



Università degli studi di Padova

---

Facoltà di Scienze Statistiche

Laurea in Scienze Statistiche, Economiche, Finanziarie ed Aziendali

**Concentrazioni di metalli nel particolato atmosferico  
presso un'acciaiera: un'analisi statistica**

Relatore:

Prof. Silvano Bordignon

Laureanda:

Tamara Feresin

Correlatori:

Dott. Francesco Pauli

Dott. Francesco Montanari

Anno Accademico 2006 – 2007



# INDICE

<b>INTRODUZIONE</b>	<b>1</b>
<b>CAPITOLO 1</b>	
<b>L'ACCIAIO: PRODUZIONE ED EMISSIONI</b>	<b>5</b>
1.1 - Le pratiche produttive	5
1.2 - La captazione dei fumi	13
1.3 - Lo stabilimento in esame	20
1.4 - Il monitoraggio	22
1.5 - Dati di letteratura	24
<b>CAPITOLO 2</b>	
<b>I DATI</b>	<b>27</b>
2.1 - Introduzione	27
2.2 - I dati di produzione	28
2.3 - I dati meteorologici	32
2.3.1 - Il vento	34
2.3.2 - La temperatura	37
2.3.3 - L'umidità relativa	40
2.3.4 - La pressione	41
2.3.5 - La pioggia	43
2.3.6 - L'irraggiamento	46
2.3.7 - Le correlazioni tra le variabili meteorologiche	47

## INDICE

---

2.4 - I dati di concentrazione dei metalli	49
2.4.1 - Manganese e Ferro	53
2.4.2 - Cadmio, Piombo e Zinco	56
2.4.3 - Nichel, Cromo, Vanadio ed Arsenico	60
2.4.4 - Correlazione tra gli elementi	64

### **CAPITOLO 3**

<b>I MODELLI</b>	<b>67</b>
3.1 - L'analisi delle componenti principali	67
3.2 - La censura dei dati	69
3.3 - I modelli binomiali	71
3.4 - I modelli lineari con variabili contemporanee	73
3.4.1 - Manganese e Ferro	74
3.4.2 - Nichel	79
3.4.3 - Cadmio, Piombo e Zinco	81
3.5 - Modelli con variabili esplicative ritardate	91
3.5.1 - Nichel	92
3.5.2 - Manganese e Ferro	94
3.5.3 - Cadmio, Piombo e Zinco	101
3.6 - La validazione incrociata	113

### **CAPITOLO 4**

<b>CONSIDERAZIONI CONCLUSIVE</b>	<b>117</b>
----------------------------------	------------

<b>APPENDICE</b>	<b>125</b>
------------------	------------

A.1 - L'implementazione di funzioni in R	125
A.2 - Elenco delle variabili	129

<b>BIBLIOGRAFIA</b>	<b>133</b>
---------------------	------------

<b>RINGRAZIAMENTI</b>	<b>135</b>
-----------------------	------------

# INTRODUZIONE

In questa Tesi si conduce un'analisi statistica dei valori di concentrazione in aria di alcuni metalli presenti nel particolato atmosferico totale<sup>1</sup> (PTS) nei pressi di un'acciaieria in prossimità della quale è posta una rete di quattro campionatori, posizionati dall'Agenzia Regionale per la Protezione dell'Ambiente del Friuli Venezia Giulia (ARPA FVG) nei pressi di un'acciaieria.

Il campionamento avviene pompando l'aria, prelevata dall'ambiente, attraverso dei filtri di cellulosa, su cui rimane adeso il particolato atmosferico.

Questi filtri vengono quindi analizzati con dei metodi chimici, con cui si determina la presenza di vari metalli, variamente correlati alle attività siderurgiche: Manganese, Ferro, Nichel, Cadmio, Piombo, Zinco, Cromo, Vanadio e Arsenico.

Alcuni di questi metalli hanno degli effetti cancerogeni, o più generalmente nocivi: per essi sono individuati valori limite di concentrazione, previsti espressamente dalla normativa o indicati dall'Organizzazione Mondiale della Sanità.

L'obiettivo di questo lavoro è quello di creare dei modelli che spieghino i valori osservati dai campionatori; in particolare, appare rilevante:

- stabilire delle correlazioni fra le concentrazioni dei metalli;

---

<sup>1</sup> in cui non viene selezionata una frazione più fine, per mezzo di impattatori aerodinamici; tipicamente, quella con diametro aerodinamico inferiore a 10  $\mu\text{m}$  (PM<sub>10</sub>)

- ottenere delle informazioni relativamente all'impatto dalle emissioni prodotte nelle diverse fasi dei processi che si svolgono nell'impianto;
- ottenere delle informazioni relativamente all'influenza sulle ricadute ambientali.

