell'aria; da costituire pericolo ovvero pregiudizio diretto o

indiretto per la salute dell'uomo; da compromettere le attività ricreative e gli altri usi legittimi dell'ambiente; da alterare le risorse biologiche e gli ecosistemi ed i beni materiali pubblici e privati”.

L'inquinamento atmosferico, nella sua natura complessa, può essere definito sulla base dei fenomeni che lo determinano.

Si distingue dunque tra inquinamento di origine naturale e di origine antropica, dipendente cioè dalla presenza dell'uomo e dalle sue attività.

Anche se al giorno d'oggi l'inquinamento di origine naturale è di entità trascurabile rispetto a quello di origine antropica, è comunque opportuno ricordare che, prima dell'avvento dell'era industriale, solamente le attività vulcaniche, i processi biologici degli organismi viventi e fenomeni di combustione spontanea rappresentavano le uniche fonti di contaminazione ambientale. Col progredire dello sviluppo tecnologico, la presenza di questi inquinanti è diminuita percentualmente nell'ambito dell'inquinamento atmosferico fino a diventare del tutto irrilevante al giorno d'oggi.

A questi inquinanti naturali si aggiungono dunque gli inquinanti derivati dalle attività umane che, se fossero uniformemente distribuiti, probabilmente non porterebbero a particolari modifiche della composizione dell'aria.

Il problema dell'inquinamento di origine antropica, per la maggior parte delle sostanze prodotte, deriva proprio dal fatto di essere particolarmente concentrato in piccole aree, soprattutto urbane, e di non essere in grado di diffondere nell'intera atmosfera per la presenza di barriere, sia geomorfologiche sia create dall'uomo, che ne limitano la diffusione laterale. Si assiste quindi ad un vero e

proprio accumulo, accentuato in alcuni giorni anche da particolari condizioni meteorologiche.

Considerando i possibili inquinanti è utile distinguere tra inquinanti primari e secondari:

- gli inquinanti primari sono presenti nelle emissioni ed intervengono direttamente sulla salute umana;
- gli inquinanti secondari invece sono frutto di reazioni tra i primari, o tra i primari e i componenti naturali dell'atmosfera.

In generale si può affermare che l'inquinamento atmosferico è certamente nocivo alla salute dell'uomo, sia pure in un quadro di estrema variabilità, dato che nessun singolo inquinante può essere ritenuto l'unico responsabile delle manifestazioni osservate e che gli effetti sembrano essere prodotti da tutto l'insieme dell'ambiente atmosferico, comprese le condizioni meteorologiche e le vicende stagionali. Per un approfondimento sul tema si veda Finzi, Brusasca (1991).

In questa visione articolata e complessa della condizione ambientale trova una valida applicazione la valutazione dell'impatto ambientale che è basata sulla nuova filosofia della "prevenzione" del danno ambientale.

Essa considera gli effetti che potrebbero manifestarsi nell'ambiente conseguenti alla realizzazione di interventi ed opere destinate ad incidere sul territorio. L'approccio valutativo deve essere strettamente connesso con l'attività di pianificazione, inducendo modifiche nei modelli di comportamento della società, con l'obiettivo di garantire il benessere e l'espansione sociale ed economica anche alle generazioni future. Ciò conduce Zavatti, Biondi, Mazzali (1994) ad affermare che valutazione di impatto ambientale e pianificazione

debbono interagire per essere in grado di contenere gli effetti negativi derivanti dalle azioni dell'uomo sul territorio.

1.2 Il controllo della qualità dell'aria

Il controllo delle condizioni ambientali prevede, innanzitutto, l'individuazione di una fascia di territorio entro cui effettuare le indagini e le misure relative agli indicatori; successivamente è opportuno distinguere, all'interno dell'ambito delimitato, le aree di diversa sensibilità e criticità ambientale per le quali si richiedono differenti livelli di densità o precisione dei rilevamenti; quindi occorre definire i criteri per l'effettuazione degli accertamenti e delle verifiche, dopo aver localizzato i punti di misura e le stazioni fisse.

Il controllo della qualità dell'aria viene effettuato rilevando i livelli di concentrazione degli inquinanti presenti nell'aria di una determinata area e, successivamente, valutando i risultati ottenuti per verificare il rispetto degli standard di qualità dell'aria.

Per un'efficace lettura della documentazione acquisita è necessario, infine, individuare tipologie opportune di metodi e tecniche per la restituzione delle informazioni raccolte.

Il prelievo dei campioni di aria comprende l'insieme delle tecniche e delle operazioni effettuate per la captazione degli inquinanti in essi contenuti.

L'analisi consiste nell'insieme delle tecniche e delle operazioni che permettono di determinare la concentrazione degli inquinanti nei campioni di aria prelevati.

Qualora venissero rilevati superamenti non occasionali dei limiti occorre impostare studi specifici e finalizzati all'individuazione delle cause di deterioramento della qualità atmosferica. Occorre inoltre valutare l'entità del superamento al fine di programmare interventi migliorativi per raggiungere situazioni ambientali di rispetto degli standard qualitativi.

Come affermano Zavatti, Biondi, Mazzali, (1994), il controllo della qualità dell'aria deve essere considerato anche come verifica dei risultati ottenuti dall'attuazione dei piani e misure adottate per il risanamento.

La rilevazione delle concentrazioni di inquinanti non deve pertanto essere finalizzata a sé stessa, ma anche all'impostazione dei piani di risanamento.

Le aree in cui è necessario effettuare il controllo della qualità dell'aria devono essere identificate in base alla dislocazione e alla intensità delle sorgenti di emissione, al trasporto degli inquinanti dalle zone di emissione, alla presenza e densità della popolazione.

Le aree prescelte devono essere, possibilmente, rappresentative di zone abbastanza estese e le rilevazioni in esse effettuate devono fornire informazioni utili per lo studio complessivo della qualità dell'aria di un intero agglomerato urbano.

La dislocazione di ogni postazione è molto importante e una scelta inadeguata può ridurre notevolmente la significatività dei risultati delle misure.

Una postazione di rilevamento è rappresentativa quando i dati da essa forniti danno informazioni sui livelli e sulle fluttuazioni delle concentrazioni degli inquinanti, caratteristici della zona.

Il numero delle postazioni di rilevamento nell'area sottoposta a controllo deve essere stabilito principalmente in funzione di:

- densità della popolazione;
- variabilità dello spazio delle concentrazioni degli inquinanti;
- densità, distribuzione e intensità delle sorgenti di emissione;
- caratteristiche topografiche, meteorologiche e orografiche dell'area sottoposta a controllo.

L'individuazione degli inquinanti da misurare deve essere effettuata sulla base di un censimento di tutte le emissioni atmosferiche di origine naturale, industriale e civile che possono determinare contaminazioni nell'area in cui si intende procedere alla verifica del rispetto degli standard di qualità dell'aria. (Zavatti, Biondi, Mazzali, 1994)

Gli inquinanti da privilegiare sono quelli per i quali è stato fissato lo standard di qualità.

Per una migliore interpretazione dei dati di concentrazione è opportuno corredare le concentrazioni rilevate con la misura di alcuni parametri meteorologici, come temperatura, direzione e velocità del vento.

I risultati delle indagini preliminari dovranno essere valutati con molta cautela; solamente nel caso che i risultati ottenuti indichino livelli di concentrazione molto al di sotto degli standard di qualità dell'aria si potrà concludere che non è necessario procedere ad un controllo sistematico nella zona oggetto dell'indagine.

Quando i risultati di un'indagine preliminare di adeguata durata indicano un rispetto sostanziale degli standard, si potrà scegliere fra

gli inquinanti presenti quello che meglio si presta come tracciante e fare di questo misurazioni periodiche.

Gli studi medici sugli effetti dell'inquinamento atmosferico non sono ancora molto noti al grande pubblico.

Le nostre vie respiratorie sono sensibili agli influssi di virus, batteri, fumo di tabacco. Gli inquinanti atmosferici provenienti dal traffico, dai riscaldamenti, dall'industria, artigianato, agricoltura rappresentano un carico che si aggiunge a quello "pre inquinamento".

Tutte le ricerche mostrano che gli effetti della polluzione atmosferica sono multifattoriali e che non c'è una specifica malattia da inquinamento. Gli effetti sono così complessi che non possono essere ricondotti ad una correlazione semplice, e come difficoltà ulteriore allo studio si è osservato che i disturbi hanno un loro corso molto lento ed in maniera nascosta.

Anche se ogni inquinante ha un effetto specifico prevalente, gli effetti dannosi dell'inquinamento sulla salute umana globalmente sono: nausea da odori, irritazioni al naso, gola e occhi, impedimenti nelle funzioni polmonari, diminuzione nelle prestazioni fisiche, aumento della suscettibilità alle malattie vere e proprie della respirazione, aumentata incidenza delle bronchiti croniche e dell'asma, aumentato rischio delle malattie cardiovascolari. Per un approfondimento vedere Zavatti, Biondi, Mazzali, (1994).

1.3 Inquinamento da ossidi di zolfo

Dalla combustione di ogni materiale contenente zolfo si producono tipi di ossidi di questo elemento: il biossido di zolfo SO_2 , e il triossido di zolfo SO_3 ; indicati con il termine generale di SO_x .

Essi sono i principali imputati dell'inquinamento atmosferico da ossidi di zolfo delle aree urbane e industriali dove l'elevata densità degli insediamenti ne favorisce l'accumulo soprattutto in condizioni meteorologiche di debole ricambio delle masse d'aria.

Le loro caratteristiche più rilevanti sono: l'assenza di colore, la natura pungente ed irritante dell'odore, la non infiammabilità e la non esplosività.

Lo zolfo presente nell'atmosfera proviene per circa due terzi da fonti naturali di origine endogena, per esempio dalle emissioni vulcaniche, mentre per la restante parte dalle varie attività dell'uomo.

Il problema principale è legato al fatto che la distribuzione uniforme e l'alta quota cui ha luogo l'emissione naturale di SO_2 , nel caso dei vulcani, fanno sì che questo contributo non abbia effetti di rilievo; mentre invece il secondo tipo di emissioni, essendo concentrato su aree urbane industriali abbastanza ristrette, non è distribuito equamente.

I maggiori imputati dell'inquinamento da SO_x prodotti in tali aree sono gli impianti di combustione fissi; una parte infatti proviene dalla combustione di carbone e la restante dall'uso di olii combustibili.

E' da notare che, in seguito alla diffusa metanizzazione degli impianti di riscaldamento domestici, il contributo inquinante degli ossidi di zolfo, prodotti in questo modo, è notevolmente diminuito nel corso degli anni.

Finzi e Brusasca (1991) sostengono che il traffico urbano non è dunque la fonte principale dell'inquinamento da SO_x e anzi vi contribuisce solo in minima parte.

La quantità di acido solforico nell'atmosfera non è dovuta unicamente alla produzione primaria di SO_3 ; infatti una volta nell'atmosfera l' SO_2 è parzialmente convertita in SO_3 e quindi in acido solforico (H_2SO_4) da processi fotolitici e catalitici. Questa conversione è influenzata da numerosi fattori, fra i quali l'umidità dell'aria, l'intensità, la durata e la distribuzione spettrale della luce solare, la presenza di diverse concentrazioni di sostanze catalizzatrici o assorbenti. Per approfondire il tema della conversione dell' SO_2 in acido solforico si veda Baird (2001).

L' SO_2 è considerato nella letteratura internazionale il più pericoloso degli inquinanti atmosferici, a causa della particolare sensibilità mostrata da alcune fasce di popolazione, come gli anziani e coloro che già soffrono, o sono soggetti, a malattie croniche dell'apparato respiratorio – vascolare; mentre non vi sono prove di una relazione diretta fra esposizione continua a SO_2 , alle concentrazioni normali presenti nell'aria ambiente, e malattie respiratorie in persone sane.

Inoltre sono rari i casi preoccupanti di esposizione a SO_2 in quanto il gas è così irritante che già a concentrazioni basse funziona esso stesso da allarme.

Esso viene assorbito dalle membrane mucose del naso e dell'apparato respiratorio superiore, mentre solo in minima parte entra nell'apparato respiratorio inferiore. Dall'apparato respiratorio entra quindi nel sangue per venire poi eliminato attraverso l'urina.

Effetti acuti sono: l'impedimento polmonare e, per gli asmatici, un aumento nella frequenza di attacchi.

Gli effetti cronici sono più frequenti di quelli sintomatici, specialmente le bronchiti croniche.

Gli ossidi di zolfo possono avere effetti rilevanti anche sulle piante.

La foglia è un recettore efficiente, essa ha peli e può avere una superficie adesiva che facilita la cattura; zuccheri, acidi organici e materiali essudano sulla sua superficie. La cera è ridotta naturalmente dai processi di deterioramento ambientale. L'acido solforico è il principale responsabile delle "piogge acide", che agiscono sulle piante necrotizzando zone delle foglie che successivamente scoloriscono e seccano.

Molti studiosi, come Baird (1994), ritengono che l'accumulo di inquinanti tossici nelle piante e nei suoli può essere evidente anche dopo 20 anni e più.

Un'ultima conseguenza negativa dell'inquinamento da ossidi di zolfo è data dall'effetto corrosivo su alcuni materiali, come i metalli e le vernici.

Nei metalli la velocità di corrosione viene accelerata notevolmente in ambienti inquinati da SO_2 , quali sono le zone industriali e quelle con emissioni dovute all'uso di combustibile per il riscaldamento. I metalli più attaccati sono il ferro, l'acciaio e lo zinco.

Nelle vernici si ha che il loro tempo di essiccamento ed indurimento aumenta dopo l'esposizione ad SO_2 , per cui alcune pellicole diventano più molli ed altre più fragili, fatto questo che influisce sulla loro durata.

1.4 La normativa

Il rilevamento della qualità dell'aria mediante sistemi automatici fissi risale alla metà degli anni settanta, principalmente con l'obiettivo di controllare gli impianti industriali.

E' solo negli anni ottanta che vengono introdotti limiti sulla qualità dell'aria.

Col DPCM del 28 Marzo 1983 vengono fissati, per la prima volta in Italia, i "limiti massimi di accettabilità delle concentrazioni e i limiti massimi di esposizione relativi ad inquinanti dell'aria nell'ambiente esterno", i cosiddetti standard di qualità.

Tali valori sono stati modificati dal successivo DPR n. 203/88, decreto che, recependo alcune Direttive Comunitarie in materia di inquinamento atmosferico, ha adeguato gli standard di qualità dell'aria alle disposizioni normative europee ed ha introdotto, accanto ai limiti massimi, i valori guida di qualità dell'aria; ovvero le concentrazioni da raggiungere progressivamente per garantire la massima tutela dell'ambiente e della salute umana.

Si definiscono:

- Valori limite di qualità dell'aria: limiti massimi di accettabilità delle concentrazioni e limiti massimi di esposizione relativi ad inquinanti nell'ambiente esterno.
- Valori guida di qualità dell'aria: "limiti delle concentrazioni e limite di esposizione relativi ad inquinanti nell'ambiente esterno destinati:
 - a. Alla prevenzione a lungo termine in materia di salute e protezione dell'ambiente;

- b. A costituire parametri di riferimento per l'istituzione di zone specifiche di protezione ambientale per le quali è necessaria una particolare tutela della qualità dell'aria."

Dal 1 gennaio 2005 i valori limite e i valori guida vengono aggiornati ulteriormente, e rimangono vigenti fino ad oggi. Tali limiti vengono inoltre precisati per ciascun tipo di inquinante. Nella tabella 1 sono esposti i limiti standard per il controllo delle emissioni di biossido di zolfo, SO₂.

Soglia di allarme	500 µg/m ³
Limite orario da non superare più di 24 volte per anno civile	350 µg/m ³
Limite protezione ecosistemi	20 µg/m ³

Tabella 1: limiti standard per la qualità dell'aria, con particolare riferimento a SO₂.

Il DM 20/05/1991 prevede per le aree urbane le tipologie delle postazioni di rilevamento su cui deve essere pianificata una rete di rilevamento:

- A. Stazioni di base o di riferimento sulle quali misurare tutti gli inquinanti primari e secondari ed i parametri meteorologici di base nonché inquinanti da valutarsi con metodologie analitiche e manuali. Tali stazioni debbono essere preferenzialmente localizzate in aree non direttamente interessate dalle sorgenti di emissione urbana (parchi, isole pedonali, ecc.);

- B. Stazioni situate in zone ad elevata densità abitativa nelle quali misurare la concentrazione di alcuni inquinanti primari e secondari;
- C. Stazioni situate in zone ad elevato traffico per la misura degli inquinanti emessi direttamente dal traffico veicolare, situate in zone ad alto rischio espositivo quali strade ad elevato traffico e bassa ventilazione. In tal caso, i valori di concentrazione rilevati sono caratterizzati da una rappresentatività limitata alle vicinanze del punto di prelievo;
- D. Stazioni situate in periferia o in aree suburbane finalizzate alla misura degli inquinanti fotochimici, da pianificarsi sulla base di campagne preliminari di valutazione dello smog fotochimico particolarmente nei mesi estivi.

Il DM 15/04/94, aggiornato ed integrato dal DM 25/11/94, definisce i livelli di “attenzione” e di “allarme”, stabilendo i criteri per l'individuazione degli stati di emergenza in funzione dei dati rilevati dai vari tipi di stazioni di monitoraggio installate nelle aree urbane, nonché gli obblighi di informazione alla popolazione sui livelli di inquinamento raggiunti.

Il superamento dei livelli comporta l'entrata nello stato di attenzione e nello stato di allarme, definiti come:

- Stato di attenzione: una situazione di inquinamento atmosferico che, se persistente, determina il rischio che si raggiunga lo stato di allarme;
- Stato di allarme: una situazione di inquinamento atmosferico che, se persistente, determina una potenziale condizione di superamento dei limiti massimi di accettabilità e di rischio sanitario per la popolazione.

I dati delle reti sono utilizzati sia dalle strutture sanitarie per definire il rischio sulla popolazione esposta e per indagini epidemiologiche, che dalle pubbliche amministrazioni per verificare i risultati di misure adottate per contenere l'inquinamento e per predisporre piani di risanamento.

Capitolo 2

Analisi esplorative

2.1 La rete di monitoraggio di Venezia

Il monitoraggio dell'inquinamento atmosferico nel territorio veneziano è oggetto di profondo interesse sin dai primi anni '70; questo in conseguenza della peculiarità dell'area, nella quale coesistono un ecosistema estremamente delicato, un'elevata densità abitativa ed una zona altamente industrializzata.

La rete di monitoraggio presente sul territorio provinciale di Venezia è attiva, nella sua veste attuale, a partire dal 1999.

Un numero rilevante di stazioni si colloca nel territorio del Comune di Venezia, come mostra la figura 1.

La rete urbana del Comune e la rete della Provincia sono gestite dal Dipartimento Provinciale di Venezia dell'Agenzia regionale per la Prevenzione e Protezione Ambientale del Veneto (ARPAV).

Nel suo complesso la rete gestita dall'ARPAV è composta da 15 stazioni di rilevamento fisse e da due laboratori mobili.

Le stazioni della rete ARPAV per il controllo dell'inquinamento atmosferico in Comune di Venezia sono classificate nel seguente modo:

- 2 stazioni di tipo A: la stazione di Parco Bissuola è ubicata in un'area verde di Mestre; la stazione di via Bottenigo è situata

in un'area di Marghera non direttamente influenzata da fonti di emissione significative ed è attrezzata anche per il rilevamento dell'inquinamento di origine industriale;

- 3 stazioni di tipo B: per la misura dell'inquinamento presente in aree densamente urbanizzate (Venezia: Sacca Fisola; Mestre: viale San Marco; via A. Da Mestre, che funziona a regime a partire dal marzo 2001);
- 4 stazioni di tipo C: le stazioni per il rilevamento dell'inquinamento da traffico autoveicolare sono situate in prossimità di strade ad elevata percorrenza (Mestre: Corso del Popolo, via Circonvallazione, via Da Verrazzano (solo fino al 16/07/2001); Marghera: via Fratelli Bandiera);
- 1 stazione di tipo D, per la rilevazione dell'inquinamento fotochimico e degli inquinanti secondari (Maerne di Martellago).

E' inoltre presente un'altra stazione preposta al controllo dell'inquinamento industriale (Malcontenta, definita di tipo I).

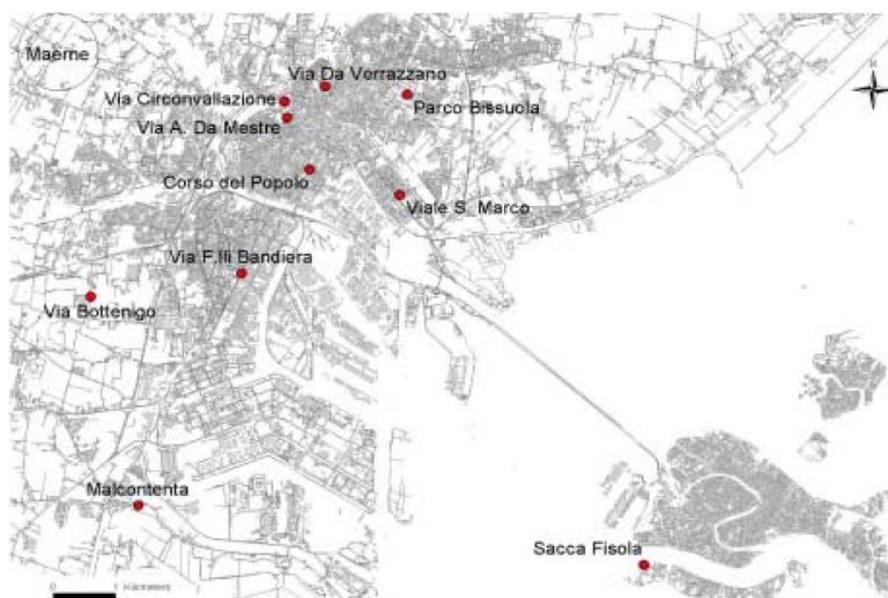


Figura 1: Localizzazione delle stazioni della rete ARPAV per il controllo dell'inquinamento atmosferico in Comune di Venezia.

2.2 Analisi dei dati

I dati utilizzati nelle analisi sviluppate per questa tesi provengono dalla rete di monitoraggio della qualità dell'aria della provincia di Venezia.