Le serie storiche disponibili hanno frequenza giornaliera e durata pluriennale: a differenza di quanto accade nel caso delle reti fisse di rilevamento della qualità dell'aria, è raro che i dati risultanti dal monitoraggio di un particolare sito industriale abbiano tali caratteristiche, venendo tipicamente acquisiti nel corso di campagne di misura più brevi, o frammentate, utilizzando strumenti montati su mezzi mobili.

Ciò può forse spiegare la difficoltà incontrata nel reperire studi analoghi nella letteratura scientifica e tecnica specialistica.

Viceversa, esiste un'ampia letteratura tecnica e normativa relativa ai presidi di contenimento, captazione ed abbattimento delle emissioni connesse con la produzione dell'acciaio per mezzo di forno elettrico ad arco (EAF), oltre a specifica documentazione relativa all'impianto di cui qui si tratta.

Esso è fornito di due forni ad arco, realizzati in momenti diversi, di tipo trifase: sono cioè muniti di tre elettrodi di grafite, che penetrano nel crogiolo attraverso altrettanti fori posti nella volta del forno. Da un quarto foro vengono captati i fumi prodotti dalla fusione (emissioni primarie).

Le emissioni secondarie, invece, si generano durante le fasi di carica del forno, che avvengono a volta aperta, e durante il riversamento (spillaggio) del metallo fuso; sono denominate secondarie anche le emissioni che sfuggono alla captazione attraverso le aperture della volta. Queste emissioni vengono captate attraverso una cappa posta sopra il forno; parte di esse può scavalcare la cappa e fuoriuscire all'esterno, a causa della sua posizione, elevata rispetto alla sorgente dei fumi.

Le emissioni convogliate vengono avviate a dei sistemi di abbattimento e, quindi, emesse a camino. Le emissioni secondarie e primarie non captate, che vengono disperse direttamente nell'aria, vengono classificate come emissioni diffuse. Da qui la necessità di affiancare e integrare la cappa

con dei sistemi di incapsulamento del forno (dog-house) o con una segregazione del reparto stesso (elephant-house).

Nello stabilimento in esame, uno dei forni è segregato con una elephant-house, l'altro con una dog-house.

In letteratura sono state reperite analisi riguardanti la composizione chimica caratteristica di campioni ottenuti dal quarto foro (primarie) e dalle cappe di aspirazione (secondarie).

A partire da tali fonti documentali è possibile raggruppare i metalli osservati in tre gruppi, a seconda della loro provenienza: uno è composto da Manganese e Ferro (entrambe metalli caratteristici di tutte le lavorazioni dell'acciaio, vengono emessi prevalentemente durante la fase fusoria, come emissioni primarie); un secondo comprende Cadmio, Zinco e Piombo (questi sono correlati tra di loro, in quanto il Cadmio è associato ai minerali di Zinco, mentre il Piombo è un'impurezza dello Zinco stesso; trattandosi di metalli basso-fondenti, vengono emessi prevalentemente già durante la fase di carica del forno, come emissioni secondarie); l'ultimo gruppo, infine, include i metalli rimanenti (tipici per la produzione di acciai speciali, vengono emessi soprattutto nella fase di affinazione).

La presenza di Cadmio, Zinco e Piombo si può ricondurre alla qualità dei rottami ferrosi utilizzati per la carica del forno.

Nello sviluppo dei modelli statistici ci si è concentrati sul campionatore più prossimo all'impianto, a valle della direzione dei venti prevalenti.

Le variabili esplicative utilizzate sono di diversa provenienza. Un primo gruppo è rappresentato dalle variabili meteorologiche, il secondo riguarda l'attività industriale dell'acciaiera. Le variabili meteorologiche sono importanti, in quanto dal tempo atmosferico dipendono i meccanismi di dispersione e deposizione delle polveri. Le variabili di produzione sono altrettanto importanti, al fine di riconoscere l'impatto dell'acciaiera rispetto ad altre possibili sorgenti.

Nei modelli sono stati inserite anche le variabili meteorologiche ritardate di un giorno; vista la loro autocorrelazione, si è dovuto porre molta

attenzione ai problemi di collinearità. Inoltre, gli errori sono risultati autocorrelati e quindi si è proceduto a modellare tale correlazione.

Il risultato finale è stato molto interessante: a seconda dell'affinità nella scelta delle variabili esplicative, si sono potuti distinguere tre gruppi di metalli, gli stessi attesi dopo l'esame della letteratura.

Infine, per testare la capacità predittiva di tali modelli, è stata effettuata una validazione incrociata e si sono confrontati con un modello di camminata casuale.



# CAPITOLO 1

## L'acciaio: produzione ed emissioni

### ***1.1 - Le pratiche produttive<sup>2</sup>***

Si definisce acciaio ogni lega del ferro che abbia un contenuto di carbonio non superiore al 2% in peso. Sopra tale tenore si parla di ghisa e non più di acciaio. La produzione dell'acciaio, dunque, è strettamente legata al controllo del carbonio in lega, oltre che a quello degli altri elementi che lo compongono ed ai quali sono da ricondurre le proprietà stesse della lega.

Gli acciai sono classificati, secondo la norma europea EN 10020, con riferimento sia alla composizione chimica, sia ai requisiti qualitativi:

- acciai di base: si tratta di acciai per i quali non è richiesta nessuna prescrizione particolare legata all'impiego, viene garantita una certa proprietà resistenziale, non è prescritto alcun trattamento termico;
- acciai di qualità: gli acciai che non presentano una regolarità di comportamento ai trattamenti termici, ma le cui condizioni di impiego richiedono accorgimenti particolari nella fase di produzione, in particolare per ridurre il rischio di rottura fragile e lo stato superficiale;

---

<sup>2</sup> Vedi [1] e [3]

- acciai speciali: destinati in genere ai trattamenti termici, con necessità di purezza, rispetto alle inclusioni gassose e non metalliche.

Le diverse qualità di acciaio sono ottenute seguendo sostanzialmente la stessa filiera produttiva, mentre la possibilità di ottenere specifiche composizioni o di controllare le caratteristiche qualitative (inclusioni non metalliche, gas) è affidata a specifiche fasi di lavorazione, in primo luogo alla metallurgia fuori forno, cioè all'elaborazione (affinazione) dell'acciaio condotta allo stato liquido in siviera<sup>3</sup> o in specifici impianti.

Le fasi di produzione si possono così schematizzare:

1. la materia prima viene scaricata e depositata nel parco rottame e suddivisa in categorie omogenee. Le materie prime, insieme a materiali ausiliari, ricicli interni e materiali energetici, vengono prelevate dallo stoccaggio e inserite in ceste e trasferite nell'area forno;
2. il materiale, approvvigionato tramite alcune ceste in successione, viene caricato nel forno. Il rottame viene fuso mediante arco elettrico. Una volta fuso il rottame, viene effettuata la scorifica della parte fusa non metallica e si conduce una breve affinazione destinata a una prima messa a punto della composizione, tramite aggiunte di materiali ausiliari (calce, ecc.) e ferroleghie approvvigionati direttamente al forno;
3. il forno viene svuotato (in realtà viene mantenuto un piede liquido per facilitare la ripresa del processo successivo e ridurre il trascinarsi di scoria) travasando l'acciaio in una siviera, dove vengono condotte ulteriori lavorazioni di affinazione della composizione. Vengono aggiunte ferroleghie e i materiali ausiliari necessari al risultato chimico e metallurgico richiesto, vengono insufflati gas inerti per il controllo di temperatura e inclusioni;
4. la siviera viene trasferita all'impianto di solidificazione: la siviera viene posizionata sopra un impianto di distribuzione dell'acciaio (paniera) a diverse unità parallele di solidificazione (lingottiere), dove si impone la sezione esterna del semilavorato e si solidifica una pelle esterna

---

<sup>3</sup> La siviera è un recipiente usato per il trasporto del metallo fuso dal forno al punto di colata

tramite importante scambio termico realizzato all'interno della lingottiera da circolazione di acqua. L'acciaio in solidificazione, estratto dalla lingottiera con continuità (colata continua), trascinato e raffreddato tramite spruzzi di acqua, viene tagliato a caldo nelle lunghezze richieste.

5. in alternativa la solidificazione avviene travasando l'acciaio liquido all'interno di contenitori statici di solidificazione (lingottiere). Al termine della solidificazione le lingottiere vengono sfilate e i lingotti avviati alle destinazioni successive.

Le lavorazioni di un'acciaieria elettrica sono schematizzate nella Figura

1.1. Si è ritenuto importante distinguere due gruppi di lavorazioni:

- a sinistra sono allineate le lavorazioni realizzate in sequenza destinate a trasformare la materia prima in semilavorati: queste attività sono riferite sostanzialmente al normale funzionamento degli impianti e a tutte le operazioni ausiliarie connesse, che sono visualizzate traslate a destra nel flusso;
- allineate a destra al di fuori della sequenza ordinata sono indicate le fasi di lavorazione e le operazioni comunque condotte nel comparto e che possono coinvolgere in misura più o meno impegnativa l'intero ciclo di lavorazione: è possibile osservare che queste attività si riferiscono principalmente a eventi non continuativi (realizzati spesso fisicamente in altro luogo rispetto alla posizione degli impianti di processo), interventi in alcuni casi implicati da malfunzionamenti e/o incidenti della normale attività di lavorazione.

Fondamentalmente, la produzione dell'acciaio avviene attraverso uno dei seguenti tre processi:

- ciclo integrale altoforno – convertitore;
- forno elettrico ad arco;
- riduzione diretta e rifusione dei preridotti.