Il *dataset* completo contiene le osservazioni relative alle medie orarie delle concentrazioni di biossido di zolfo (SO₂) nell'aria, raccolte dalla stazione di rilevamento della rete ARPAV situata a Sacca Fisola, durante il periodo che va dal 1 gennaio 1999 al 31 dicembre 2002.

Le misurazioni dei parametri meteorologici quali la temperatura, l'umidità relativa e il vento, sono disponibili solo a partire dal 1 gennaio 2001.

anno	1999	2000	2001	2002
totale dati	8760	8784	8760	8760
mancanti	3832	412	1082	785
% dati mancanti	43,7	4,7	12,4	9,0
dati meteo	no	no	si	si

Tabella 2: Statistiche riassuntive.

Fin dal principio del processo di analisi è emerso il problema dell'elevato numero di valori mancanti nella matrice dei dati, dovuti a errori nella fase di rilevazione delle misurazioni.

Come si evince dalla tabella 2, per l'anno 1999 non ci sono le rilevazioni meteorologiche e il 43,7% dei dati manca.

Per tutte le analisi statistiche seguenti si considerano solo gli anni 2001 e 2002, per i quali la percentuale di dati mancanti è

relativamente contenuta e il *dataset* è provvisto anche dei rilievi meteorologici.

Le tecniche adoperate per stimare i valori mancanti sono la media e il metodo delle *splines*.

Con la media si deduce il valore mancante dai valori conosciuti, stimandolo con la media aritmetica semplice calcolata con il valore precedente e successivo.

Questo tipo di sostituzione è applicata solo quando il dato mancante è isolato; una volta individuata la sua posizione si inserisce il valore più probabile che è quindi la media.

Dati mancanti di questo genere sono presenti nel *dataset* quasi ogni 24 ore e sono dovuti alla calibrazione automatica dello strumento di rilevazione, che viene effettuata periodicamente in un'ora notturna quando tipicamente le concentrazioni di inquinante disciolto nell'aria sono minori.

Il metodo delle *splines* (Wahba, 1990) è stato adottato per coprire più ampi tratti di dati mancanti con le loro previsioni.

In analisi numerica una *spline* è uno speciale tipo di curva composta a tratti da polinomi, selezionati per formare un'unica funzione che disegni, attraverso i punti dati, una curva liscia.

Per ottenere questa funzione liscia, con derivata prima poco variabile, si deve minimizzare la quantità:

$$\min_f \sum_{i=1}^n (y(t_i) - f(t_i))^2 + \lambda \int_a^b (f''(t))^2 dt$$

Specificando che:

- $y(t_i)$ è la serie temporale dei dati;
- il primo termine $\sum_{i=1}^n (y(t_i) - f(t_i))^2$ rappresenta l'errore di previsione;
- λ è un valore non negativo e la funzione risulta maggiormente liscia al crescere di λ . Casi estremi sono:
 - $\lambda = 0$ la funzione f interpola i dati;
 - $\lambda = \infty$ la f è una retta;
- gli estremi dell'integrale sono $a = t_1 < \dots < t_n = b$;
- $(f''(t))^2$ è la misura della variabilità della funzione tra a e b .

Applicando questo metodo ad alcuni tratti delle serie temporali in esame, che includevano dati mancanti, si è potuto ricostruire l'andamento della serie anche in questi intervalli e quindi prevedere i valori.

2.3 La variabile SO₂

Per la rilevazione delle concentrazioni di biossido di zolfo gli apparecchi per la misurazione fanno riferimento all'ora solare.

La media oraria delle ore 01:00 rappresenta la media di tutti i valori elementari, ritenuti validi da sistema, misurati dalle ore 00:01 alle ore 01:00 per un determinato parametro. La media oraria delle ore 24:00 rappresenta la media di tutti i valori elementari, ritenuti validi da sistema, misurati dalle ore 23:01 alle ore 24:00 per un determinato parametro.

In generale gli stessi criteri sono applicati anche nei processi di rilevamento dei dati meteorologici.

L'unità di misura per le concentrazioni medie orarie di SO₂ è $\mu\text{g}/\text{m}^3$.

La serie storica delle concentrazioni medie orarie del biossido di zolfo, rappresentata per i due anni considerati nei grafici sottostanti (figure 2 e 3), non presenta un trend facilmente riconoscibile.

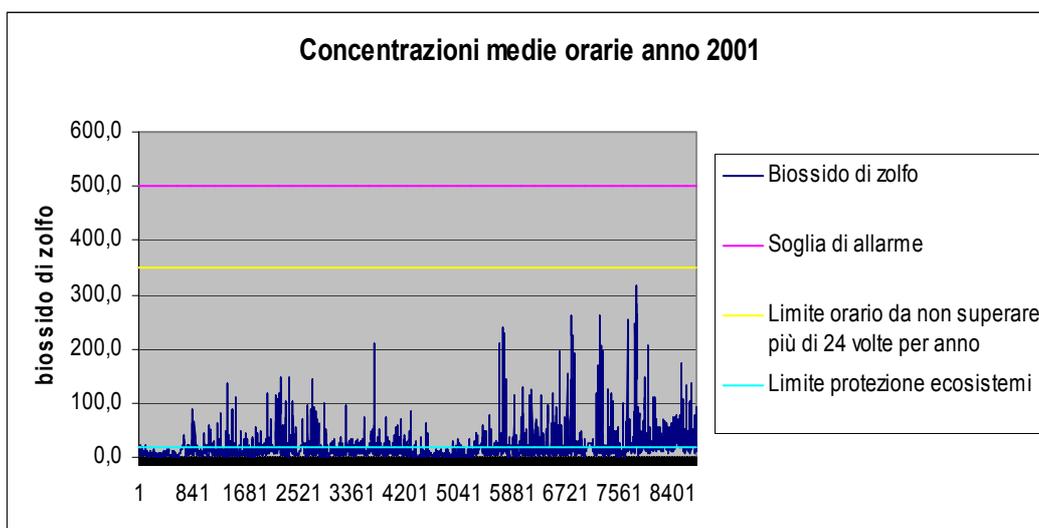


Figura 2: Concentrazioni medie orarie di SO₂ per l'anno 2001.

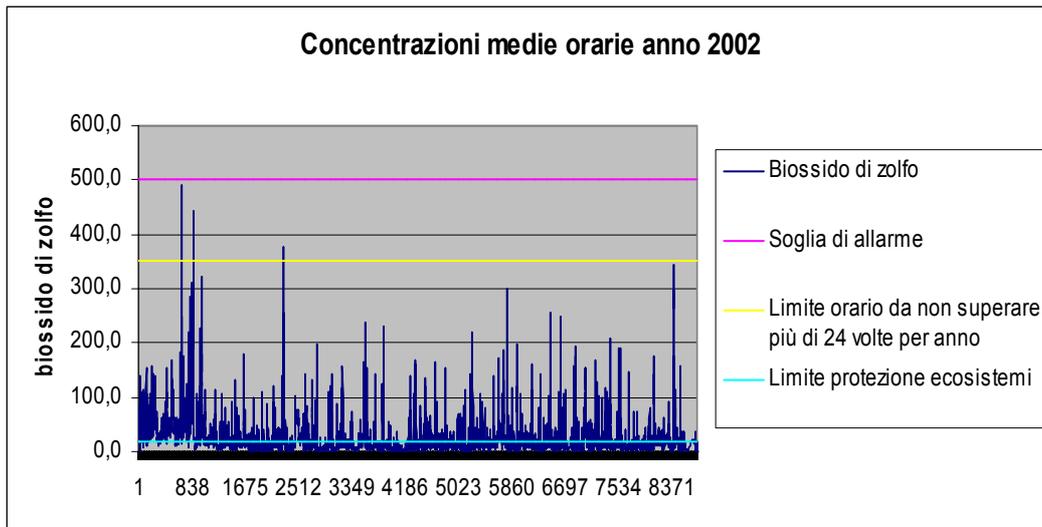


Figura 3: Concentrazioni medie orarie di SO₂ per l'anno 2002.

Nelle figure 2 e 3 sono anche evidenziati i limiti per la qualità dell'aria. Come si può notare, la soglia per la protezione degli ecosistemi ($20 \mu\text{g}/\text{m}^3$) è praticamente sempre superata, il limite orario ($350 \mu\text{g}/\text{m}^3$) viene oltrepassato solo quattro volte nel 2002, mentre la soglia di allarme ($500 \mu\text{g}/\text{m}^3$) è quasi raggiunta, ma mai oltrepassata, solo una volta nel mese di gennaio del 2002.

La tabella 3 riassume alcune statistiche descrittive. La percentuale di valori mancanti è di circa il 10 % per entrambi gli anni, mentre i valori della deviazione standard sono abbastanza elevati, i valori massimi raggiungono quasi il limite orario da non superare più di 24 volte l'anno ($350 \mu\text{g}/\text{m}^3$) e la soglia di allarme ($500 \mu\text{g}/\text{m}^3$).

Le medie mensili (figura 4) calcolate per i due anni consecutivi sembrano identificare un aumento nelle concentrazioni medie di SO₂ nei mesi invernali tra ottobre 2001 e febbraio 2002.

anno	2001	2002
media	17,5	22,9
deviazione standard	23,1	30,5
% dati mancanti	12,4	9,0
minimo	0	0
massimo	317,7	490,5
mediana	11,5	14,0
98° percentile	86,4	115,8

Tabella 3: Statistiche descrittive della serie del biossido di zolfo.

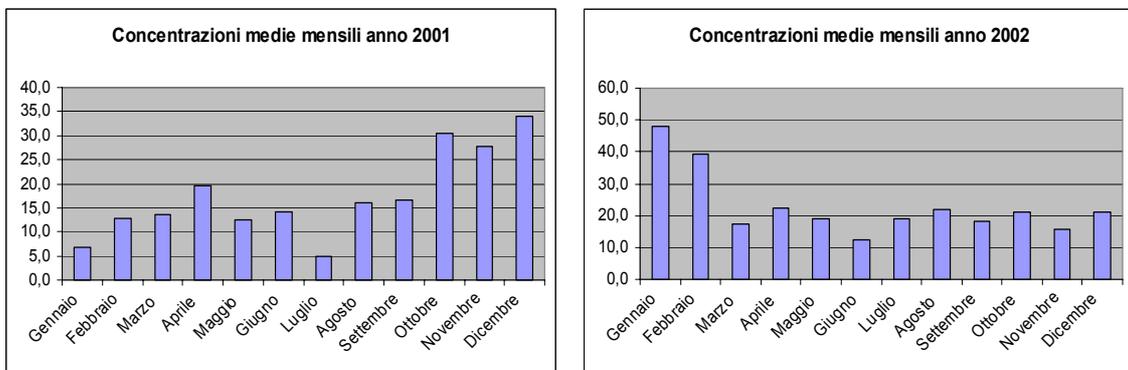


Figura 4: Concentrazioni medie mensili di SO₂ per il 2001 e 2002.

Analizzando giornalmente le concentrazioni di SO₂ si nota che i valori medi più elevati si presentano nelle ore centrali della giornata, cfr. figura 5.

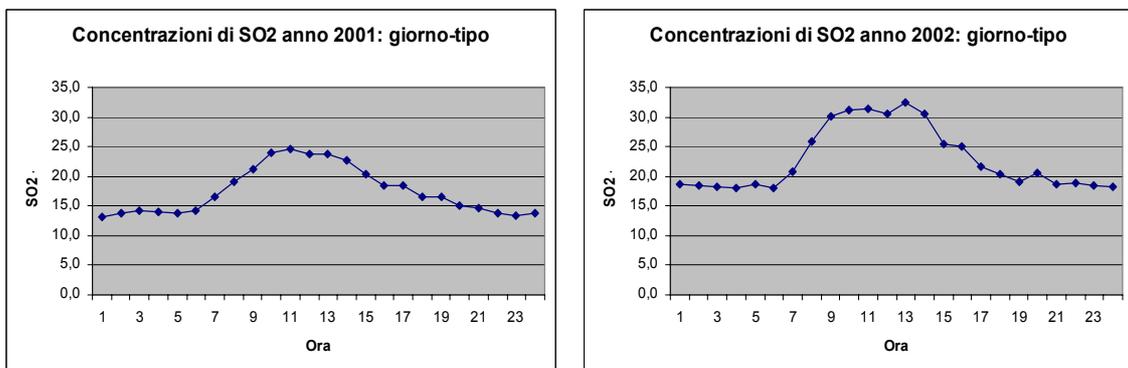


Figura 5: giorno-tipo per le concentrazioni di SO₂ per il 2001 e 2002.

2.4 Le variabili meteorologiche

È necessario sottolineare la massima importanza delle condizioni meteorologiche per la comprensione della nascita, della gravità e dello sviluppo nel tempo di un fenomeno di inquinamento atmosferico.

Su scala locale il fattore che più influenza il trasporto e la diffusione degli inquinanti è l'intensità del vento; inoltre un ruolo notevole è svolto dalle precipitazioni atmosferiche che contribuiscono a dilavare l'aria dalle contaminazioni.

In genere le concentrazioni dei contaminanti dell'aria sono minori quando il vento è moderato e l'atmosfera è instabile nei bassi strati. Al contrario, le concentrazioni degli inquinanti sono elevate in presenza di nebbia persistente oppure in assenza di vento o quando si è in presenza di inversioni termiche.

Il fenomeno dell'inversione termica crea degli strati d'aria più calda a qualche decina o centinaia di metri d'altezza per cui lo strato sottostante non sale e ristagna al suolo, il tutto comporta inevitabili processi di accumulo delle sostanze inquinanti. Le inversioni termiche si formano solitamente nelle notti limpide subito dopo il tramonto, a causa del rapido raffreddamento del suolo, che a sua volta provoca il raffreddamento dell'aria con cui è a contatto. Se questo fenomeno non termina col riscaldamento mattutino della superficie terrestre, gli inquinanti si possono accumulare anche per più giorni consecutivi, con tutti i problemi che ciò comporta.

Per un approfondimento sul tema dell'inversione termica si vedano Baird (2001) e Bruzzi (1994).

Le variabili meteorologiche a disposizione per lo studio sono: temperatura, umidità relativa e vento.

Si è suddiviso l'anno meteorologico in semestre "caldo", comprendente i mesi da aprile a settembre, e semestre "freddo", comprendente i mesi da gennaio a marzo e da ottobre a dicembre.

La descrizione dell'ambiente atmosferico su base stagionale si presenta come una rappresentazione più vicina ai fenomeni naturali, rispetto agli andamenti annuali, e favorisce i confronti sui comportamenti degli inquinanti che risentono delle variazioni stagionali.

2.4.1 Temperatura

La temperatura dell'aria viene espressa in gradi centigradi, °C, e secondo le norme dell'Organizzazione Meteorologica Mondiale (World Meteorological Organization) il sensore di rilevamento si deve collocare tra m 1,25 e m 2 dal suolo (WMO,1983).

Si riportano di seguito alcune delle statistiche descrittive di maggior interesse svolte sui dati, tabella 4.

anno	2001	2002
media	14,8	14,9
deviazione standard	8,4	8
% dati mancanti	2,4	1,2
minimo	-3,9	-3,8
massimo	34,6	34,4

Tabella 4: Statistiche descrittive della temperatura.

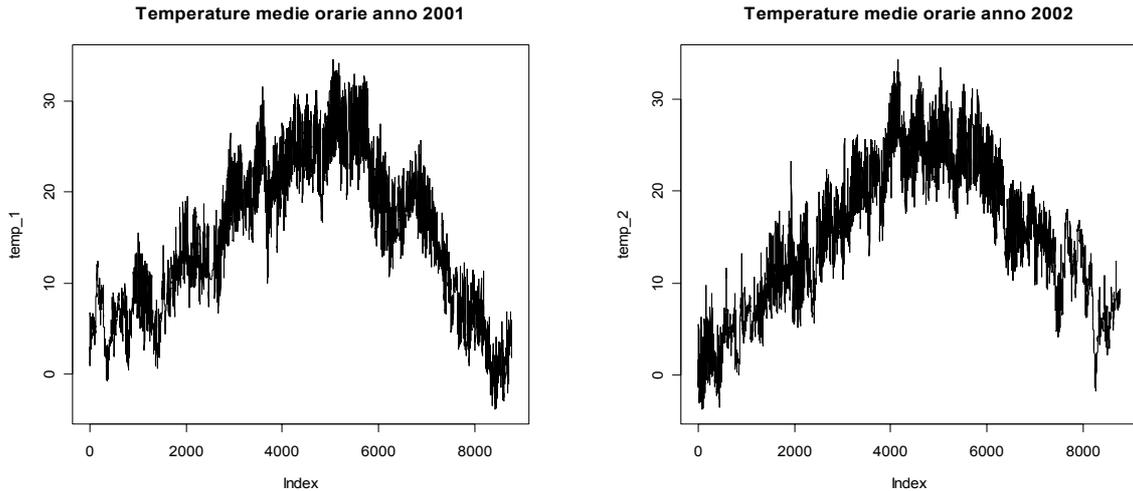


Figura 6: serie delle temperature medie orarie per il 2001 e 2002.

L'andamento stagionale delle temperature è facilmente intuibile. Dalla figura 6 si nota che l'innalzamento della temperatura corrisponde ai mesi estivi, ciò è visibile anche dallo sviluppo delle medie mensili (figura 7).

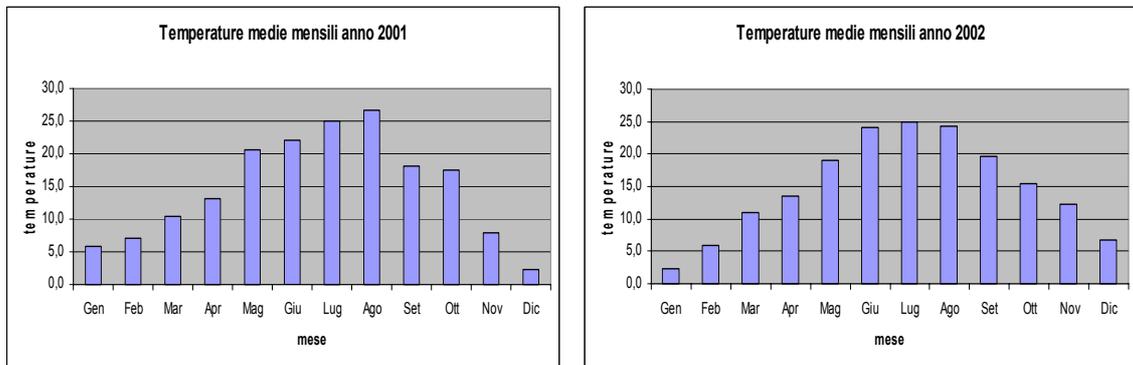


Figura 7: Temperature medie mensili del 2001 e 2002.

L'andamento della temperatura dell'aria per il "giorno-tipo" risulta quasi completamente sovrapponibile nei due periodi, salvo per l'aumento del valore assoluto nel semestre caldo. Il giorno tipo

presenta un trend in crescita in corrispondenza delle ore centrali della giornata, a causa dell'insolazione diurna, risulta quindi leggermente anticipato e prolungato nella fase estiva, cfr. figura 8 e 9.

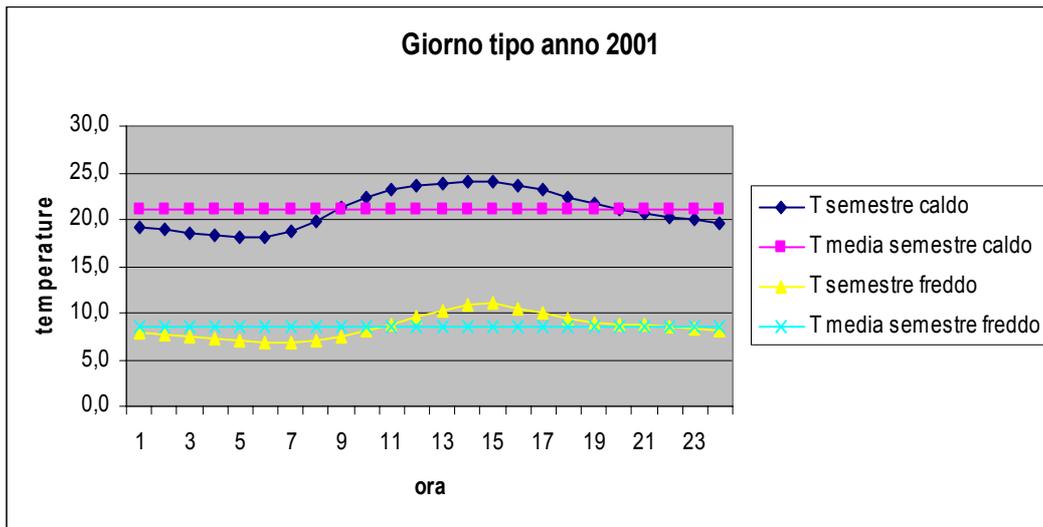


Figura 8: Giorno-tipo delle temperature del 2001.

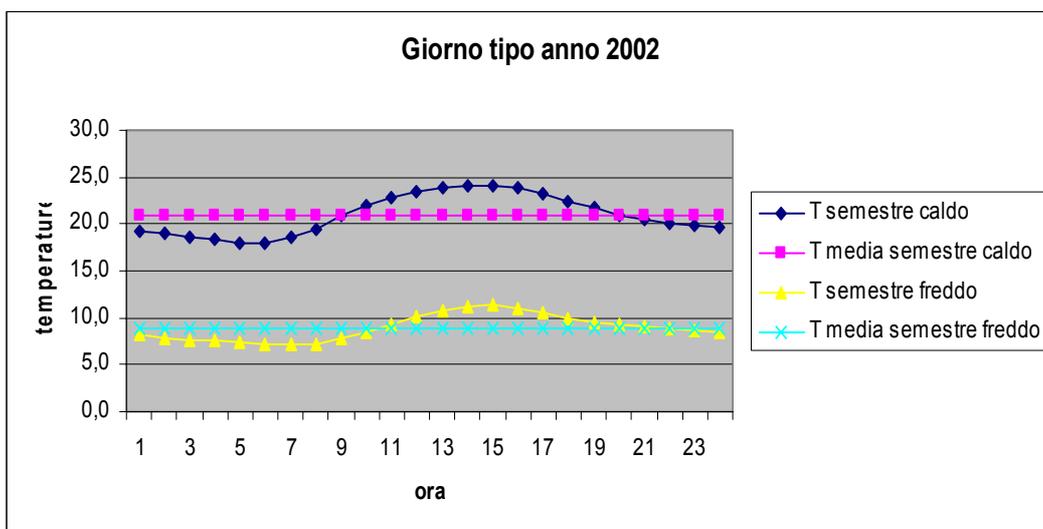


Figura 9: Giorno-tipo delle temperature del 2002.