La scoperta dell'arco che scocca tra due elettrodi di carbonio risale al 1800 per opera di H. Davy ed il primo tentativo di fondere il ferro sfruttando il ca-

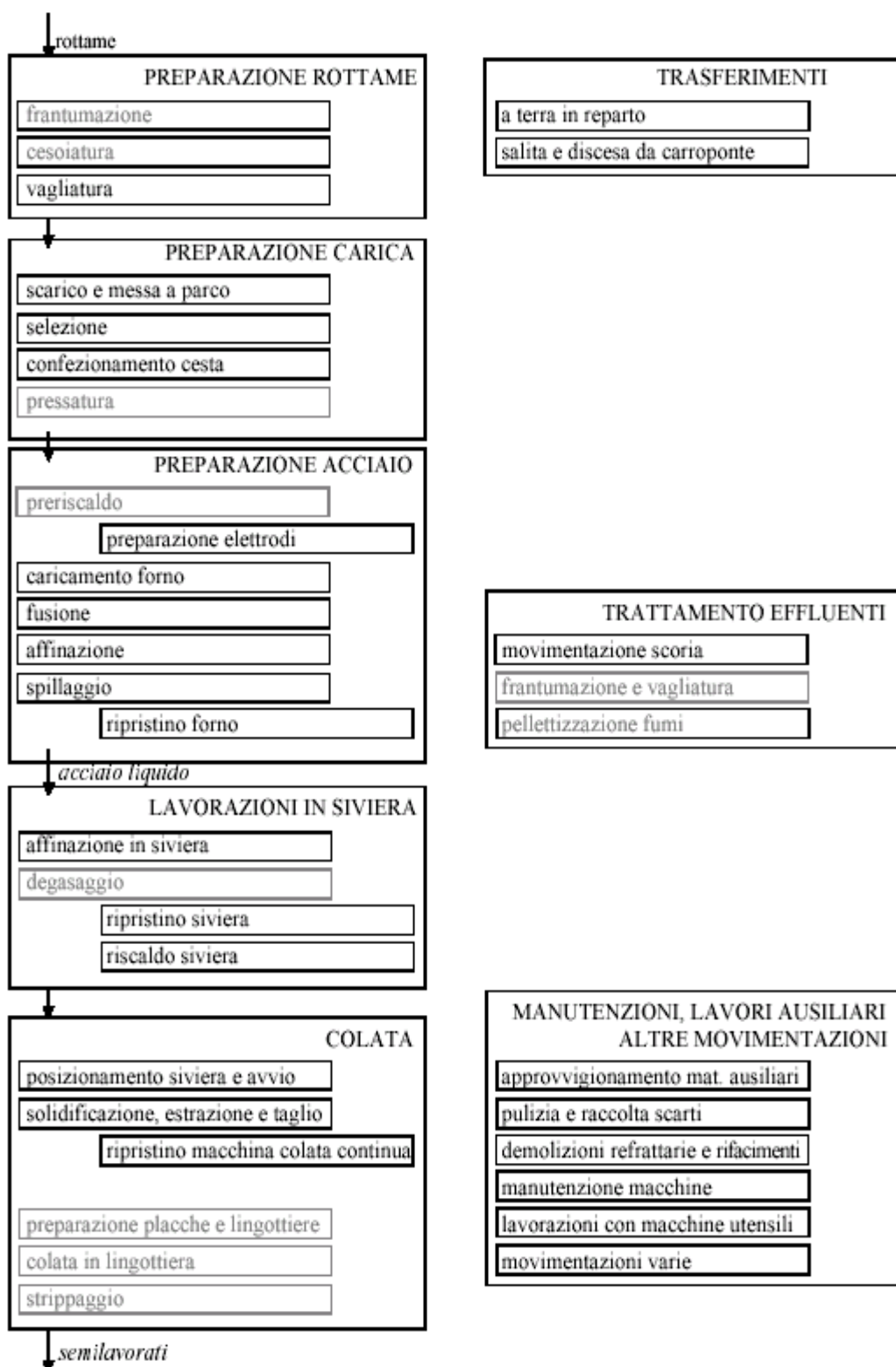


Figura 1.1 - Schema a blocchi acciaieria: principali fasi di lavorazione e relazione tra le fasi.

In grigio sono indicate le fasi di lavorazione presenti solo in alcune attività produttive

lore che l'arco irradia ad altissima temperatura fu operato da Siemens nel 1878. All'inizio del secolo scorso (1900), Héroult in Francia, Stassano in Italia e Kjellin in Svezia idearono e costruirono tre diversi tipi di forni elettrici per produrre acciaio su scala industriale. Mentre il forno di Héroult rappresenta il primo esemplare di forno ad arco diretto, nel quale anche la carica è attraversata dalla corrente elettrica, quello di Stassano fu il prototipo dei forni ad arco indiretto o radiante ed il Kjellin di quelli ad induzione.

I forni ad induzione sono costituiti da un avvolgimento primario, che viene alimentato dalla rete, e da un circuito secondario, rappresentato dalla carica metallica che si vuole fondere, in cui avviene la trasformazione dell'energia elettrica in calore, per effetto Joule di resistenza alle correnti indotte. L'energia è trasmessa dal primario al secondario per induzione elettromagnetica. I forni ad induzione sono largamente usati nella produzione delle leghe non ferrose e delle ghise speciali, ma anche per la fabbricazione di acciai speciali ad alto tenore di lega.

Oggi, il rottame di ferro viene rifuso nei cosiddetti *minimill* dotati di forno elettrico ad arco diretto (Electric Arc Furnace o EAF). Di seguito vengono spiegati solo questi impianti, in quanto l'acciaieria in analisi utilizza principalmente il rottame per la produzione dell'acciaio.

Gli EAF sono caratterizzati da ridotti costi di investimento e maggiori costi di esercizio, con capacità produttive dell'ordine delle centinaia di migliaia di tonnellate annue. L'alimentazione del forno elettrico è costituita principalmente da rottame ferroso di varia qualità. Il mezzo fusorio è essenzialmente costituito da un forno trifase equipaggiato con tre elettrodi in grafite: l'arco scocca tra l'estremità di ciascuno dei tre elettrodi e la carica metallica, trasformando l'energia elettrica in calore in uno spazio limitato raggiungendo temperature superiori ai 2000° C. L'EAF non è usato per estrarre ferro dal minerale ossidato, ed ha una limitata capacità di affinazione. In generale gli impianti che utilizzano l'EAF sono costituiti da un parco rottami, una sezione di preparazione della carica (cesoiatura, pressatura) nonché di preriscaldamento della carica (realizzato con

l'utilizzo di parte del calore sensibile dei fumi), dalla sezione del forno, dalla linea di trattamento dei fumi captati dal forno e dagli impianti ausiliari. Completano l'impianto le linee di trasformazione dell'acciaio in prodotti finiti piani o lunghi, con prevalenza di questi ultimi. Gli impianti basati sull'EAF sono di taglia generalmente medio bassa, l'unità media è costituita da un forno avente una capacità di bacino di circa 70 ton di acciaio liquido, la produzione media annuale si aggira sulle 400.000÷500.000 ton/anno e gli investimenti sono molto minori di quelli di un impianto a ciclo integrato.

Attualmente la tecnologia dell'arco elettrico ha raggiunto la piena maturità industriale e non sono quindi prevedibili ulteriori evoluzioni che possano influire o modificare radicalmente l'attuale quadro d'interazioni ambientali. Nella Figura 1.2 sono riportate, in forma grafica, le tappe salienti della più recente evoluzione del forno ad arco.

Un'acciaiera elettrica può essere suddivisa, per comodità descrittiva, nelle varie sezioni che costituiscono il complesso industriale che la ospita.

La Figura 1.3 ne schematizza le più importanti. Come si può osservare, procedendo da destra verso sinistra, si incontrano il parco rottami, la campata forni e la fossa di colata. In un'altra sezione, non visibile in figura, si effettuano sia le operazioni ausiliarie, quali la preparazione delle lingottiere nel caso in cui non sia presente un impianto di colata continua,