2.4.2 Umidità relativa

L'umidità relativa si definisce come rapporto espresso in percentuale tra la quantità di vapor acqueo contenuto nell'atmosfera e la quantità massima di vapore che, alla stessa temperatura, potrebbe esservi contenuto.

Anche per questo parametro meteorologico sono riassunte alcune statistiche descrittive nella tabella 5.

anno	2001	2002
media	77,6	80,6
deviazione standard	15,2	15,7
% dati mancanti	2,4	1,2
minimo	30,8	21,4
massimo	98,0	98,7

Tabella 5: Statistiche descrittive per l'umidità relativa.

Le medie mensili riprodotte graficamente denotano una certa stabilità dell'umidità relativa nel corso dell'anno, cfr. figura 10.

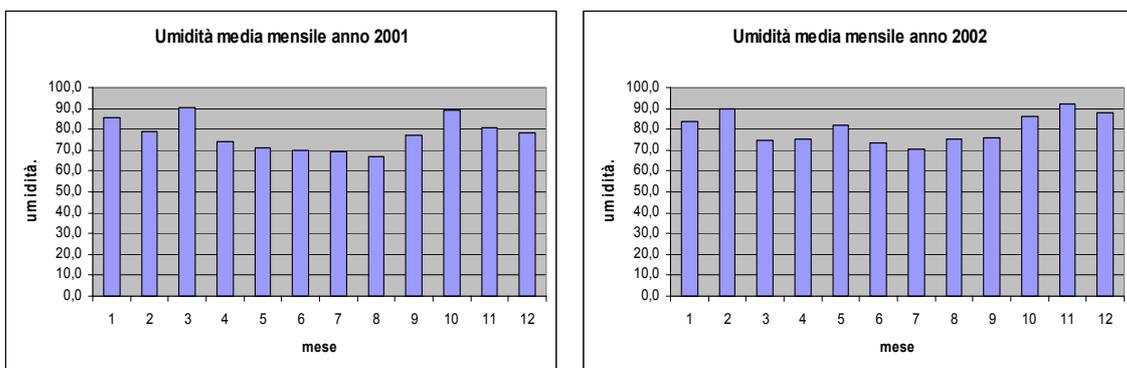


Figura 10: Umidità medie mensili per gli anni 2001 e 2002.

2.4.3 Vento

La velocità e direzione del vento sono misurate rispettivamente in metri/secondo e in gradi sessagesimali rispetto al nord. La quota standard per le misure al suolo è m 10 (WMO, 1983).

Per il calcolo della direzione di provenienza del vento vengono utilizzati prevalentemente tre metodi:

1. la direzione prevalente, cioè il settore nel quale si verifica la massima frequenza dei venti durante l'intervallo di tempo dell'osservazione;
2. la media vettoriale della velocità del vento;
3. la media aritmetica della direzione del vento.

Condizioni di calma e di vento variabile vengono attribuite rispettivamente quando l'intensità del vento è molto bassa e quando la deviazione standard della direzione del vento risulta molto elevata.

Le distribuzioni congiunte del valor medio della direzione e della velocità del vento vengono sintetizzate nei caratteristici diagrammi della rosa dei venti.

Dall'interpretazione delle rose dei venti (figura 11), tratte dal Rapporto annuale 2001 dell'Arpav sulla qualità dell'aria nel comune di Venezia, si afferma che il semestre caldo presenta prevalentemente venti da NE (nord – est), frequenza 15,6%, e NNE (nord – nord – est), frequenza 11,4%, con una forte componente da SSE (sud – sud – est), frequenza 12,4%, e una percentuale superiore al 55% di velocità comprese tra i 2 e 4 m/s.

Anche nel semestre freddo l'intervallo di velocità prevalente è tra i 2 e 4 m/s, nel 46% dei casi, e permane come principale la componente NNE, 16,0%, assieme alla direzione NE, 14,2%.

Si nota che nel semestre freddo la componente del vento da SSE non è presente con la stessa frequenza riscontrata nel semestre caldo.

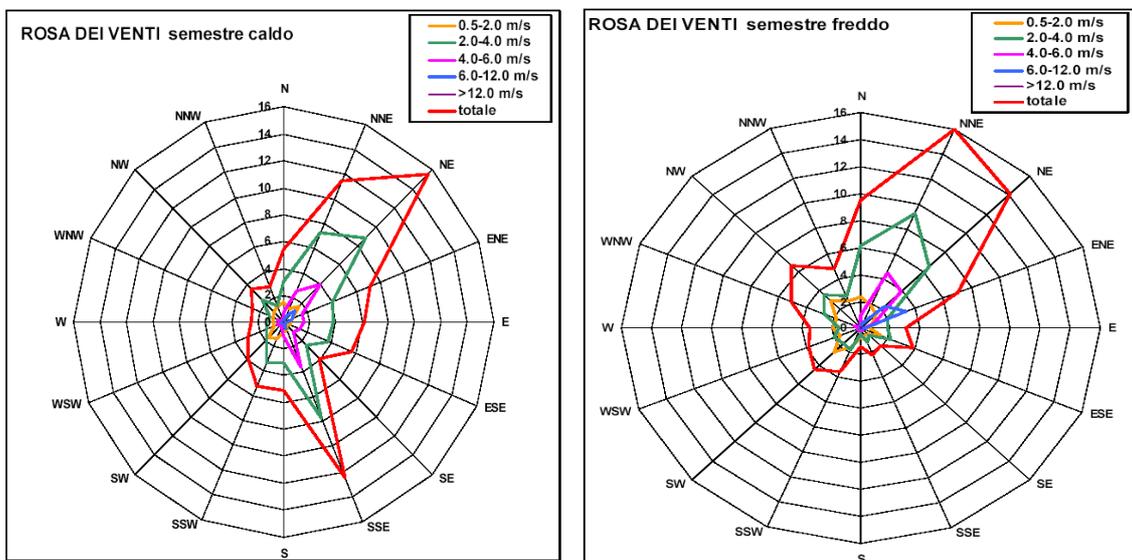


Figura 11: Rose dei venti dei semestri caldo e freddo dell'anno 2001.

La velocità del vento nella giornata tipo del semestre caldo è caratterizzata in generale da un incremento nelle ore centrali, durante il quale si verifica un maggiore grado di rimescolamento dell'atmosfera. Questo fenomeno non si osserva nei mesi invernali per i quali la velocità oscilla in modo relativamente contenuto attorno alla media, vedi figura 12.

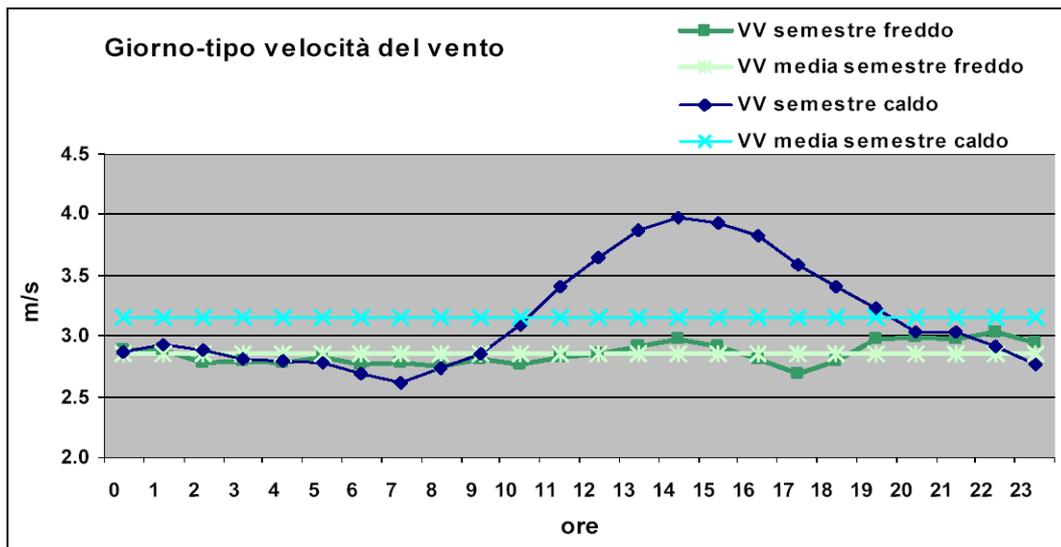


Figura 12: Giorno-tipo per il vento nell'anno 2001.

Al fine di esaminare se le concentrazioni di biossido di zolfo risentono dell'influenza del vento il problema è stato scomposto in due fasi.

Sono stati scartati dal *dataset* del vento tutti i dati mancanti poiché sarebbe stato impossibile prevederli con qualsiasi metodo a causa della complessità del tipo di rilevazioni, comprendenti non solo la velocità ma anche il settore di direzione prevalente del vento.

Si sono calcolate velocità e concentrazioni medie dell'inquinante per ciascuna direzione del vento prevalente, e si sono considerati i dati secondo classi ordinali di velocità del vento.

La calma di vento (velocità < 0,5 m/s) e la variabilità (condizione in cui non è possibile stabilire con precisione da che settore soffia il vento) hanno formato due classi separate, mentre per la scomposizione dei dati rimanenti si sono prese in considerazione le stesse divisioni valutate per il disegno della rosa dei venti, cioè classi di velocità da 0,5 a 2 m/s, da 2 a 4 m/s, da 4 a 6 m/s e da 6 a 12 m/s.

Mettendo a confronto i dati così suddivisi si è evidenziato quello che si avrebbe potuto intuire con facilità: quanto più la velocità del vento è elevata tanta meno quantità di inquinante è presente nell'aria.

Le velocità per le quali si osservano valori contenuti di SO_2 sono molto superiori alla media di 2 – 4 m/s, infatti, dopo i 6 m/s e particolarmente per l'anno 2001, si evince un notevole abbassamento nelle concentrazioni dell'inquinante, cfr. figure 13 e 14.

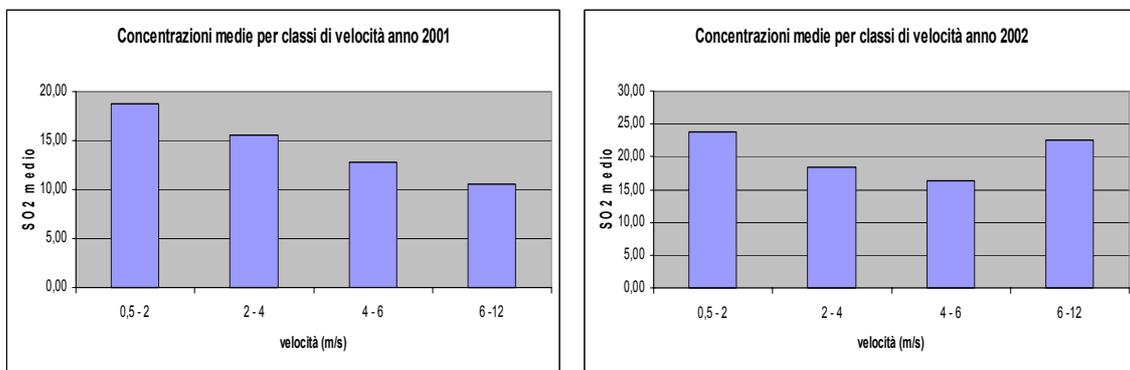


Figura 13: Concentrazioni medie di SO_2 per classi di velocità del vento, anni 2001 e 2002.

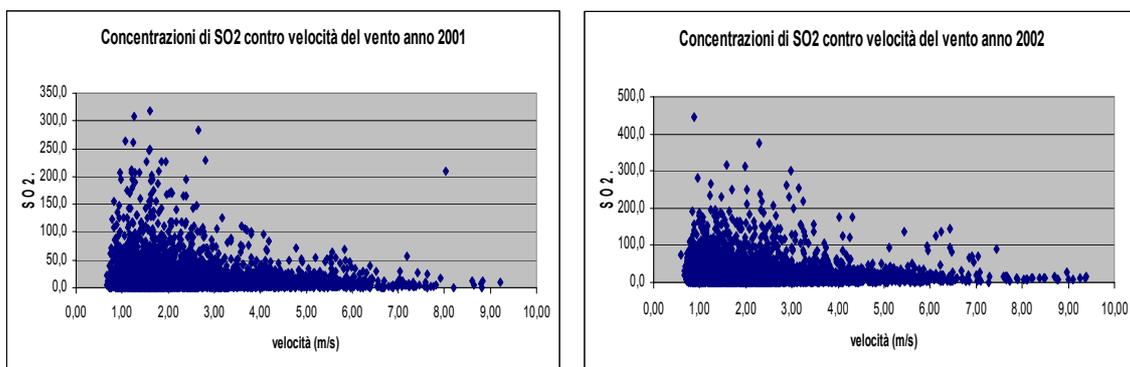


Figura 14: Concentrazioni di SO_2 al variare della velocità del vento, anni 2001 e 2002.

Si nota che dalla direzione NO (306° circa), indipendentemente dalla velocità del vento, le concentrazioni di inquinante si mantengono molto al di sopra della media, vedi figura 15.

Questo accade probabilmente per la vicinanza del sito di rilevamento Sacca Fisola, alla zona di Mestre - Marghera, situata proprio in direzione NO, ricca di industrie che bruciano oli combustibili, densamente popolata e perciò fonte molto probabile di inquinamento da SO₂.

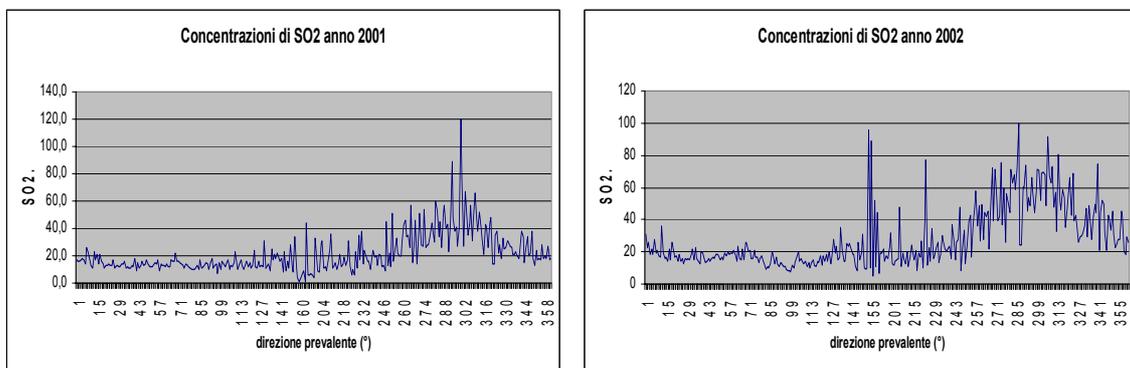


Figura 15: Concentrazioni di SO₂ al variare della direzione del vento, anni 2001 e 2002.

2.5 Correlazioni

Da queste prime analisi esplorative condotte sulla matrice dei dati a disposizione si calcolano le correlazioni che intercorrono tra l'inquinante SO₂, temperatura e umidità relativa, tabelle 6 – 7 e figura 16.

	SO2	temp	umr
SO2	1,00	-0,14	0,05
temp	-0,14	1,00	-0,43
umr	0,05	-0,43	1,00

Tabella 6: correlazioni per l'anno 2001.

	SO2	temp	umr
SO2	1,00	-0,20	0,04
temp	-0,20	1,00	-0,43
umr	0,04	-0,43	1,00

Tabella 7: correlazioni per l'anno 2002.

Nel corso dei due anni le correlazioni tra le variabili si mantengono quasi costanti, cioè da un anno all'altro non cambia l'influenza di temperatura e umidità sull'SO₂.

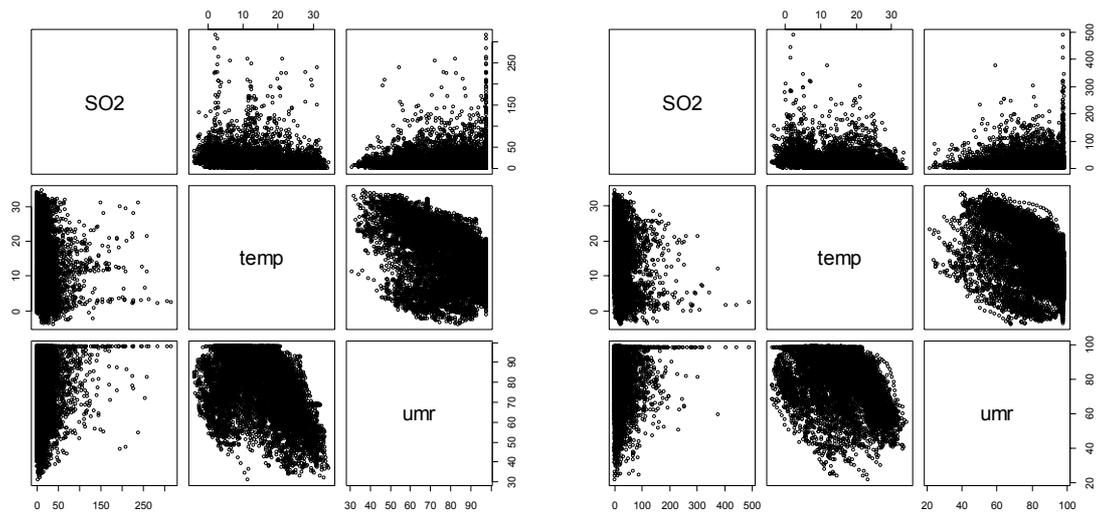


Figura 16: Grafici di dispersione per gli anni 2001 e 2002.

Non è stato possibile calcolare la correlazione tra biossido di zolfo e vento a causa dell'elevato numero di valori mancanti per il vento.

Capitolo 3

Identificazione dei punti di cambiamento in una serie temporale

3.1 Introduzione

Gli algoritmi per l'individuazione di salti di livello nello sviluppo delle serie temporali sono già ben conosciuti nelle analisi di sistemi dinamici controllati che descrivono processi industriali.

Un nuovo campo di interesse in cui queste metodologie possono essere utilizzate e sviluppate è il monitoraggio dei processi di inquinamento ambientale e, in particolare, di inquinamento dell'aria.

Problemi molto frequenti nell'analisi di dati sulla qualità dell'aria sono la mancanza e/o l'imprecisione dei dati stessi. Infatti questi possono portare in alcuni casi a scorrette interpretazioni dei risultati ottenuti nelle varie fasi delle analisi.

Al fine di identificare anomalie nei dati, con un ritardo ragionevole rispetto al momento in cui queste avvengono realmente, metodologie definite di tipo "on line", Shirayev (1963), Pollack (1985), Willsky e Jones (1976), Lai (1995), propongono di studiare le serie con algoritmi di controllo di processo e modellazioni di tipo stocastico.

Recentemente, allo sviluppo di queste tecniche applicate alla sfera ambientale hanno contribuito Bordignon, Scagliarini (2000), Nirel (2000), Capizzi (2001).

Questa tesi si occupa invece di studiare le serie temporali dal punto di vista “*off line*”, metodo che rappresenta un approccio in cui l'identificazione dei cambi di livello viene studiata a posteriori.

Quando ci si trova nella situazione in cui il momento del cambiamento è noto, si può ricorrere all'utilizzo della funzione di trasferimento opportunamente strutturata e ai principi della *intervention analysis*, affrontata per la prima volta da Box e Tiao (1975).

Se invece il momento del cambiamento non è noto, allora bisogna seguire la procedura per l'identificazione dei cambi di livello o degli *outliers* proposta da Chen e Liu (1993).

Un caso particolare del momento del cambiamento non noto è quando in più ci sono alcune variabili esplicative che possono essere inserite nel modello. Quindi, prima di risolvere il problema del riconoscimento del punto di cambiamento con il metodo di Chen e Liu, trattato nel paragrafo 3.6, è necessario determinare un modello che includa le variabili esplicative più adatte alla stima del fenomeno. Tutto ciò è risolto tramite l'utilizzo congiunto degli algoritmi genetici e di un algoritmo denominato *ISGN, Iterative Stepwise Gauss Newton*.

3.2 Modelli a funzione di trasferimento

Date k serie di *input* $X_i(t)$, $i=1, \dots, k$ e una serie di *output* $Y(t)$ la relazione che lega le due serie è rappresentata da:

$$Y(t) = \sum_{i=1}^k v_i(B) X_i(t) + n(t)$$

Dove:

- $v_i(B) = \sum_{j=-\infty}^{+\infty} v_{ij} B^j$ è la funzione di trasferimento che trasforma gli *input* in *output*, rappresentabile anche nella forma razionale $v_i(B) = \frac{\omega_i(B)}{\delta_i(B)}$;
- B è l'operatore *backshift* tale che $BY(t) = Y(t-1)$;
- $n(t)$ è la serie di rumore additivo indipendente dagli *input*.

Un esempio molto semplice di modello a funzione di trasferimento è il modello stazionario ARMA(p,q):

$$\begin{aligned} \phi(B)Y_t &= \vartheta(B)Z_t \\ Y_t &= \frac{\vartheta(B)}{\phi(B)}Z_t = \psi(B)Z_t \end{aligned}$$

che rappresenta un sistema dinamico in cui l'input è un processo *white noise* e la funzione di trasferimento è $\psi(B) = \frac{\vartheta(B)}{\phi(B)}$.

In relazione alle varie forme che può assumere la serie di rumore $n(t)$ Box e Tiao (1975) stimano un unico modello ARMA con interventi che considera insieme cambiamenti strutturali e *outliers* (cfr. paragrafo successivo):

$$Y(t) = \sum_{i=1}^k \frac{\omega_i(B)}{\delta_i(B)} I_{t_i}(t) + \frac{\theta(B)}{\phi(B)} e(t)$$

La variabile indicatrice $I_{t_i}(t)$ evidenzia il verificarsi, o meno, di un intervento al tempo t_i .

Un *outlier* al tempo t_i è modellato fissando $\frac{\omega_i(B)}{\delta_i(B)} = \omega_i$, mentre un salto

di livello al tempo t_i è modellato fissando $\frac{\omega_i(B)}{\delta_i(B)} = \frac{\omega_i}{1-B}$.