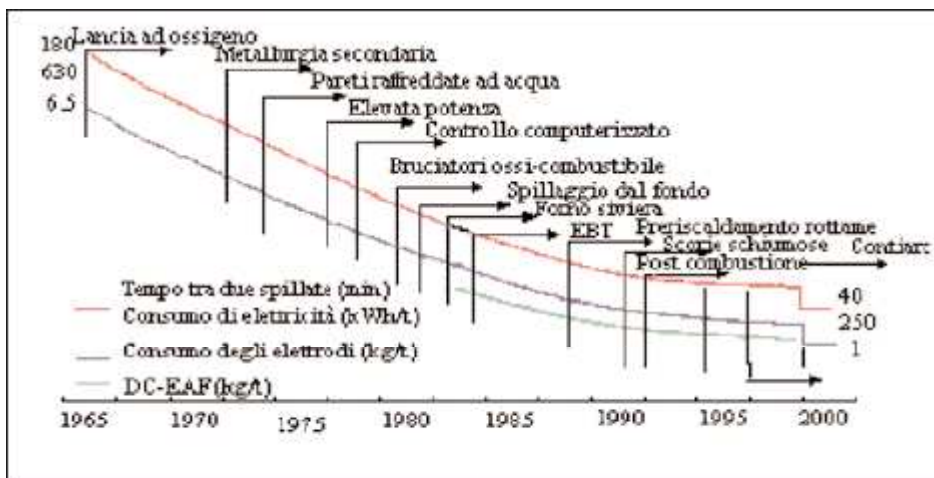


Figura 1.2 - Sviluppo dei forni elettrici ad arco

sia l'immagazzinamento dei refrattari, delle ferroleghie ed eventualmente degli elettrodi. Il parco rottame è servito da carroporti che selezionano il materiale e lo trasferiscono entro ceste di carica con fondo apribile, che possono avere un volume pari anche a 80% di quello del forno. Impianti di cesoiatura e pressatura consentono di tagliare i rottami e di pressare nella cesta quelli di bassa densità apparente: quest'ultima operazione permette di diminuire il numero complessivo di ceste e dunque le dispersioni termiche, e le emissioni, determinate dall'apertura della volta nella fase di caricamento. Vicino al forno vi è la cabina elettrica che contiene il trasformatore ed i meccanismi per la regolazione automatica del movimento degli elettrodi.

Ogni forno è inoltre dotato di un sistema di captazione ed abbattimento fumi e di una presa per l'ossigeno che viene ormai usato in tutte le acciaierie.

I forni elettrici ad arco diretto sono costituiti da un crogiolo cilindrico ad asse verticale composto da un fasciame in lamiera interamente rivestito in refrattario il cui fondo ha la forma di una calotta sferica.

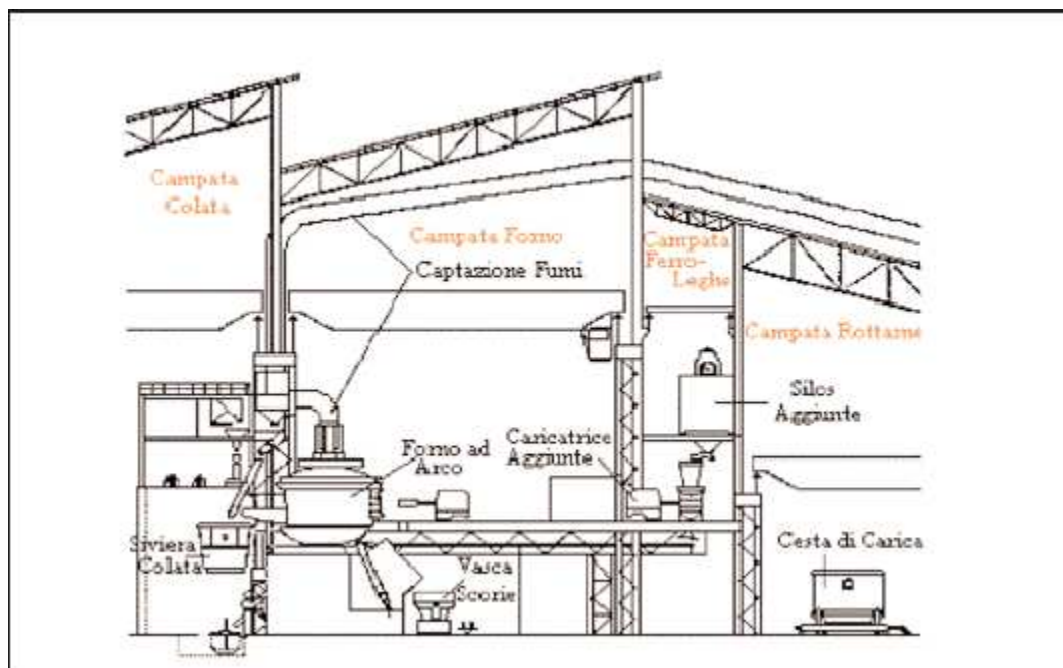


Figura 1.3 - Sezione semplificata di un'acciaieria elettrica con forno trifase da 150 t