3.3 Intervention analysis

Spesso capita di dover valutare l'effetto di un intervento esterno sull'andamento di una serie storica, dove per intervento si intende la comparsa di un qualsiasi fenomeno che ad un dato periodo T , o da un dato periodo T e poi per un certo tempo, influenza la serie stessa.

Esempi di simili interventi possono essere il passaggio dal carbone all'olio combustibile nell'Inghilterra del 1965 e il conseguente cambiamento riscontrato nei livelli di inquinamento della regione (Bennet, 1976), oppure la costruzione di una nuova autostrada e le variazioni nella composizione della benzina a Los Angeles che hanno portato ad alterazioni nei livelli di ozono (Box e Tiao, 1975).

Ulteriori esempi sono l'effetto di una campagna pubblicitaria sulla serie temporale delle vendite di un certo prodotto o le conseguenze dell'obbligatorietà delle cinture di sicurezza sulla serie storica delle morti per incidente stradale (Bhattacharyya e Layton, 1979; Abraham, 1987); ecc.

Dal punto di vista metodologico, ciò che differenzia l'analisi d'intervento dalle procedure più generali di identificazione e stima di cambiamenti di livello e di *outliers* è la conoscenza a priori della posizione e della natura delle anomalie che appaiono sulla serie.

Si supponga che al tempo t_i ($t_i < n$) avvenga un intervento, e che quindi la serie $z(t)$ (generata dal processo stocastico stazionario $Z(t)$ e con valore atteso nullo) sia rappresentabile attraverso un modello ARMA solamente per $t \in [1, (T-1)]$ ma non per il resto della serie.

Box e Tiao (1975), nel lavoro anticipatore della *intervention analysis*, proposero il seguente modello a k interventi:

$$Y(t) = \sum_{i=1}^k v_i(t) + Z(t) = \sum_{i=1}^k \frac{\omega_i(B)}{\delta_i(B)} I_{t_i}(t) + \frac{\theta(B)}{\phi(B)} e(t),$$

dove:

- $\omega(B)$ e $\delta(B)$ sono i polinomi in B , operatore di ritardo, di grado rispettivamente s e r :

$$\begin{aligned} \omega(B) &= \omega_0 - \omega_1 B - \omega_2 B^2 - \dots - \omega_s B^s, \\ \delta(B) &= 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r, \end{aligned}$$

- $v(t)$ è la funzione di trasferimento, che trasferisce gli effetti degli *input* sull'*output* $Y(t)$, ed è data dalla soluzione dell'equazione:

$$\delta(B)v(t) = \omega(B)I_{t_i}(t).$$

- $I_{t_i}(t)$ è una variabile indicatrice, detta "variabile d'intervento", che assume solo valore 0 e 1 , e che configura l'intervento attraverso una variabile "scalino" che assume valore 1 dall'istante t_i dell'intervento in poi e valore 0 altrove:

$$I_{t_i}(t) = S_{t_i}(t) = \begin{cases} 0, & t < t_i; \\ 1, & t \geq t_i \end{cases};$$

oppure attraverso una variabile "impulso" che assume valore 1 nell'istante dell'intervento e valore 0 altrove:

$$I_{t_i}(t) = P_{t_i}(t) = \begin{cases} 0, & t \neq t_i \\ 1, & t = t_i \end{cases}.$$

Un modello d'intervento può essere rappresentato indifferentemente per mezzo dell'una o dell'altra forma, perché tra le due variabili sussiste la seguente relazione:

$$(1 - B)S_{t_i}(t) = P_{t_i}(t).$$

Nel caso in cui, per esempio, l'intervento provochi un salto di livello di ampiezza ω , ritardato di k periodi, la funzione di trasferimento, con funzione scalino o impulso, adatta è:

$$\begin{aligned} v_i(t) &= \omega B^k S(t), \\ v_i(t) &= \frac{\omega B^s}{1 - B} P(t). \end{aligned}$$

Vari altri tipi di funzione di trasferimento descrivono la molteplicità di interventi che possono modificare la serie con nuovi andamenti.

Alcuni esempi tra i più noti sono: quando il cambio di livello avviene gradualmente a partire dal periodo $t+1$; quando si desidera rappresentare una situazione in cui l'intervento provoca un balzo verso l'alto della serie che rientra poi gradualmente sino a tornare al livello di partenza; quando la serie, dopo il salto iniziale, ritorna gradualmente verso un livello più alto di quello pre-intervento; o nel caso in cui la serie subisce un incremento graduale a differenze decrescenti in $[t_1, t_2]$ per poi tornare asintoticamente verso il livello di partenza $t_2 + 1$.

3.4 Iterative Stepwise Gauss Newton algorithm (ISGN)

Quando l'istante in cui ha luogo il cambio di livello non è noto e si ha a disposizione più di una variabile esplicativa, il problema risulta composto di due parti: bisogna infatti scegliere le variabili esplicative da includere nel modello di stima e quindi identificare il momento e il tipo di cambiamento avvenuto.

La fase iniziale, la stima dei modelli possibili, può essere condotta impiegando insieme un *Iterative Stepwise Gauss Newton algorithm (ISGN)* e un algoritmo genetico, come in Chiogna, Gaetan, Masarotto (2003).

L'obiettivo è ottenere un modello a funzione di trasferimento M della forma:

$$Y(t) = \sum_{i=1}^k \frac{\omega_i(B)\Omega_i(B)}{\delta_i(B)\Delta_i(B)} X_i(t) + \frac{\mathcal{G}(B)\Theta(B^g)}{(1-B)^d(1-B^g)^D\phi(B)\Phi(B^g)} e(t)$$

Dove:

- g rappresenta la stagionalità della serie;
- $e(t)$ è il processo *white noise* $WN(0, \sigma^2)$ indipendente dagli *input*.

Inizialmente si fissano:

- $r_i^*, R_i^*, s_i^*, S_i^*, q^*, Q^*, p^*, P^*, i = 1, \dots, k$ i gradi massimi dei polinomi $\omega_i(B), \Omega_i(B), \delta_i(B), \Delta_i(B), \mathcal{G}(B), \Theta(B), \phi(B), \Phi(B)$;
- $m^* = p^* + q^* + \sum_{i=1}^k (r_i^* + s_i^* + R_i^* + S_i^*)$ il numero di geni in un cromosoma, dove ogni gene è un parametro del modello M stimato.

Il vettore dei parametri ignoti è ottenuto minimizzando la quantità:

$$s(\alpha) = \frac{\sum_{t=\tilde{t}+1}^T e(t; \alpha)^2}{\tilde{T}},$$

dove:

- $e(t; \alpha) = Y(t) - \hat{Y}(t; \alpha)$, e $\hat{Y}(t; \alpha)$ è la previsione a un passo avanti di $Y(t)$;
- $\tilde{T} = T - \tilde{t}$ con $\tilde{t} = p + d + Dg + \max_{i=1, \dots, I} \left\{ \sum_{j=1}^k (s_j + S_j), r_i + R_i + \sum_{j \neq i}^k (s_j + S_j) \right\}$.

Per minimizzare questa quantità si utilizza un algoritmo di tipo *Gauss – Newton*. Il passo $k+1$ di questo algoritmo è:

$$\hat{\alpha}^{(k+1)} = \hat{\alpha}^{(k)} - \lambda^{(k)} \left[\sum_{t=\tilde{t}+1}^T J(t; \hat{\alpha}^{(k)}) J(t; \hat{\alpha}^{(k)})^T \right]^{-1} \sum_{t=\tilde{t}+1}^T J(t; \hat{\alpha}^{(k)}) e(t; \hat{\alpha}^{(k)})$$

dove:

- $0 < \lambda^{(k)} \leq 1$;
- $J(t; \hat{\alpha}^{(k)})$ è il vettore jacobiano $\frac{\partial \hat{Y}(t; \alpha)}{\partial \alpha}$.

Si scelgono casualmente i geni della popolazione di partenza, secondo una distribuzione normale.

La discendenza $M^{(0)}$ di due genitori $M^{(1)}$, $M^{(2)}$ è generata secondo:

$$\theta_j^{(0)} = \frac{\theta_j^{(1)} + \theta_j^{(2)}}{2}, \quad j = 1, \dots, m^* .$$

Mentre la fase di mutazione provvede, poi, a cambiare casualmente alcuni geni della discendenza. Si veda il paragrafo 3.5.

L'algoritmo *ISGN* si applica alla discendenza: se il modello così identificato è diverso da quelle che sono le caratteristiche della popolazione di modelli attuale allora bisogna sostituirci gli individui, cioè le variabili esplicative, che minimizzano il criterio BIC (*Bayesian Information Criterion*):

$$BIC = \log s(\hat{\alpha}) + m \frac{\log \tilde{T}}{\tilde{T}},$$

con:

- α è il vettore di m parametri ignoti;

- $s(\alpha)$ è la quantità $\frac{\sum_{t=\tilde{T}+1}^T e(t; \alpha)^2}{\tilde{T}}$.

Altrimenti, si calcola un altro modello che è a sua volta confrontato nuovamente con la popolazione di modelli già presente.

3.5 Algoritmi genetici

Gli algoritmi genetici sono una classe particolare di algoritmi evolutivi che sfruttano le tecniche ispirate alla biologia evolutiva, come: mutazione, selezione naturale e ricombinazione.

Sono molto usati per cercare le soluzioni dei problemi di ottimizzazione e di combinazione.

Il punto di forza del loro utilizzo è che non è necessario specificare per quale tipo di problema di ottimizzazione si stanno usando e che quindi non dipendono dalle ipotesi fatte sulla funzione obiettivo.

L'esempio più semplice di algoritmo genetico rappresenta ogni individuo, o cromosoma, in una stringa di *bit*; ogni cromosoma è composto di unità più piccole, chiamate geni.

La capacità di adattamento degli individui costituisce la funzione obiettivo che deve essere ottimizzata.

I processi di selezione, riproduzione e mutazione, agendo iterativamente, creano nuovi individui per la popolazione.

La selezione degli individui con maggior capacità di adattamento organizza la popolazione che può procedere attraverso la fase di riproduzione, nella quale da due individui scelti a caso viene creata una discendenza che, a sua volta, è casualmente modificata.

Nella pratica si adopera un algoritmo genetico per trovare il modello che meglio si adatta al problema in esame, avendo a disposizione varie potenziali soluzioni. L'algoritmo calcola una combinazione di questi modelli con una probabilità di mutazione definita a priori.

Se per esempio, il modello iniziale G è dato da:

$$Y(t) = \frac{\omega(B)}{\delta(B)} X(t) + e(t),$$

con:

$$\begin{aligned} \omega(B) &= \omega_0 - \omega_1 B - \omega_2 B^2 \\ \delta(B) &= 1 - \delta_1 B \end{aligned} .$$

I pesi dei parametri sono tutti uguali a 1, e quindi devono essere tutti stimati:

$$\begin{array}{cccc} \omega_0 & \omega_1 & \omega_2 & \delta_1 \\ 1 & 1 & 1 & 1 \end{array}$$

Ma se alcuni hanno peso 0, come per esempio:

$$\begin{array}{cccc} \omega_0 & \omega_1 & \omega_2 & \delta_1 \\ 1 & 0 & 1 & 0 \end{array}$$

Allora si stimano solo ω_0 e ω_2 , e i polinomi $\omega(B)$ e $\delta(B)$ diventano:

$$\begin{aligned} \omega(B) &= \omega_0 - \omega_2 B^2 \\ \delta(B) &= 1 \end{aligned} .$$

Da due genitori G e G' che hanno pesi rispettivamente:

$$\begin{array}{cccc} 1 & 0 & 1 & 0 \\ \omega_0 & 0 & \omega_2 & 0 \end{array}$$

e

$$\begin{array}{cccc} 0 & 0 & 1 & 1 \\ 0 & 0 & \omega_2' & \delta_1' \end{array}$$

La loro discendenza ha nei suoi geni i valori:

$$1 \quad 0 \quad 1 \quad 1$$

e, al posto di $\frac{\omega(B)}{\delta(B)}$, presenta la ricombinazione:

$$\frac{\omega_0 - \frac{\omega_2 + \omega_2'}{2} B^2}{1 - \delta_1' B} .$$

L'effetto di una mutazione può manifestarsi quando uno o più geni della discendenza cambiano valore, per esempio:

$$1 \quad 0 \quad 1 \quad 1$$

muta in:

$$1 \quad 0 \quad 0 \quad 1$$

e quindi il modello che rappresenta la discendenza diventa:

$$Y(t) = \frac{\omega_0}{1 - \delta_1' B} X(t) + e(t) .$$

3.6 Procedura di Chen e Liu

Stimato il modello che si adatta meglio al problema in esame, si deve procedere con il metodo di Chen e Liu (1993) per l'identificazione dei cambi di livello o degli *outliers* quando il momento del cambiamento è ignoto.

Questo metodo prende origine dall'opera di Box e Tiao (1975), che considera insieme cambiamenti strutturali e *outliers* in un unico modello ARMA con interventi, già trattato nel paragrafo 3.3:

$$Y(t) = \sum_{i=1}^k \frac{\omega_i(B)}{\delta_i(B)} I_{t_i}(t) + \frac{\mathcal{G}(B)}{\phi(B)} e(t).$$

Contrariamente a quest'ultimo caso della *intervention analysis*, ora il tipo di intervento, il numero di interventi e le loro posizioni sull'asse temporale non sono dati noti.

Le stime ai minimi quadrati di $\omega_{a,i}$, con $a = AO$ (*additive outlier*) sono:

$$\hat{\omega}_{AO,i} = \tau_{AO,i}^2 \tilde{\pi}(F) Y(t_i)$$

con varianza: $\text{var}(\hat{\omega}_{AO,i}) = \tau_{AO,i}^2 \sigma^2$,

dove:

$$\begin{aligned} \tau_{AO,i}^{-2} &= 1 + \pi_1^2 + \dots + \pi_{T-t_i}^2 \\ \tilde{\pi}(B) &= 1 - \pi_1 F - \dots - \pi_{T-t_0} F^{T-t_1} \\ F &= B^{-1} \end{aligned}$$

ed π_j è il coefficiente di B^j nel polinomio $\pi(B) = \frac{\phi(B)}{\mathcal{G}(B)} = 1 - \pi_1 B - \pi_2 B^2 - \dots$

Le stime ai minimi quadrati di $\omega_{a,i}$, con $a = LC$ (*level change*) sono:

$$\hat{\omega}_{LC,i} = \tau_{LC,i}^2 \tilde{\eta}(B) Y(t_i)$$

con varianza: $\text{var}(\hat{\omega}_{LC,i}) = \tau_{LC,i}^2 \sigma^2$,

dove:

$$\begin{aligned} \tau_{LC,i}^{-2} &= 1 + \eta_1^2 + \dots + \eta_{T-t_i}^2 \\ \tilde{\eta}(F) &= 1 - \eta_1 F - \dots - \eta_{T-t_0} F^{T-t_1} . \\ F &= B^{-1} \end{aligned}$$

ed η_j è il coefficiente di B^j nel polinomio $\eta(B) = \frac{\pi(B)}{(1-B)} = 1 - \pi_1 B - \pi_2 B - \dots$.

La statistica test che aiuta a rintracciare queste atipicità è data da:

$$\lambda_a(t_i) = \frac{|\hat{\omega}_i^a|}{\sqrt{\text{var}(\hat{\omega}_{a,i})}} .$$

Quando non è noto se il tipo di disturbo è un *outlier* o un cambio di livello, è più comodo definire la statistica test come:

$$\lambda_{t_i}^* = \max \{ \lambda_{AO}(t_i), \lambda_{LC}(t_i) \} .$$

Fissato a priori un certo valore c , che rappresenta una soglia oltre la quale i punti vengono classificati come punti di cambiamento della serie:

- se $\lambda_{t_i}^* = \lambda_{AO}(t_i) > c$ allora esiste un *outlier* al tempo t_i ;

- se $\lambda_{t_i}^* = \lambda_{LC}(t_i) > c$ allora il livello della serie ha subito un cambiamento al tempo t_i .

In funzione del tempo t_i ignoto si calcola $\lambda^* = \max_t \lambda_t^*$:

- se $\lambda^* = \lambda_{t_i}^* = \lambda_{AO}(t_i) > c$, allora al tempo t_i esiste un *outlier*;
- altrimenti se $\eta = \eta_{t_i} = \lambda_{LC}(t_i) > c$, allora al tempo t_i è intervenuto un cambiamento di livello.

Capitolo 4

Applicazione ai dati

4.1 Introduzione

Nel capitolo precedente, dedicato alla rassegna di alcuni metodi per la ricerca di punti di cambiamento nello studio “*off line*” delle serie temporali, sono state poste le basi formali per l’applicazione ai dati reali.

Per identificare queste anomalie si possono adottare due tipi di algoritmo: uno non guidato da un algoritmo genetico, oppure uno guidato da un algoritmo genetico.

Il *dataset*, già preso in esame per le statistiche descrittive nel capitolo 2, include le osservazioni delle concentrazioni di biossido di zolfo disciolto nell’aria e le rilevazioni di temperatura e umidità relativa, raccolte a Sacca Fisola, (Venezia) dal gennaio 2001 al dicembre 2002.

A causa del numero di osservazioni, due anni di dati orari per circa 17500 rilevazioni per ogni variabile, le serie sono state suddivise in tratti più brevi, corrispondenti a trimestri.

Nel seguito saranno presi in considerazione quattro trimestri campione: da luglio a settembre, e da ottobre a dicembre, sia per

l'anno 2001 sia per il 2002. Nell'appendice 1 i grafici relativi agli altri quattro trimestri.

Operativamente il lavoro si esegue con l'aiuto del programma *R*, versione 1.8.1, ed in particolare con alcune funzioni della libreria "*smiso*", di cui è autore il professor Guido Masarotto.

4.2 Stima di un modello a funzione di trasferimento

La scelta di un modello a funzione di trasferimento per le serie considerate, le concentrazioni di biossido di zolfo come *output* e temperatura e umidità relativa come variabili esplicative di *input*, è il primo passo per verificare se le condizioni meteorologiche influenzino la quantità di inquinamento da SO₂ presente nell'aria.

Nelle valutazioni iniziali dei modelli a funzione di trasferimento sono state incluse le variabili esplicative con un ritardo massimo possibile di 6, e la stagionalità, considerata di ventiquattro ore, con un ritardo massimo di 2.

Per il trimestre luglio – settembre 2001 il modello ottenuto è:

$$Y(t) = \frac{1 + 0,023B - 0,032B^4}{1 - 0,978B^{48}} U(t) + \frac{(1 - 0,586B^3 + 0,125B^6)(1 + 0,081B^{48})}{1 - 0,637B - 0,532B^3 + 0,306B^4} e(t)$$

dove:

- $U(t)$ è la variabile umidità relativa.

Il modello stimato risulta soddisfacente.

Infatti il test di Box-Ljung, il cui valore è 14.33, gdl: 17, p-value: 0.643, accetta l'ipotesi di incorrelazione dei residui. La funzione di autocorrelazione è dentro le bande di confidenza, che sono di tipo $\frac{2}{\sqrt{n}}$; il periodogramma è anch'esso all'interno delle bande di confidenza e non segnala comportamenti anomali nella sua distribuzione.

La figura 17 e la tabella 8 ne danno evidenza.

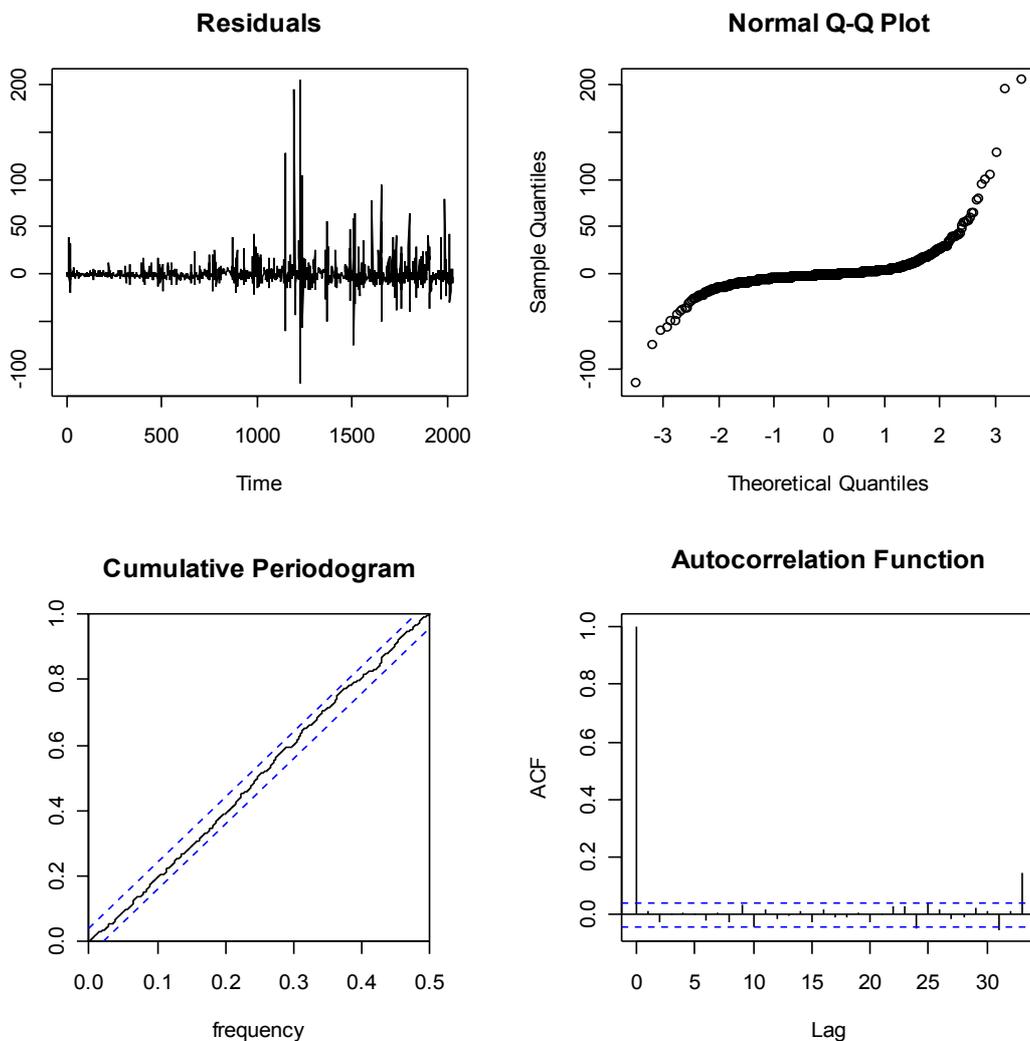


Figura 17: Grafici della distribuzione dei residui, quantile – quantile, periodogramma e ACF per il trimestre luglio – settembre 2001.

Coefficienti	Stima	Std.Error	z value	Pr(> z)
U.phi[1]	-0,023	0,006	-3,567	3,606e-04
U.phi[4]	0,032	0,008	3,835	1,254e-04
U.THETA[48]	0,978	0,017	56,896	0,000e+00
Noise.phi[3]	0,586	0,116	5,043	4,580e-07
Noise.phi[6]	-0,125	0,023	-5,416	6,108e-08
Noise.PHI[48]	-0,081	0,023	-3,609	3,074e-04
Noise.theta[1]	0,637	0,019	34,349	0,000e+00
Noise.theta[3]	0,532	0,116	4,599	4,249e-06
Noise.theta[4]	-0,306	0,082	-3,715	2,031e-04

Tabella 8: Stime dei coefficienti del modello a FT per il trimestre luglio – settembre 2001.

Per il trimestre ottobre – dicembre 2001 il modello stimato è:

$$Y(t) = (1 - 0,381B^4)U(t) + \frac{1 - 0,112B^2}{1 - 0,785B - 0,161B^4 + 0,088B^6} e(t)$$

Anche questo modello risulta soddisfacente per le stesse considerazioni fatte nel caso precedente, si vedano tabella 9 e figura 18.

Coefficienti	Stima	Std.Error	z value	Pr(> z)
U.phi[4]	0,381	0,031	12,134	0,000e+00
Noise.phi[2]	0,120	0,026	4,594	4,349e-06
Noise.theta[1]	0,785	0,018	42,699	0,000e+00
Noise.theta[4]	0,160	0,019	8,326	0,000e+00
Noise.theta[6]	-0,088	0,017	-5,084	3,694e-07

Tabella 9: Stime dei coefficienti del modello a FT per il trimestre ottobre – dicembre 2001.

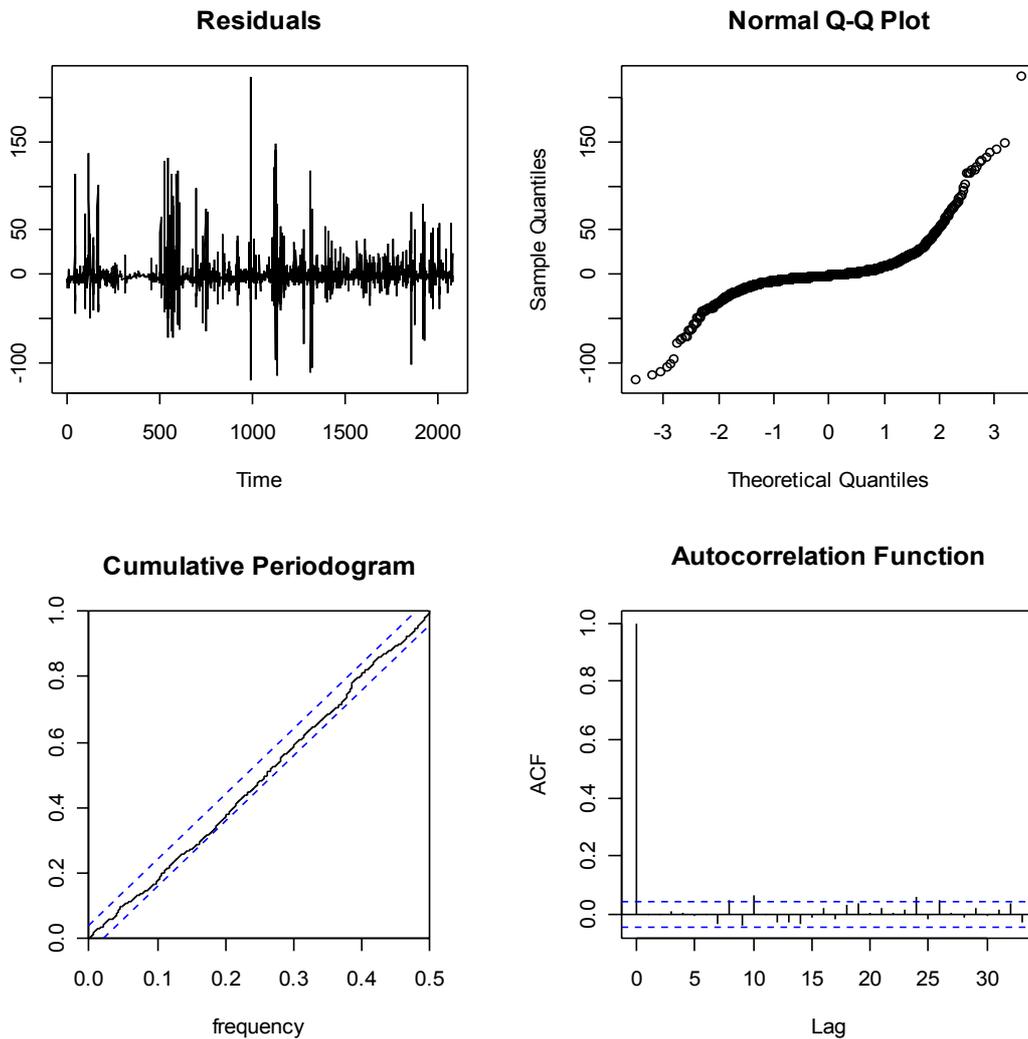


Figura 18: Grafici della distribuzione dei residui, quantile – quantile, periodogramma e ACF per il trimestre ottobre – dicembre 2001.

Si considerano ora i due trimestri campione per l'anno 2002.

Per il trimestre luglio – settembre il modello ottenuto è:

$$Y(t) = (1 - 1.583B + 1.472B^5)T(t) + (1 - 0.233B)U(t) + \left(\frac{1 + 0.232B}{1 - 0.57B}\right)e(t)$$

dove:

- $T(t)$ è la variabile temperatura;

- $U(t)$ è la variabile umidità relativa.

I risultati ottenuti sono soddisfacenti. Con il test di Box-Ljung, valore della statistica test 13.05, gdl: 17, p-value: 0.733, si accetta l'ipotesi di incorrelazione dei residui. La funzione di autocorrelazione è completamente entro le bande di confidenza; il periodogramma è anch'esso all'interno delle bande di confidenza e non segnala comportamenti anomali nella sua distribuzione. Cfr. figura 19 e tabella 10.

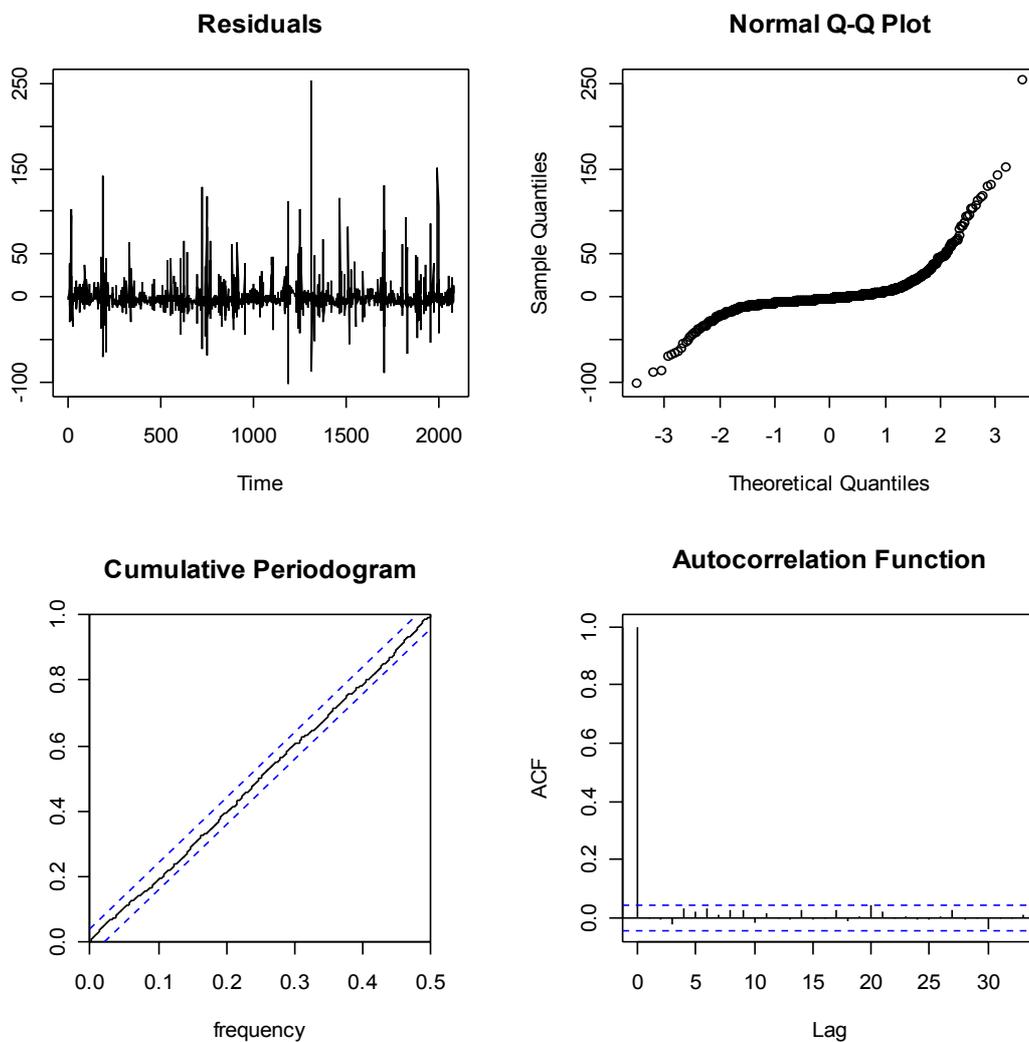


Figura 19: Grafici della distribuzione dei residui, quantile – quantile, periodogramma e ACF per il trimestre luglio – settembre 2002.

Coefficienti	Stima	Std.Error	z value	Pr(> z)
T.phi[1]	1,583	0,269	5,885	3,99E-09
T.phi[5]	-1,472	0,303	-4,846	1,26E-06
U.phi[1]	0,233	0,042	5,581	2,39E-08
Noise.phi[1]	-0,232	0,030	-7,697	1,40E-14
Noise.theta[1]	0,570	0,026	22,229	0,00E+00

Tabella 10: Stime dei coefficienti del modello a FT per il trimestre luglio – settembre 2002.

Per il trimestre ottobre – dicembre 2002 invece, il modello è risultato:

$$Y(t) = 21.998 + (1 - 3.858B + 2.328B^2 + 1.813B^4)T(t) + \left(\frac{1 - 0.102B^6}{1 - 0.71B - 0.105B^3} \right) e(t)$$

I grafici della figura 20 dimostrano la bontà del modello scelto.

Per questo modello il valore ottenuto con la statistica test di Box-Ljung è 12,18, gdl:17, p-value: 0,789.

Coefficienti	Stima	Std.Error	z value	Pr(> z)
Intercetta	21,997	4,42	4,977	6,45e-07
T.phi[1]	3,858	0,734	5,259	1,450e-07
T.phi[2]	-2,328	0,835	-2,789	5,290e-03
T.phi[4]	-1,813	0,572	-3,170	1,523e-03
Noise.phi[6]	0,102	0,023	4,454	8,422e-06
Noise.theta[1]	0,710	0,018	39,722	0,000e+00
Noise.theta[3]	0,105	0,018	5,738	9,591e-09

Tabella 11: Stime dei coefficienti del modello a FT per il trimestre ottobre – dicembre 2002.

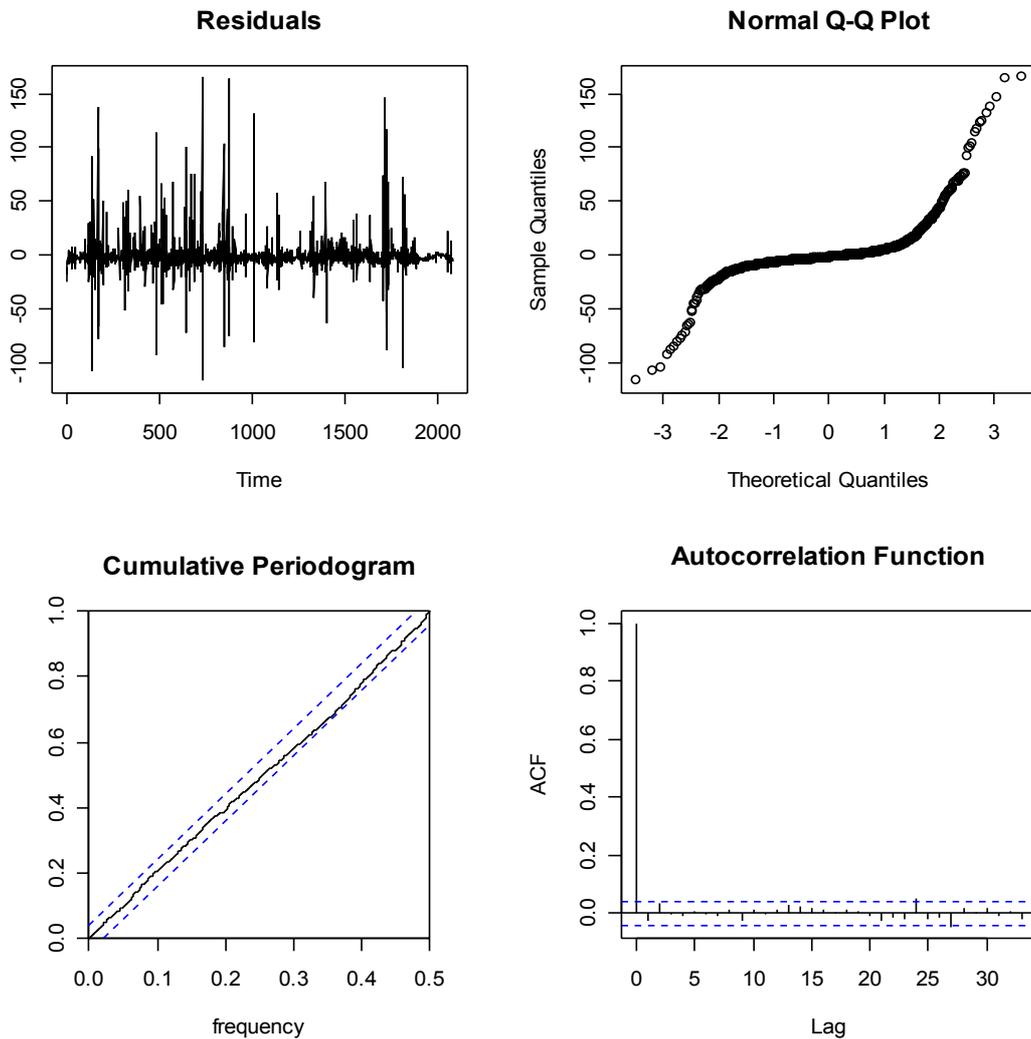


Figura 20: Grafici della distribuzione dei residui, quantile – quantile, periodogramma e ACF per il trimestre ottobre – dicembre 2002.

In generale si nota che per i due campioni dell'anno 2002 la temperatura risulta praticamente l'unica variabile esplicativa, mentre l'umidità relativa non entra quasi a far parte del modello; fatta eccezione per il periodo luglio – settembre dove è presente con un solo coefficiente al ritardo 1. Nei campioni dell'anno precedente, invece, c'è solo l'umidità come variabile esplicativa.

4.3 Punti di cambiamento

Per iniziare la ricerca dei punti di cambiamento si considerino le figure 21, 22, 23, 24.

Gli ultimi due grafici di ciascuna figura evidenziano che, proprio dove le serie fuoriescono dalle bande di confidenza, o dove hanno andamento a salti, si può ipotizzare la presenza di *outliers* e di salti di livello.

I calcoli realizzati si sono presentati molto lunghi in termini di tempo, in quanto l'insieme dei possibili modelli è di cardinalità elevata, 2^{m^*} . Nel capitolo 3, m^* è stato espresso come il numero di geni in un cromosoma, secondo la formula:

$$m^* = p^* + q^* + \sum_{i=1}^k (r_i^* + s_i^* + R_i^* + S_i^*)$$

e ogni gene è un parametro del modello M stimato.

Con un computer AMD Sempron™ Processor 3000+ 1,8 GHz 512 MB di RAM, con sistema operativo Microsoft Windows XP Home Service Pack 2, i tempi di calcolo sono stati di circa 20 - 30 minuti per ogni procedura eseguita.

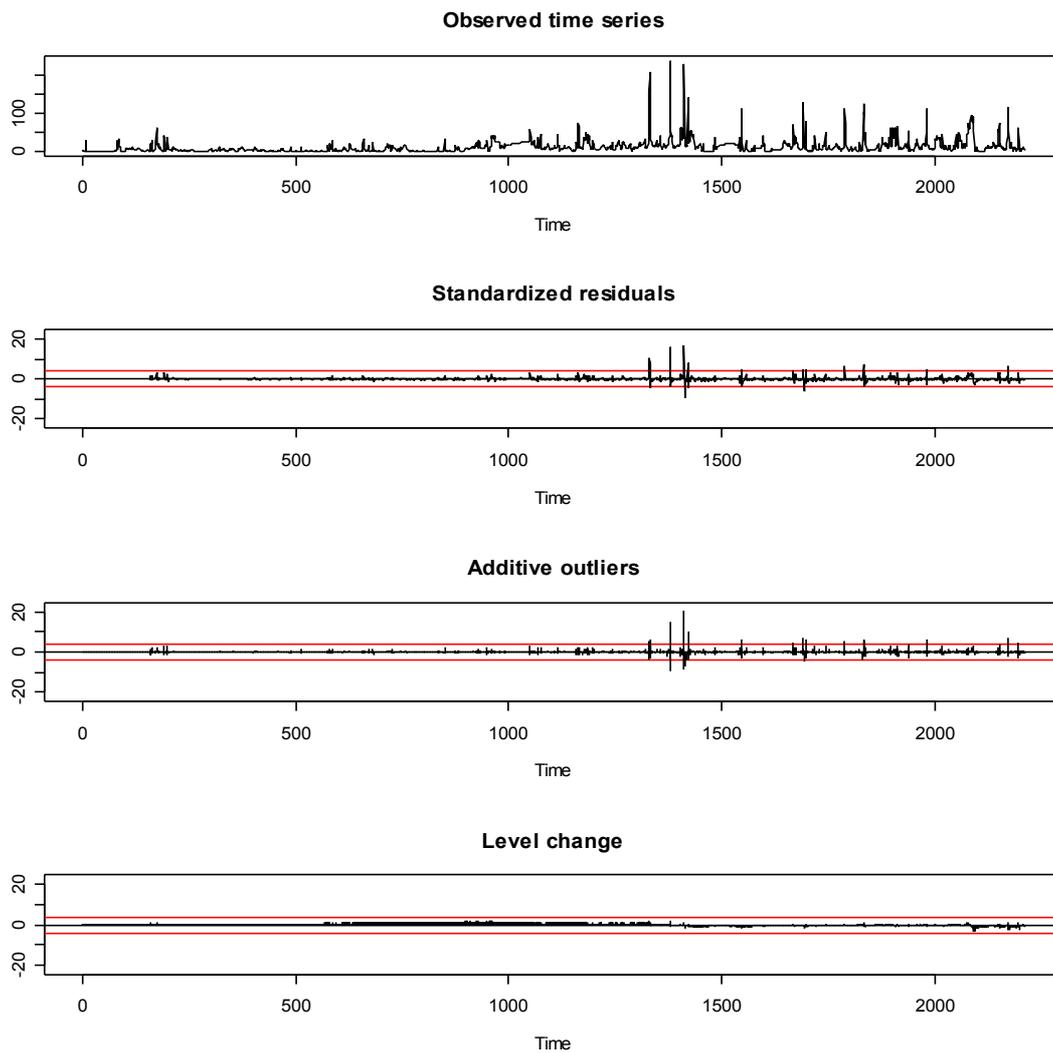


Figura 21: Grafici della serie originale, residui standardizzati e punti di cambiamento per il trimestre luglio – settembre 2001.

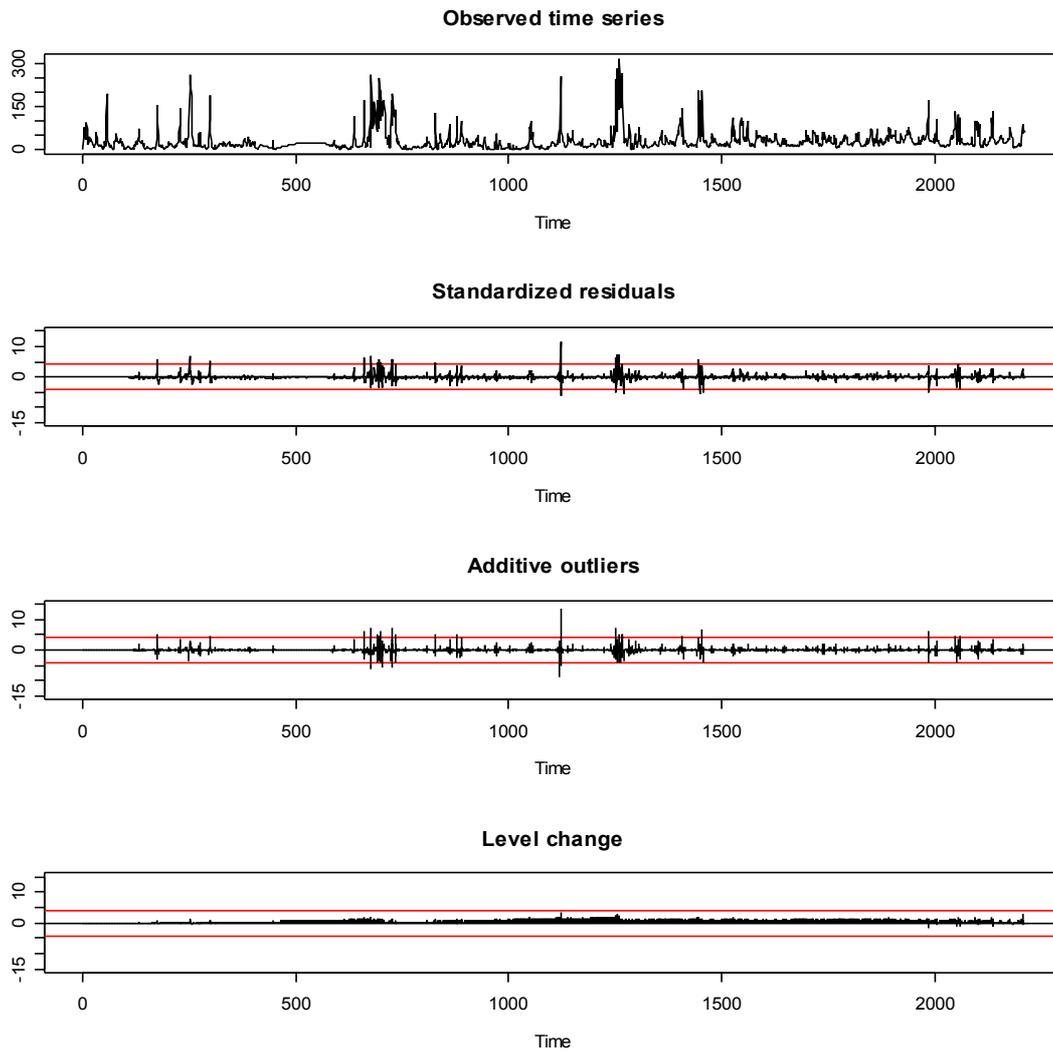


Figura 22: Grafici della serie originale, residui standardizzati e punti di cambiamento per il trimestre ottobre - dicembre 2001.