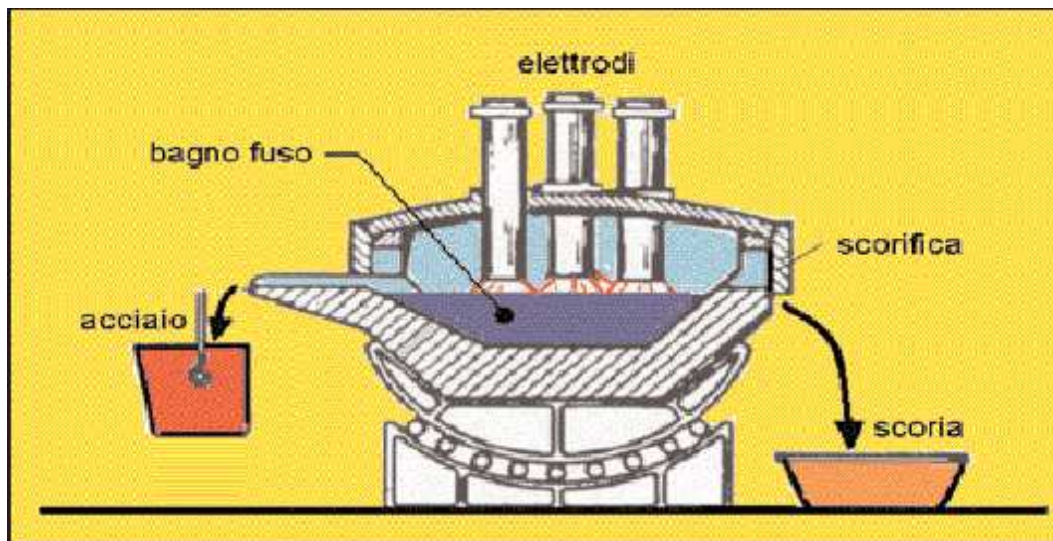


Figura 1.4 - Rappresentazione schematica della sezione di un forno elettrico ad arco diretto

Esso è coperto da una volta mobile che ne consente il caricamento rapido dall'alto mediante le ceste di carica. Il crogiolo ha la possibilità di basculare ruotando attorno ad un asse orizzontale: tale movimento consente sia di versare lentamente (*spillare*) la colata versandola, attraverso un foro ed un canale di colata in siviera, sia di agevolare le operazioni di scorifica effettuate attraverso delle apposite porte.

I forni ad arco sono generalmente di tipo trifase e sono quindi muniti di tre elettrodi di grafite, come mostra la Figura 1.4, che penetrano nel crogiolo attraverso tre fori praticati nella volta e disposti secondo i vertici di un triangolo equilatero. Gli elettrodi sono serrati da morse conduttrici collegate elettricamente al secondario del trasformatore. A questi tre elettrodi, nella volta del forno, si aggiunge un ulteriore foro (quarto foro) da cui vengono aspirati i fumi che si producono nella fase di fusione a forno chiuso.

L'arco scocca tra l'estremità di ciascuno dei tre elettrodi e la carica metallica. Questa è percorsa dalla corrente ed il circuito secondario si chiude attraverso la carica. L'energia elettrica viene trasformata in calore per effetto Joule ed il calore trasmesso al resto della carica principalmente per irraggiamento.



I moderni forni possono essere equipaggiati con lance super-soniche, che sono dei bruciatori ausiliari, sistemi di post combustione dei fumi, mentre il rottame può essere preriscaldato (soluzione non molto diffusa in Italia) facendo uso proprio del calore sensibile dei fumi. In definitiva l'apporto energetico elettrico viene integrato in maniera sempre più significativa da altri non elettrici, ed i fumi sono divenuti parte integrante del processo fusorio.

### **1.2 - La captazione dei fumi<sup>4</sup>**

Per proteggere l'ambiente all'interno e nei dintorni dell'acciaieria elettrica occorre intervenire sui fumi prodotti.

Le emissioni di un forno elettrico vengono usualmente distinte in:

- primarie le emissioni che si generano durante le fasi di fusione del rottame e di affinazione del bagno condotte a forno chiuso;
- secondarie le emissioni che si generano durante periodi relativamente brevi, cioè le fasi di caricamento del rottame condotte a forno aperto e durante lo spillaggio in siviera, ma caratterizzate da importanti volumi. Nelle emissioni secondarie confluiscono anche le primarie che sfuggono attraverso le aperture (alveoli degli elettrodi, anello fra tino e volta, porta) alla captazione che estrae direttamente dal volume del forno.

La caratterizzazione dei fumi, e delle polveri che da essi si ottengono, è strettamente dipendente da diversi fattori quali:

- le caratteristiche della carica: la carica risulta sostanzialmente "pulita" quando è costituita da scarti dei processi siderurgici e semilavorati, mentre all'estremo opposto si situa la carica contenente composti clorurati, batterie, torniture, oli e grassi, ecc.;
- la tipologia del forno;

---

<sup>4</sup> Vedi [2] e [3]

- il tipo di acciaio prodotto.

Il volume e la temperatura dei fumi risultano, dunque, notevolmente influenzati dalla particolare combinazione dei fattori sopra menzionati. Per quanto attiene al volume, fattori di incremento durante la fusione sono costituiti da:

- gli apporti di ossigeno che vengono a loro volta bilanciati dalle tecniche di post combustione;
- l'incremento delle potenza elettrica che determina una più rapida combustione delle sostanze organiche presenti (oli, plastiche, ecc.) ed un più rapido innalzamento delle temperature dei gas sviluppati;
- il preriscaldamento della carica contribuisce invece ad una loro riduzione.

Per le fasi di carica e di spillaggio, l'ottimizzazione dei tempi, l'elevata carburazione, ecc. portano a notevoli aumenti dei volumi delle emissioni.