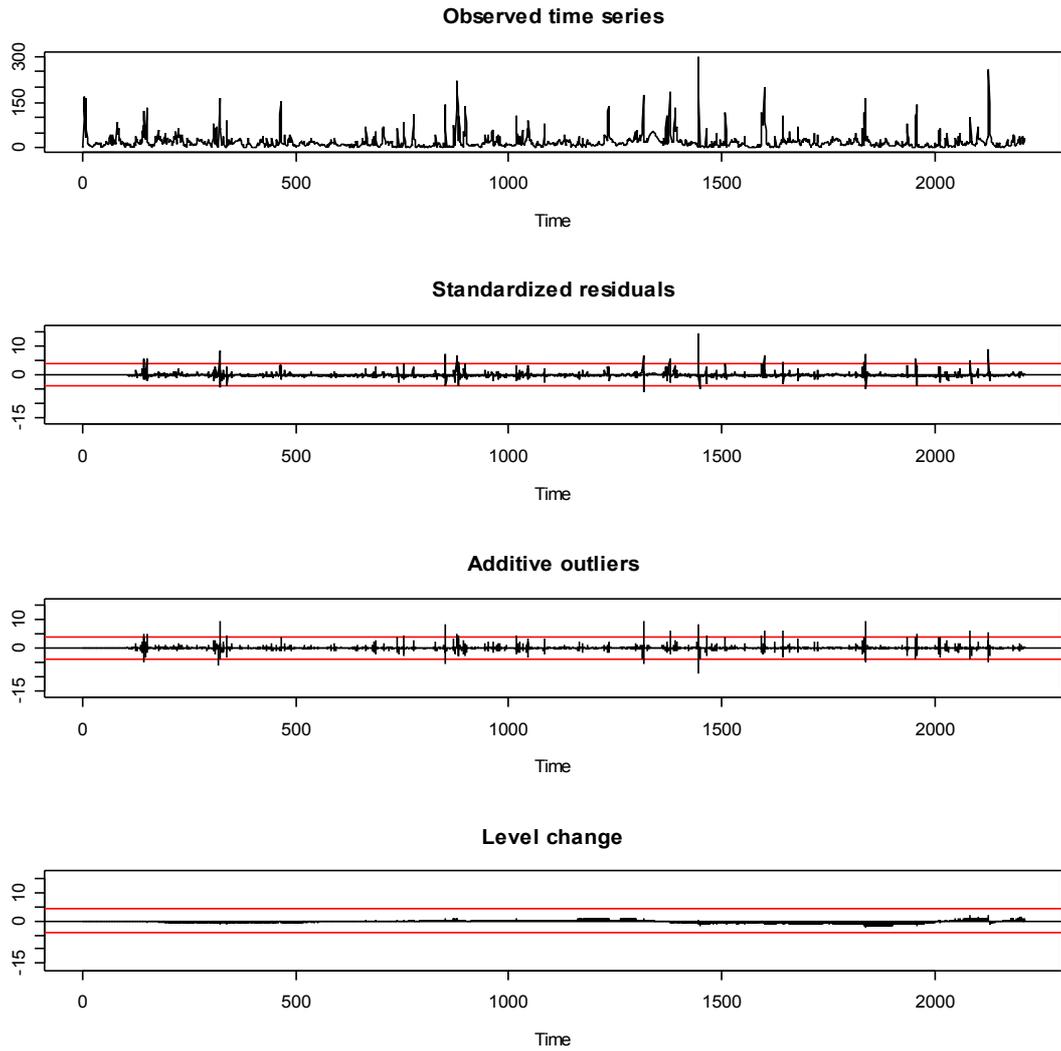


Figura 23: Grafici della serie originale, residui standardizzati e punti di cambiamento per il trimestre luglio – settembre 2002.

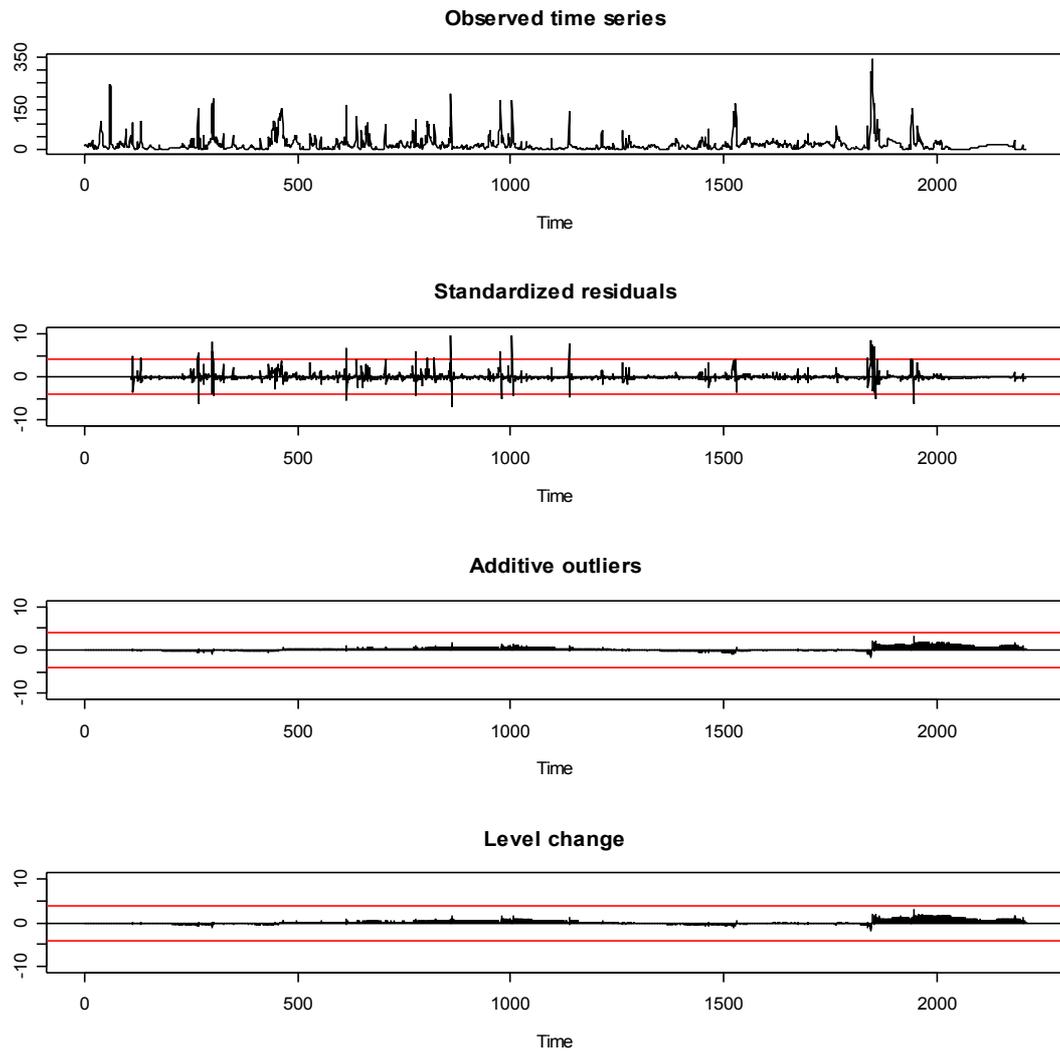


Figura 24: Grafici della serie originale, residui standardizzati e punti di cambiamento per il trimestre ottobre - dicembre 2002.

Considerando il trimestre ottobre – dicembre 2002, la figura 25 rappresenta nel primo grafico la serie originale, e la serie depurata dai cambi di livello avvenuti, e nel secondo grafico i cambi di livello e gli *outliers*.

Qui è evidente, dalla metà di ottobre, un netto distacco tra la serie originale, in nero, e quella depurata, in rosso. Il salto di livello e la presenza di *outliers* è ben visibile anche nella funzione a gradini.

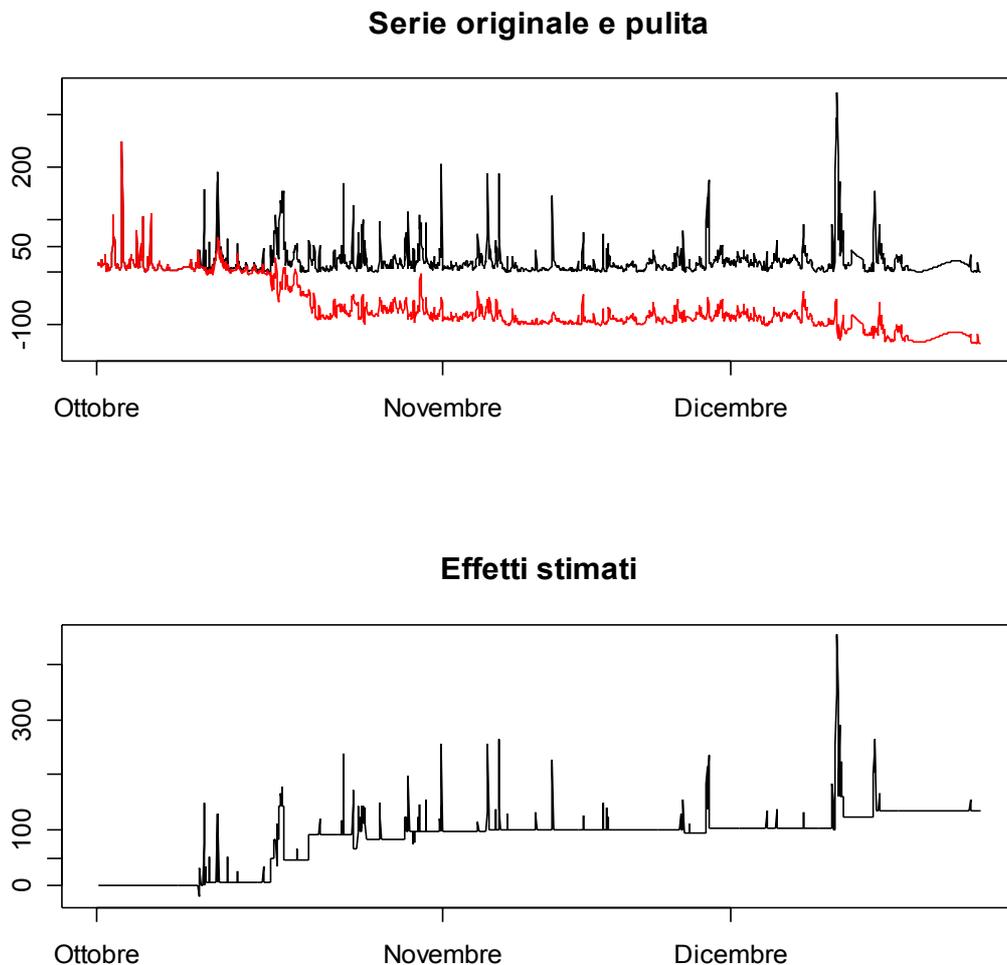


Figura 25: Nel primo grafico la serie originale, in nero, e la serie depurata, in rosso; nel secondo grafico i punti di cambiamento per il trimestre ottobre – dicembre 2002.

Un caso differente invece si evidenzia nella figura 26, che fa riferimento al periodo luglio – settembre 2002.

Come si nota, le due serie, originale e depurata, sono praticamente sovrapposte, e il grafico dei punti di cambiamento raccoglie infatti solo *outliers*: il livello rimane stabile senza salti, e i picchi presenti sono rilevati in corrispondenza di osservazioni dal valore particolarmente elevato già nella serie originale.

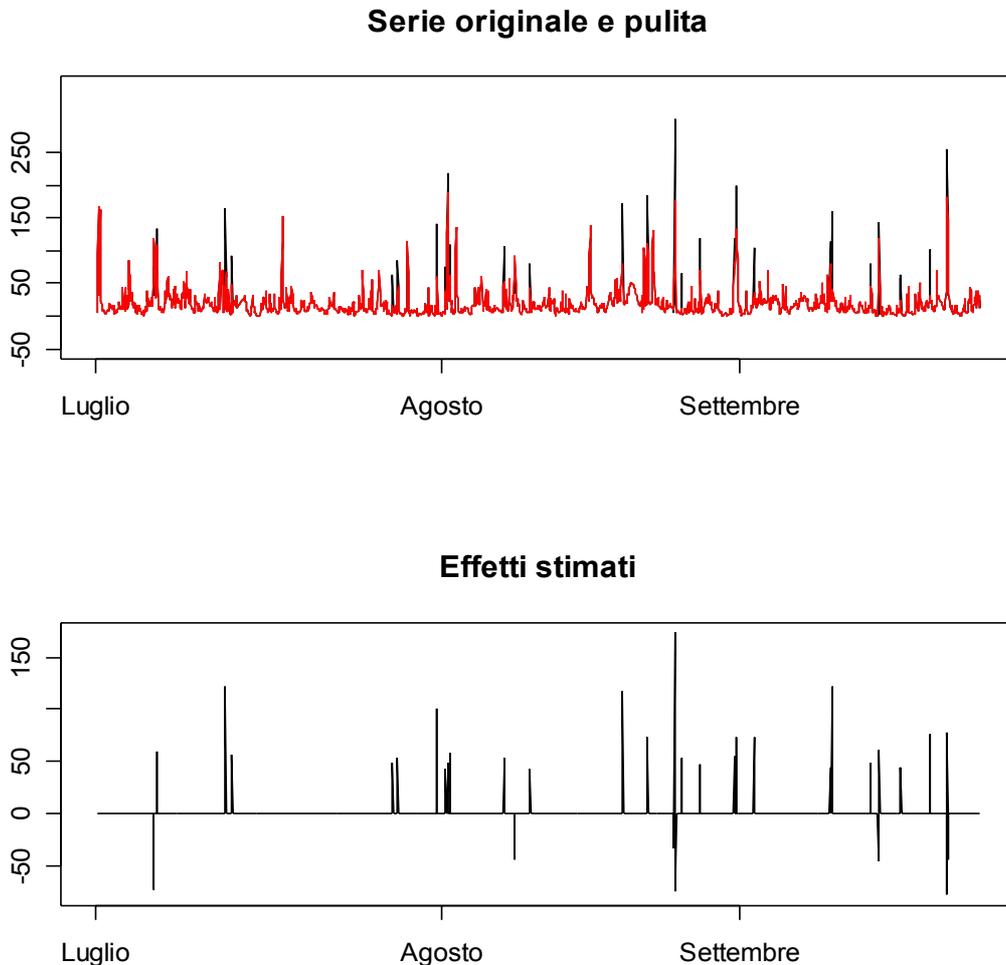


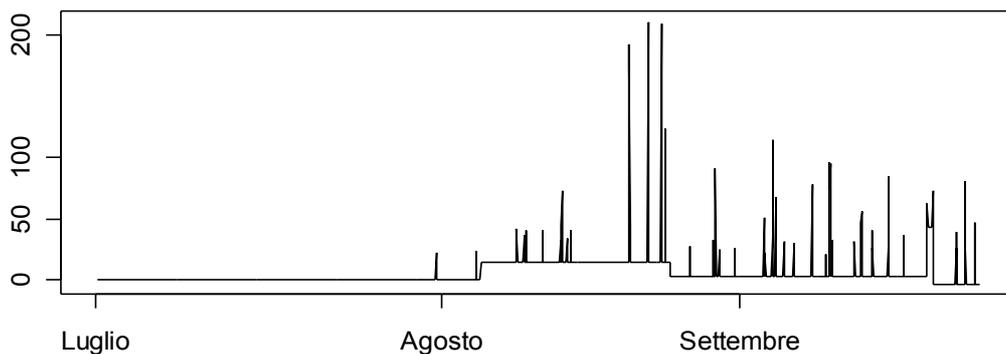
Figura 26: Nel primo grafico la serie originale, in nero, e la serie depurata, in rosso; nel secondo grafico i punti di cambiamento per il trimestre luglio – settembre 2002.

Nelle figure seguenti invece, 27 e 28, si presenta solo il grafico dei punti di cambiamento perché, come per il caso della figura 26, la serie originale e la serie depurata coincidono.

Il primo grafico di ogni figura è ricavato con la procedura “ISGN” e il secondo con la procedura basata sull’algoritmo genetico.

Confrontando i grafici ottenuti con il metodo “ISGN”, procedimento iterativo non guidato da un algoritmo genetico, con quelli che derivano dalla procedura guidata da un algoritmo genetico, sembrano non esserci grandi differenze.

Effetti stimati con 'istep' per il trimestre Luglio - Settembre 2001



Effetti stimati con 'genstep' per il trimestre Luglio - Settembre 2001

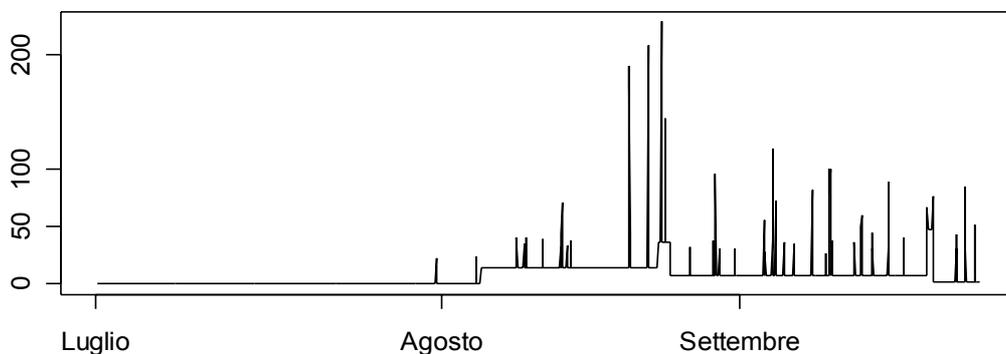
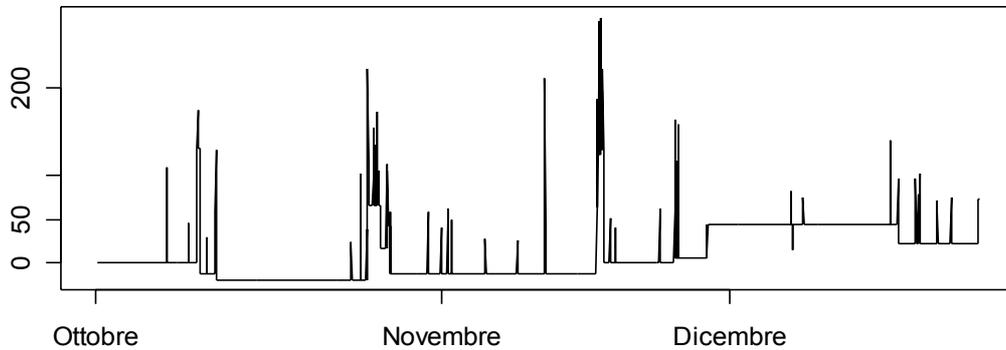


Figura 27: Grafici a gradini generati con i metodi “istep” (ISGN) e “genstep” (genetico) per il trimestre luglio – settembre 2001.

Effetti stimati con 'istep' per il trimestre Ottobre - Dicembre 2001



Effetti stimati con 'genstep' per il trimestre Ottobre - Dicembre 2001

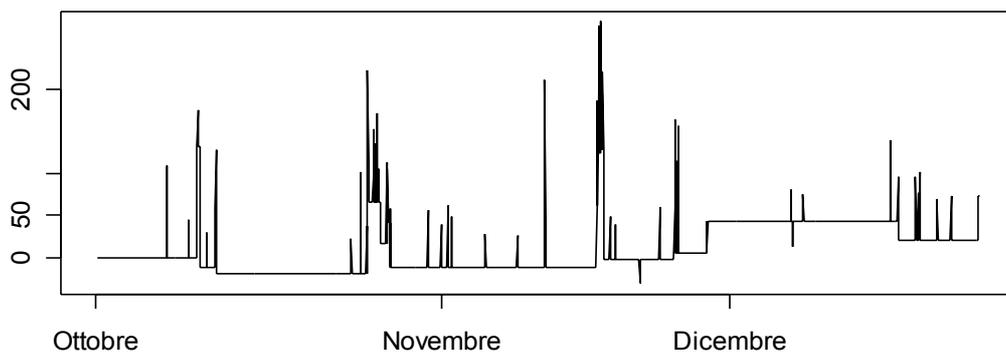


Figura 28: Grafici a gradini generati con i metodi "istep" (ISGN) e "genstep" (genetico) per il trimestre ottobre – dicembre 2001.

Per ciò che riguarda gli esiti riportati, non si sono riscontrate notevoli diversità tra le procedure: i grafici relativi all'uno e all'altro metodo sono pressoché uguali, e anche i risultati numerici non si differenziano. L'unico vantaggio a favore del metodo *ISGN* è di avere tempi di calcolo più brevi del caso genetico, tra 10 e 15 minuti contro mediamente 30 minuti. Infatti bisogna riconoscere che la procedura genetica procede con l'ausilio di un algoritmo genetico che ha un'esecuzione molto più complessa dello sviluppo iterativo della metodologia *ISGN*.

4.4 La serie annuale

Dopo aver illustrato l'individuazione di punti di cambiamento nelle serie di quattro intervalli campione composti da dati orari, si passa ora all'analisi delle medie giornaliere delle concentrazioni di SO₂.

Si tengono quindi solo 365 dati per anno contro 8760 rilevazioni di medie orarie per variabile e per anno.

Le concentrazioni medie giornaliere hanno mantenuto l'aspetto generale della serie iniziale, ma l'effetto lisciante della media ha levigato leggermente alcuni dei picchi che la serie presentava originariamente. La figura 29 presenta la serie così trasformata.

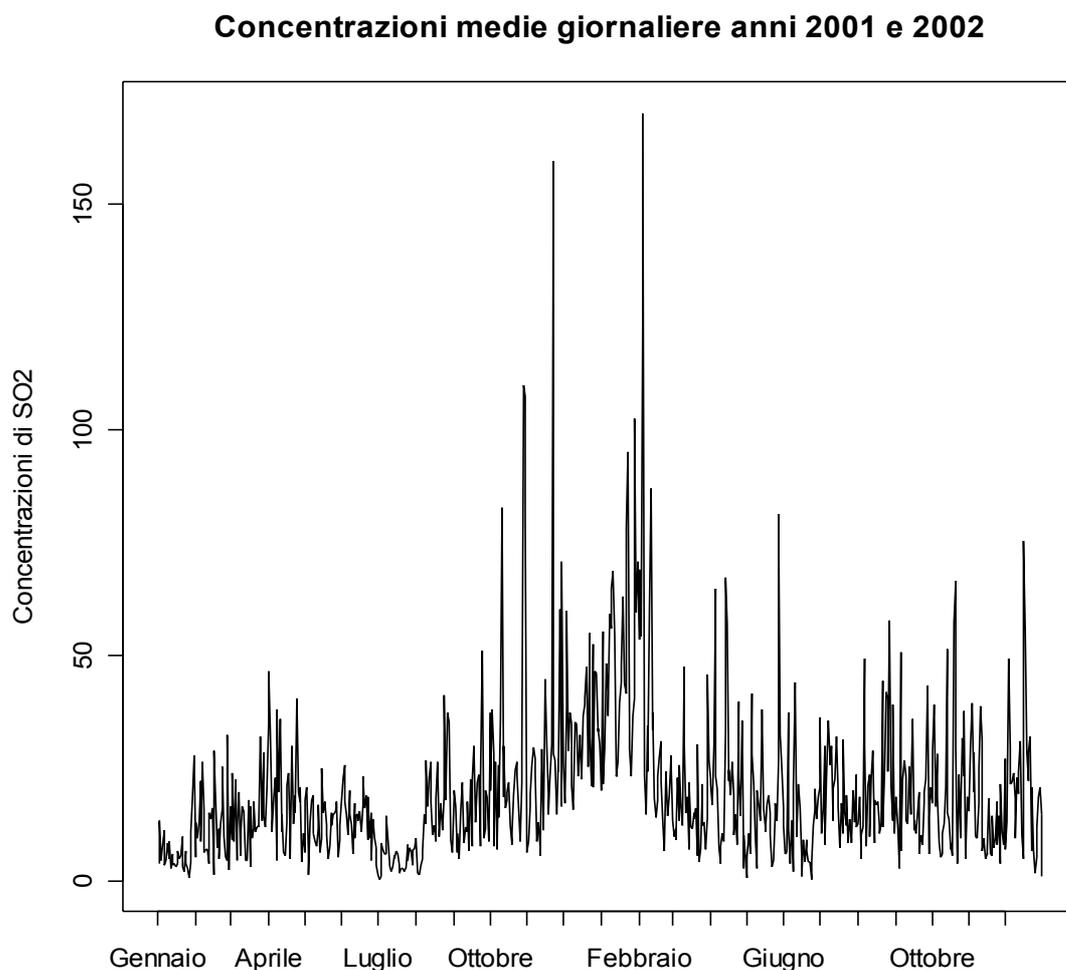


Figura 29: Serie delle concentrazioni medie giornaliere di SO₂ per gli anni 2001 e 2002.

Si stima ora un modello a funzione di trasferimento per le osservazioni medie giornaliere di SO₂, e delle variabili esplicative, temperatura e umidità relativa:

$$Y(t) = (1 + 0,465B)T(t) + \frac{(1 - 0,596B - 0,233B^4)(1 - 0,953B^{12})}{1 - 0,899B - 0,086B^5}e(t)$$

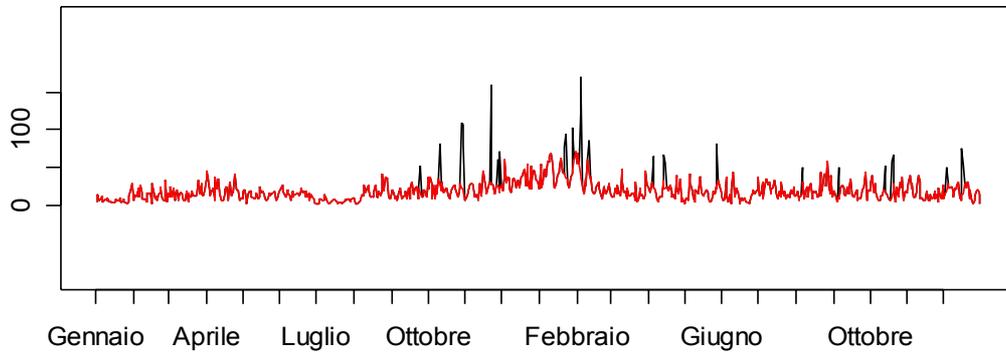
La tabella 12 mostra le stime dei coefficienti del modello ottenuto. Il valore del test di Box-Ljung è 9,717, gdl: 14 e p-value: 0,783.

Coefficienti	Stima	Std.Error	z value	Pr(> z)
T.phi[1]	-0,465	0,164	-2,836	0,005
Noise.phi[1]	0,597	0,047	12,684	0,000
Noise.phi[4]	0,233	0,034	6,904	0,000
Noise.PHI[12]	0,953	0,012	76,365	0,000
Noise.theta[1]	0,899	0,035	25,916	0,000
Noise.theta[5]	0,086	0,031	2,775	0,006

Tabella 12: Stime dei coefficienti del modello a funzione di trasferimento.

La figura 30 mostra i risultati della ricerca dei punti di cambiamento nella serie delle rilevazioni medie giornaliere. Come si nota la serie depurata dai cambi di livello è sovrapposta alla serie originale, e quindi si identificano solo *outliers*.

Serie osservata e depurata



Effetti stimati

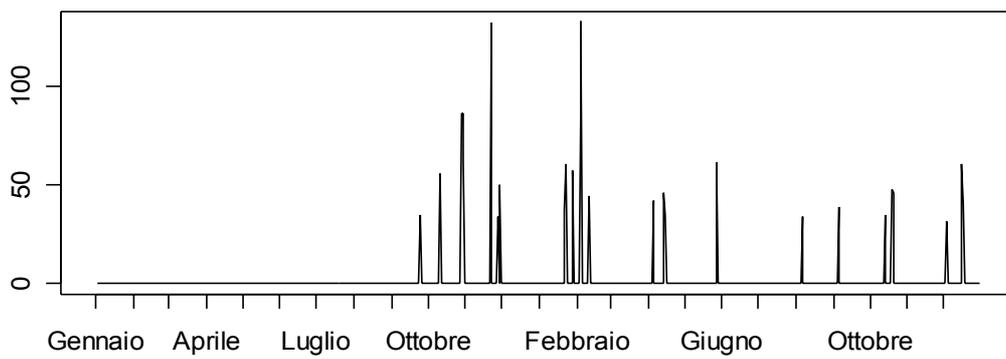


Figura 30: Nel primo grafico la serie originale, in nero, e la serie depurata, in rosso; nel secondo grafico gli outliers.

4.5 Confronti

A causa della trasformazione dei dati orari in medie giornaliere, sono stati rilevati molti meno punti di cambiamento.

La tabella 13 riassume questi risultati e specifica: se il punto di cambiamento è un *outlier* (AO) o un salto di livello (LC), indica il giorno preciso dell'anno in cui è avvenuto, e mostra le stime.

AO/LC	Giorno	Stima	Std.Error	zvalue	Pr(> z)
AO	25-set	35,084	8,730	4,019	0,000
AO	11-ott	56,006	8,746	6,403	0,000
AO	29-ott	86,610	9,156	9,460	0,000
AO	30-ott	86,085	9,206	9,351	0,000
AO	31-ott	61,648	9,115	6,763	0,000
AO	22-nov	132,179	8,785	15,046	0,000
AO	28-nov	34,242	8,834	3,876	0,000
AO	30-nov	49,752	8,838	5,629	0,000
AO	22-gen	36,889	8,942	4,126	0,000
AO	23-gen	60,569	8,961	6,759	0,000
AO	29-gen	57,470	8,798	6,532	0,000
AO	04-feb	84,412	9,026	9,352	0,000
AO	05-feb	132,918	8,984	14,794	0,000
AO	11-feb	44,687	8,813	5,071	0,000
AO	05-apr	42,344	8,766	4,830	0,000
AO	14-apr	46,031	8,930	5,155	0,000
AO	15-apr	34,484	8,931	3,861	0,000
AO	28-mag	61,069	8,751	6,978	0,000
AO	06-ago	34,109	8,762	3,893	0,000
AO	05-set	38,982	8,780	4,440	0,000
AO	13-ott	34,475	8,809	3,914	0,000
AO	19-ott	48,046	9,001	5,338	0,000
AO	20-ott	45,849	8,993	5,098	0,000
AO	03-dic	31,281	8,815	3,549	0,000
AO	16-dic	60,538	9,057	6,684	0,000
AO	17-dic	40,233	9,042	4,450	0,000

Tabella 13: differenza tra *outlier* (AO) e cambio di livello (LC), giorno dell'anno in cui è avvenuto il cambio, stime.

La tabella 13 è stata confrontata con la tabella nell'appendice 2.

La prima fa riferimento alle analisi svolte sulla serie delle concentrazioni medie giornaliere di SO₂; la seconda si riferisce alle concentrazioni medie orarie del SO₂.

Entrambe contengono le stime e i giorni precisi dell'anno in cui hanno avuto luogo *outliers* e salti di livello.

Si è notato che i punti di cambio avvenuti, si possono osservare nello stesso giorno sia con le analisi annuali, che con le analisi dei dati orari.

L'unica differenza è che un'anomalia che si presenta solo in un giorno nella serie delle concentrazioni giornaliere, può estendersi per qualche ora dello stesso giorno nelle serie dei dati orari, e presentarsi nell'arco delle ore con natura diversa da quella rilevata nelle analisi annuali.

Ad esempio, nel giorno 30 ottobre 2001, a causa dell'effetto aggregante della media aritmetica e della diversità dei modelli scelti per i due tratti messi a confronto, le analisi giornaliere rilevano un *outlier*, le analisi orarie riscontrano quattro *outliers* e un salto di livello; nel giorno 5 aprile 2002 le analisi annuali hanno rilevato un'anomalia, mentre non è avvenuta secondo i risultati sui dati orari.

Capitolo 5

Conclusioni

I quattro trimestri, dei quali sono state riportate le analisi integralmente, sono rappresentanti dell'andamento della serie dei due anni, e sono stati scelti a titolo d'esempio tra gli otto periodi in cui era stata suddivisa la serie originale. Tutte le caratteristiche riscontrate in questi campioni si possono notare sia negli altri intervalli sia nella serie annuale di dati giornalieri, quindi si considera attendibile la generalizzazione all'intera serie dei risultati ottenuti per questi tratti.

Dopo aver svolto le analisi sui dati a disposizione, ci si domanda se le condizioni atmosferiche contribuiscano in modo particolare all'aumento di inquinamento da SO_2 nell'aria.

Le variabili meteorologiche a disposizione per l'indagine erano soltanto temperatura, umidità relativa e vento.

Le misurazioni di temperatura e umidità relativa sono state inserite come variabili esplicative per l'andamento del biossido di zolfo nel modello a funzione di trasferimento.

Il fatto che si siano divisi i due anni in trimestri ha certamente fatto diminuire quello che poteva essere l'elemento stagionalità e quindi, in un certo senso, ha indebolito l'effetto delle temperature e dell'umidità relativa sulla serie dell'inquinante.

Per i quattro periodi considerati, come anche per gli altri, queste due variabili sono entrate senza molto peso nella stima del modello; le stime hanno infatti valori contenuti: mediamente di 0,196 per i coefficienti e di 0,136 per la stima dello *standard error*.

Per quanto riguarda il vento, come già si era concluso nel capitolo 2, questo ha il merito di spazzare via o, viceversa, trasportare, buona parte dell'inquinamento dell'aria. La causa delle elevate concentrazioni di SO₂ rilevate a Sacca Fisola è probabilmente da ricercarsi nelle emissioni inquinanti delle industrie dell'area di Mestre – Marghera, situata a non molti chilometri di distanza dalla stazione di rilevamento, trasportate da un vento che soffia particolarmente forte proprio da quella direzione.

La ricerca dei punti di cambiamento, obiettivo di questa tesi, ha portato all'identificazione di anomalie, sia negli intervalli di dati orari sia nella serie di dati giornalieri, che non sono spiegabili solo con i dati a disposizione; inoltre, non si distingue una relazione tra le condizioni meteorologiche e salti di livello individuati.

Per ipotizzare le cause di questi punti di cambiamento nella serie del biossido di zolfo nel sito di Sacca Fisola potrebbe essere utile confrontare i risultati ottenuti con quelli che si ottengono dalle informazioni su SO₂ tratte da altre stazioni di rilevamento.

Questo confronto, però, non è stato possibile perché i *dataset* che contenevano questi dati erano insufficienti per quanto riguarda le rilevazioni sulle variabili meteorologiche, presentavano lunghi tratti di dati mancanti, e sarebbe stato impossibile prevederli.

Questa tesi, pertanto, si presenta come uno studio non definitivo e può dare spazio a nuove future analisi.

Appendice 1

Ulteriori risultati

Si riportano qui di seguito i risultati raggiunti nelle analisi per l'identificazione dei punti di cambiamento per i periodi non esposti nella trattazione del capitolo 4.

Per ogni trimestre si mostrano: il modello a funzione di trasferimento scelto, la tabella con le stime dei parametri del modello, i grafici con la serie originale e la serie pulita, il grafico con gli effetti stimati.

Trimestre gennaio – marzo 2001.

$$Y(t) = 6,507 + (1 - 0,608B)I(t) + \left(\frac{(1 - 0,126B^2 + 0,099B^3)}{(1 - 0,815B)(1 - 0,1B^{48})} \right) e(t)$$

Coefficienti	Stima	Std.Error	z value	Pr(> z)
Intercetta	6,507	1,670	3,896	9,78E-02
T.phi[1]	0,608	0,176	3,459	5,43E-01
Noise.phi[2]	0,126	0,026	4,807	1,53E-03
Noise.phi[3]	-0,099	0,025	-3,988	6,67E-02
Noise.theta[1]	0,815	0,017	48,528	0,00E+00
Noise.THETA[48]	0,100	0,023	4,420	9,89E-03

Tabella 1: Stime dei coefficienti del modello a FT per il trimestre gennaio – marzo 2001.

Box-Ljung test: 18,07, gdl: 17, p-value: 0,384.

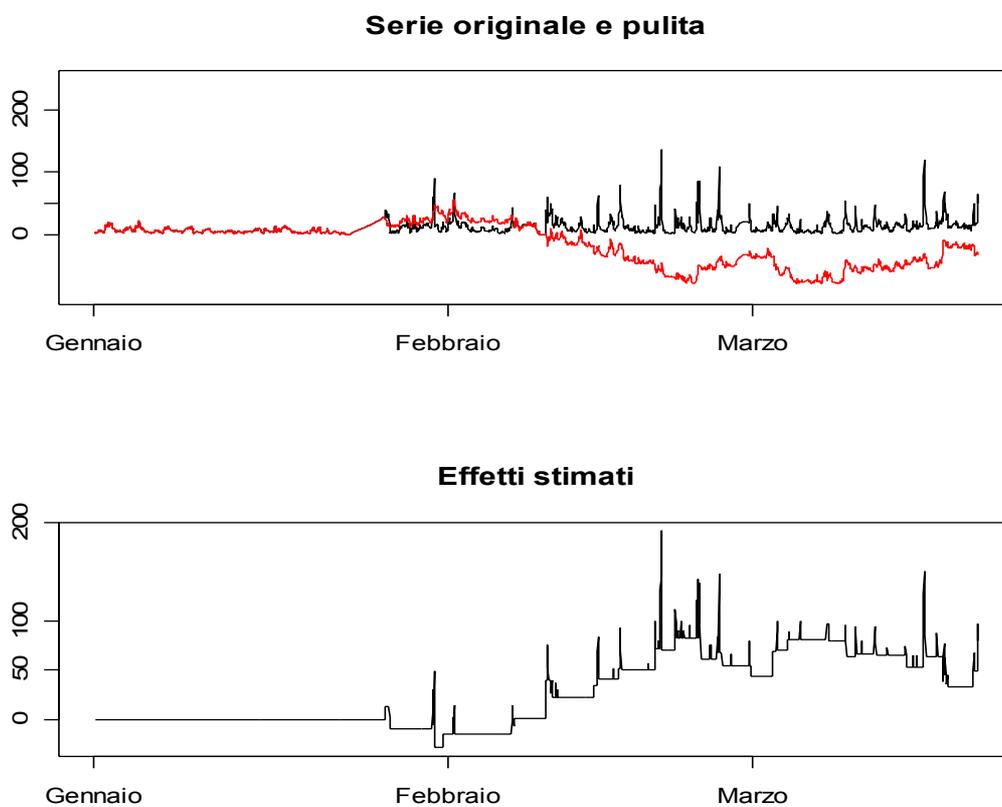


Figura 1: Nel primo grafico la serie originale, colore nero, e la serie pulita, colore rosso; nel secondo grafico i punti di cambiamento per il trimestre gennaio – marzo 2001.

Trimestre aprile – giugno 2001.

$$Y(t) = \left(\frac{(1 - 0,663B^2)}{(1 + 0,683B^4)} \right) T(t) + \left(\frac{(1 - 0,983B)}{(1 - 0,593B)} \right) e(t)$$

Coefficienti	Stima	Std.Error	z value	Pr(> z)
T.phi[2]	0,663	0,207	3,205	1,35E-03
T.theta[4]	-0,683	0,156	-4,371	1,24E-05
Noise.phi[1]	0,983	0,004	240,802	0,00E+00
Noise.theta[1]	0,593	0,019	31,360	0,00E+00

Tabella 2: Stime dei coefficienti del modello a FT per il trimestre aprile – giugno 2001.

Box-Ljung test: 21,55, gdl: 17, p-value: 0,203.

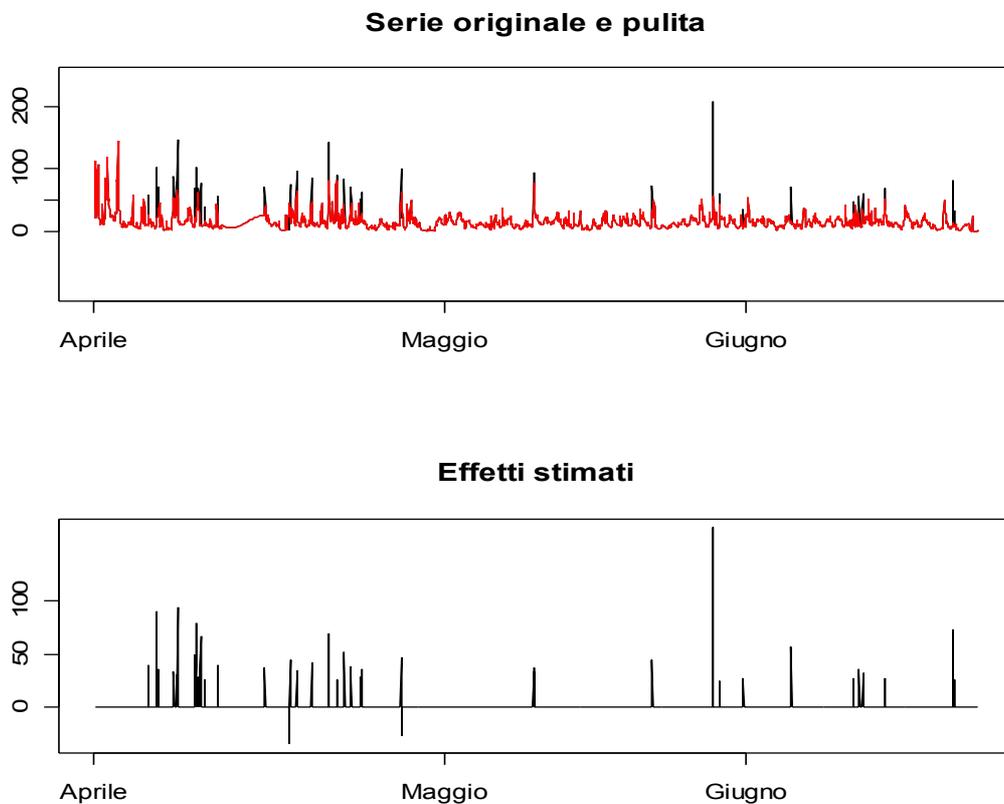


Figura 2: Nel primo grafico la serie originale, colore nero, e la serie pulita, colore rosso; nel secondo grafico i punti di cambiamento per il trimestre aprile – giugno 2001.

Trimestre gennaio – marzo 2002.

$$Y(t) = (1 + 1,970B^5)I(t) + \left(\frac{(1 - 0,216B - 0,282B^2 - 0,101B^3 - 0,377B^6)(1 + 0,082B^{24})}{(1 - 0,298B^6)} \right) e(t)$$

Coefficienti	Stima	Std.Error	z value	Pr(> z)
T.phi[5]	-1,970	0,590	-3,341	8,36E-04
Noise.phi[1]	0,216	0,021	10,190	0,00E+00
Noise.phi[2]	0,282	0,021	13,366	0,00E+00
Noise.phi[3]	0,101	0,022	4,687	2,77E-06
Noise.phi[6]	0,377	0,027	13,808	0,00E+00
Noise.PHI[24]	-0,082	0,022	-3,668	2,45E-04
Noise.theta[6]	0,298	0,032	9,337	0,00E+00

Tabella 3: Stime dei coefficienti del modello a FT per il trimestre gennaio – marzo 2002.

Box-Ljung test: 11,67, gdl: 16, p-value: 0,766.

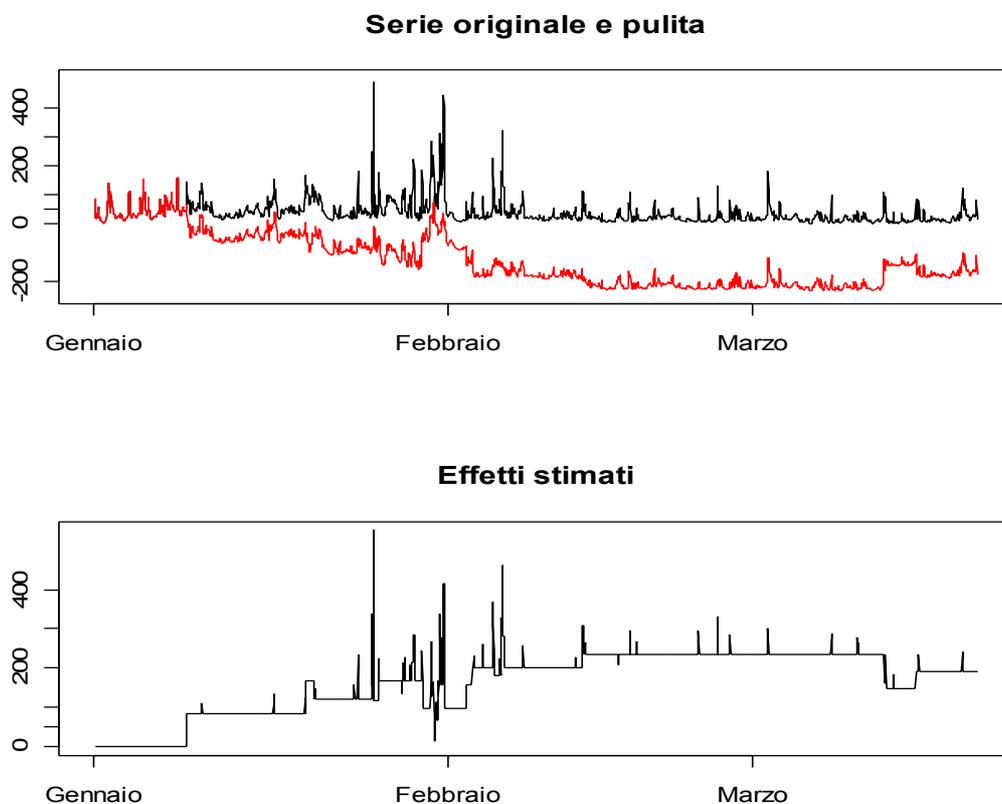


Figura 3: Nel primo grafico la serie originale, colore nero, e la serie pulita, colore rosso; nel secondo grafico i punti di cambiamento per il trimestre gennaio – marzo 2002.

Trimestre aprile – giugno 2002.

$$Y(t) = 35,075 + \left(\frac{(1 - 1,822B + 2,382B^3)}{(1 - 0,407B^2)} \right) T(t) + \left(\frac{(1 - 0,111B^5)}{(1 - 0,734B - 0,103B^5)} \right) e(t)$$

Coefficienti	Stima	Std.Error	z value	Pr(> z)
Intercetta	35,075	5,911	5,934	2,96E-09
T.phi[1]	1,822	0,326	5,589	2,28E-08
T.phi[3]	-2,382	0,332	-7,176	7,18E-13
T.theta[2]	0,407	0,107	3,791	1,50E-04
Noise,phi[5]	0,111	0,029	3,850	1,18E-04
Noise,theta[1]	0,734	0,016	46,996	0,00E+00
Noise,theta[5]	0,103	0,020	5,177	2,26E-07

Tabella 4: Stime dei coefficienti del modello a FT per il trimestre aprile – giugno 2002.

Box-Ljung test: 15,34, gdl: 17, p-value: 0,571.

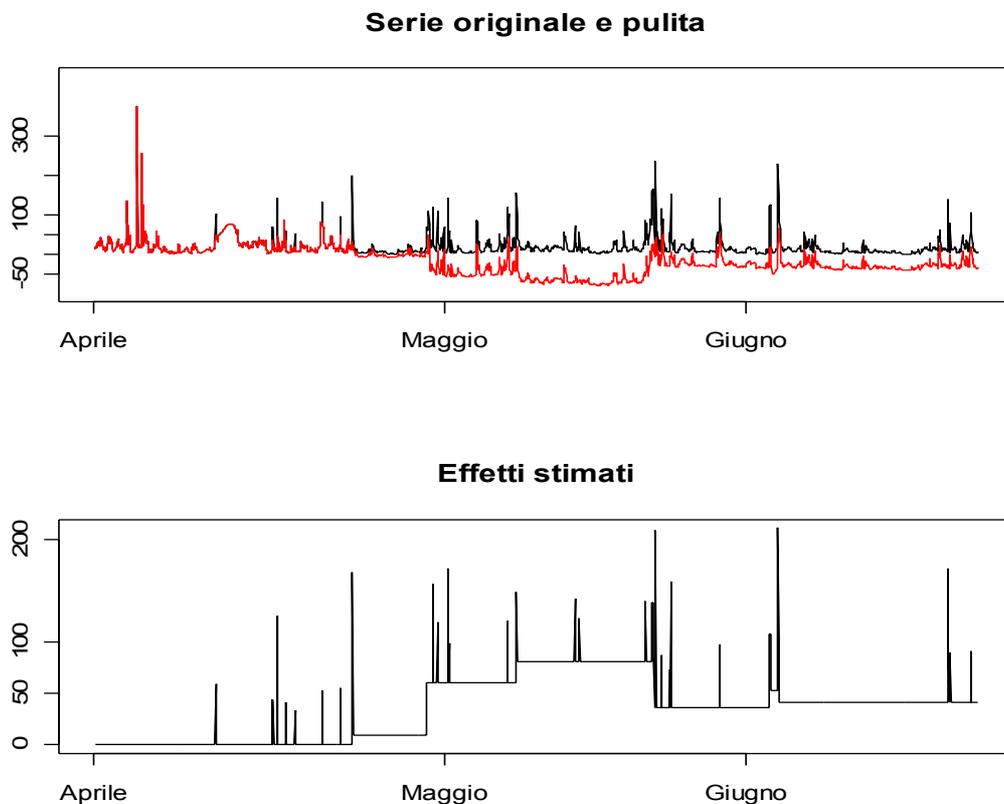


Figura 4: Nel primo grafico la serie originale, colore nero, e la serie pulita, colore rosso; nel secondo grafico i punti di cambiamento per il trimestre aprile – giugno 2002.

Appendice 2

Tabella dei punti di cambio

Tabella che mostra *outliers* (AO) o cambi di livello (LC), identificati nelle serie delle concentrazioni orarie di biossido di zolfo suddivise in trimestri, avvenuti negli stessi giorni in cui si sono evidenziati punti di cambiamento nella serie delle concentrazioni medie giornaliere di SO₂ per gli anni 2001 e 2002.

giorno	AO/LC	Stima	Std.Error	zvalue	Pr(> z)
11-ott	LC	130,799	9,069	14,422	0,000
	AO	43,604	7,050	6,185	0,000
	LC	-76,523	8,988	-8,514	0,000
	LC	-66,148	8,700	-7,603	0,000
29-ott	AO	56,816	6,969	8,153	0,000
	LC	132,807	9,394	14,137	0,000
	AO	108,104	7,195	15,024	0,000
	LC	-47,555	9,329	-5,098	0,000
	AO	37,425	6,960	5,378	0,000
	AO	87,162	7,059	12,347	0,000
30-ott	AO	68,568	7,072	9,696	0,000
	AO	105,433	7,127	14,794	0,000
	AO	38,006	6,987	5,440	0,000
	AO	39,959	6,951	5,749	0,000
	LC	-49,170	8,434	-5,830	0,000
31-ott	AO	95,501	7,129	13,396	0,000
	LC	87,806	9,662	9,088	0,000
	LC	-62,889	9,966	-6,311	0,000
	LC	-54,101	9,185	-5,890	0,000
	AO	70,622	7,013	10,070	0,000

22-nov	LC	75,021	9,078	8,264	0,000
	AO	123,661	7,055	17,529	0,000
	LC	157,514	9,868	15,963	0,000
	AO	55,383	7,352	7,533	0,000
	LC	-96,133	10,381	-9,261	0,000
	LC	155,520	10,534	14,763	0,000
	LC	-85,472	10,287	-8,309	0,000
	LC	-66,309	10,204	-6,498	0,000
	AO	53,988	9,210	5,862	0,000
	AO	93,043	8,962	10,382	0,000
	LC	-129,222	9,296	-13,902	0,000
30-nov	LC	60,773	9,460	6,424	0,000
	LC	103,739	9,522	10,895	0,000
	LC	-158,170	10,371	-15,251	0,000
	LC	110,146	10,306	10,687	0,000
	LC	-110,428	9,711	-11,372	0,000
	AO	87,026	8,958	9,715	0,000
	AO	152,272	8,968	16,980	0,000
23-gen	LC	-44,555	9,459	-4,710	0,000
	AO	26,522	6,345	4,180	0,000
29-gen	LC	128,774	9,839	13,089	0,000
	AO	89,835	6,733	13,342	0,000
	LC	-78,696	9,876	-7,968	0,000
	LC	-53,808	9,837	-5,470	0,000
	AO	433,062	8,640	50,125	0,000
	AO	154,314	8,430	18,306	0,000
	LC	106,649	9,678	11,020	0,000
	LC	-55,986	9,660	-5,795	0,000
04-feb	LC	67,440	9,791	6,888	0,000
	AO	102,964	6,681	15,412	0,000
	AO	-35,682	6,473	-5,513	0,000
	LC	-96,443	9,748	-9,893	0,000
	AO	-56,196	6,531	-8,605	0,000
	AO	44,631	6,316	7,066	0,000
	LC	100,207	9,568	10,474	0,000
05-feb	AO	-38,103	6,548	-5,819	0,000
	LC	51,951	10,141	5,123	0,000
	LC	117,622	9,962	11,807	0,000
	LC	-53,789	9,874	-5,447	0,000
	LC	-123,874	9,928	-12,477	0,000
	AO	118,547	6,358	18,645	0,000
	LC	87,815	9,770	8,989	0,000
	LC	168,514	9,976	16,892	0,000

Appendice 2 - Tabella dei punti di cambio

	LC	-122,898	10,129	-12,134	0,000
	LC	-141,111	9,718	-14,520	0,000
	LC	-52,703	9,767	-5,396	0,000
28-mag	LC	46,004	7,773	5,918	0,000
	LC	-60,688	7,835	-7,745	0,000
	LC	-60,014	7,826	-7,668	0,000
	AO	173,174	5,207	33,260	0,000
05-set	AO	53,851	7,996	6,735	0,000
	AO	75,208	8,006	9,394	0,000
16-dic	LC	79,205	6,110	12,962	0,000
	LC	-31,016	6,005	-5,165	0,000
	LC	-51,798	6,238	-8,304	0,000
	LC	49,438	6,299	7,848	0,000
	LC	103,862	6,195	16,765	0,000
	LC	92,587	6,381	14,510	0,000
	AO	106,606	4,493	23,729	0,000
17-dic	LC	-92,446	6,253	-14,785	0,000
	LC	-93,220	6,066	-15,367	0,000
	AO	129,600	4,458	29,071	0,000
	AO	63,208	4,455	14,187	0,000
	LC	-37,730	5,844	-6,456	0,000

Appendice 3

Principali funzioni del software *R* utilizzate

Per precisare le analisi descritte nel capitolo 4, si descrive il funzionamento delle funzioni del programma *R* usate per la ricerca dei punti di cambiamento.

È stata adoperata la libreria “*smiso*” della versione 1.8.1 del programma *R*.

La funzione “*smisofit*”, con l’opzione “*strange=TRUE*”, determina le stime di tutti i cambiamenti di livello.

La specifica “*fit.by=genstep*” si riferisce all’algoritmo guidato da un procedimento genetico, invece la specifica “*fit.by=istep*” si riferisce a un algoritmo che lavora iterativamente ma senza l’ausilio di un algoritmo genetico, *ISGN*.

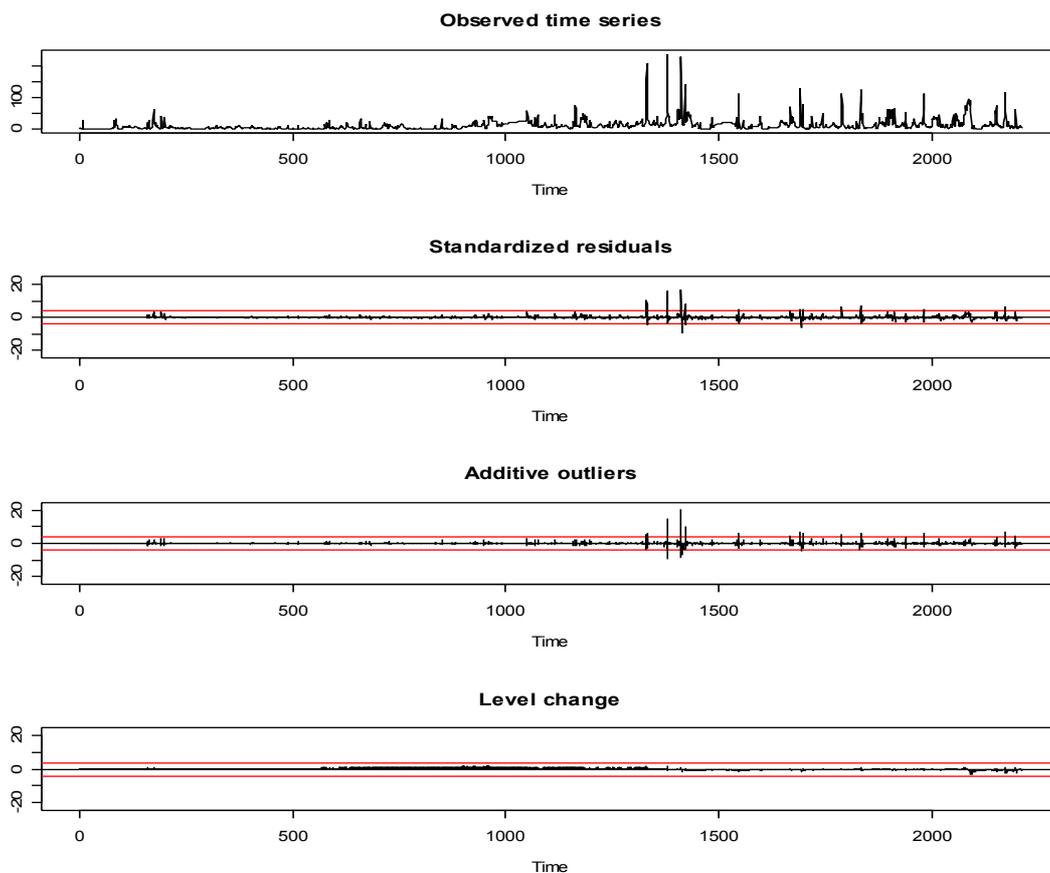
In sequenza le funzioni applicate per ottenere le stime di *outliers* e cambi di livello sono:

- `m<-smisofit(r(SO)~r(T,4,2,24)+r(U,4,2,24)+r(4,2,24),diff=c(1,24),
choose.diff=TRUE)`

Questa procedura stima un modello a funzione di trasferimento. Le variabili sono inserite in un modello di forma razionale, con un ritardo massimo possibile di 4, e la stagionalità di 24 ore, con un ritardo massimo di 2.

- *smisostrange(m)*

Con questa si ottiene il grafico della serie originale, i residui standardizzati e i punti di cambiamento. Dove le serie fuoriescono dalle bande di confidenza indicate, o dove hanno andamento a salti, si può ipotizzare la presenza di *outliers* e di salti di livello. Il grafico di esempio è il seguente:



- $mi \leftarrow smisofit(r(SO) \sim r(T, 4, 2, 24) + r(U, 4, 2, 24) + r(4, 2, 24), \text{diff} = c(1, 24), \text{choose.diff} = TRUE, \text{fit.by} = "istep", \text{strange} = TRUE)$

Dove le opzioni *fit.by="istep"* e *strange=TRUE* permettono la stima simultanea del modello a funzione di trasferimento e dei punti di cambiamento. Qui è utilizzata la tecnica *ISGN*.

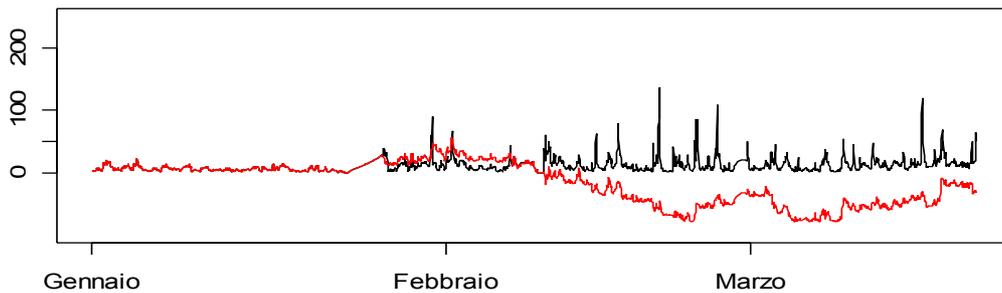
- `mg<-smisofit(r(SO)~r(T,4,2,24)+r(U,4,2,24)+r(4,2,24),diff=c(1,24),
choose.diff=TRUE, fit.by="genstep",strange=TRUE)`

In questo caso invece la stima del modello e la ricerca dei punti di cambio sono effettuate con la procedura genetica.

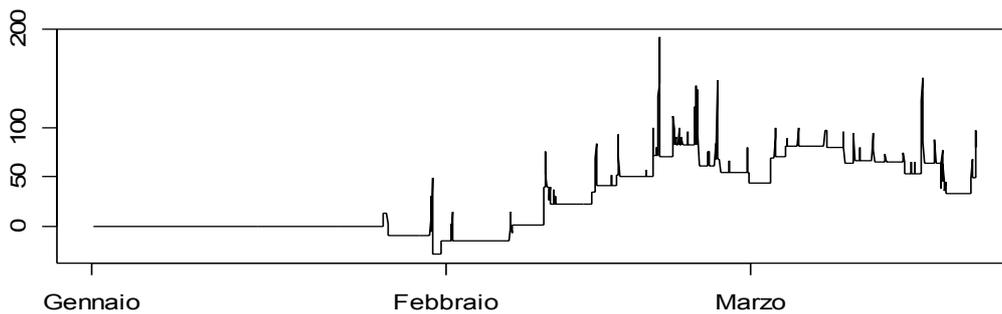
- `smisostrangeclean(mi)`

Permette di calcolare la serie pulita dai cambi di livello avvenuti ed inoltre, disegna in un grafico la serie originale (in nero) e depurata (in rosso) e in un altro la funzione a gradini con i cambi di livello; entrambi sono riportati nella figura in basso.

Serie originale e pulita



Effetti stimati



Bibliografia

Abraham, B. (1987). Application of Intervention Analysis to a Road Fatality Series in Ontario. *Journal of Forecasting*, **6**, 211-219.

Arpav (2001). *Qualità dell'aria nel comune di Venezia. Rapporto annuale 2001.*

Arpav (2002). *Qualità dell'aria nel comune di Venezia. Rapporto annuale 2002.*

Arpav (2003). *Qualità dell'aria nel comune di Venezia. Rapporto annuale 2003.*

Bahattacharyya, M. N. e Layton, A. P. (1979). Road Toll, an Australian Case Study in Intervention Analysis. *JASA*, **74**, 596-603.

Baird, C. (2001). *Chimica ambientale*. Zanichelli

Bordignon, C. e Scagliarini, M. (2000). Monitoring algorithms for detecting changes in the ozone concentrations. *Environmetrics*, **11**, 125-137.

Bruzzi, L. (1994). *L'uomo e l'ambiente: rischi e limiti dell'accettabilità*. Pitagora, Bologna.

Capizzi, C. (2001). Design of change detection algorithms based on the generalized likelihood ratio test. *Environmetrics*, **12**, 749-756.

Chiogna, M., Gaetan, C., Masarotto, G. (2003). Automatic identification of seasonal transfer models by means of iterative stepwise and genetic algorithms. Working paper.

Finzi, G., Brusasca, G. (1991). *La qualità dell'aria*. Masson, Milano.

Iqbal, M. e Yunus, M. (1996). *Plant Response to Air Pollution*. Wiley.

Jahnke, J. A. (2000). *Continuous Emission Monitoring*, 2nd Edition, Wiley.

Lai, T. L. (1995). Sequential change point detection in quality control and dynamic systems. *Journal of Royal Statistical Society*, **B 57**, 613-658.

McGraham, G. e Murray, F. (2003). *Air Pollution and Health in Rapidly Developing Countries*. Earthprint.

Mitchell, M. (1996). *Introduction to genetic algorithms*. MIT Press, Cambridge.

Nirel, R. (2000). Developing an efficient surveillance scheme for assessing compliance with air quality standards. Nrcse-trs No. **044**, Jerusalem.

Pollack, M. (1985). Optimal detection of a change in distribution. *Ann. Statist.*, **13**, 206-227.

Shiryayev, A. N. (1963). On optimum methods in quickest detection problems. *Theory Probab. Appl.*, **13**, 22-46.

Wahba, G. (1990). Spline models for observational data. *Society for Industrial and Applied Mathematics*, Philadelphia.

Willsky, A. S. Jones, H. L. (1976). A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems. *IEEE Transactions on Automatic Control*, **21**, 108-112.

Zavatti, C., Biondi, Mazzali, P. (1994). *Il controllo dell'ambiente: sintesi delle tecniche di monitoraggio ambientale*. Pitagora, Bologna.

Siti Internet di interesse

www.ambiente.it

www.ambientediritto.it

www.arpa.veneto.it

www.minambiente.it/Sito/home

www.nonsoloaria.com

Un ringraziamento sincero a tutte le persone che mi sono state vicine durante questi anni di studio e di vita.

Erico