Il contenuto di polveri nei fumi è a sua volta condizionato da diversi fattori quali i contributi di ossigeno e le tecniche di preriscaldamento ma, in generale, li si può stimare in circa 1-2 % del peso della carica solida.

Lo schema base è costituito da:

- sistema di captazione (cappe di aspirazione, dog house);
- condotte di convogliamento dei fumi;
- gruppo ventilatori di estrazione;
- sistema di filtraggio costituito attualmente, nella grande maggioranza dei casi, da batterie di filtri a manica;
- eventuale trattamento delle polveri (sinterizzazione).

Il sistema di estrazione, convogliamento ed abbattimento dei fumi, provenienti dal quarto foro, è dimensionato al fine di ottimizzare i diversi fattori che condizionano il funzionamento del forno:

- la velocità di estrazione dei fumi, e quindi la pressione del forno influenza non solo l'efficienza termica del forno ma anche il consumo degli elettrodi. Una velocità troppo elevata raffredda il forno ed aumenta il consumo degli elettrodi, mentre una velocità troppo bassa

non permette una combustione completa dei gas combustibili prodotti dal processo di acciaieria ed il loro raffreddamento;

- l'ottimizzazione della combustione rende necessaria un'apposita camera (che garantisca il miscelamento ottimale e la combustione in condizioni di sicurezza) da collocarsi in prossimità del forno; nella fase iniziale della fusione ed in presenza di rottami ricchi di grassi, si possono sviluppare notevoli quantità di gas combustibili che rendono necessaria l'installazione di bruciatori pilota;
- il sistema di captazione ed abbattimento dei fumi deve inoltre proteggere il forno dal rischio di esplosioni che si presenta, in assenza di uno stadio di pre-combustione, quando scintille o materiale caldo vengono a contatto con piccole quantità di gas rimaste nei condotti;
- una particolare condizione di pericolo si genera quando si miscelano, prima dell'ingresso nell'unità di abbattimento, i gas di estrazione diretta provenienti da due forni diversi (gas ricchi di ossigeno, provenienti da uno, incontrano gas ricchi di combustibile, provenienti dall'altro);
- i condotti devono essere dimensionati per ridurre le resistenze al moto ed ottenere velocità dell'ordine della decina di m/s, necessaria per evitare la deposizione delle polveri nei condotti.

Il controllo dell'inquinamento aerodisperso va condotto con soluzioni tecnologiche e impiantistiche che tengono conto delle modalità di emissione, che si determinano in particolare nelle fasi di carica, nelle fasi di fusione interessate da regime non regolare e in misura minore nella fase di spillaggio.

Nella realtà industriale si osserva un sistema consolidato di captazione delle emissioni primarie, realizzato tramite quarto foro e l'inserimento generalizzato di sistemi di abbattimento tramite filtrazione.

Con la captazione tramite quarto foro le emissioni primarie vengono prelevate direttamente dal volume del forno, tramite un condotto di aspirazione connesso alla volta del forno. Questo sistema risulta efficace esclusivamente per le fasi di lavoro condotte a forno chiuso.

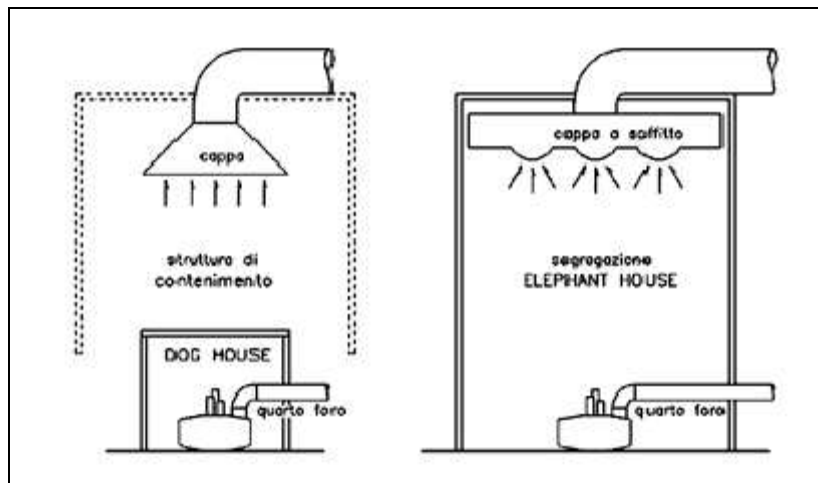


Figura 1.5 - Esempi dei sistemi di captazione delle emissioni primarie e secondarie negli EAF

In integrazione alla captazione tramite quarto foro è stata sviluppata una configurazione che prevede una cappa di aspirazione (aspirazione a pressione bilanciata) avvolgente la volta del forno, tale da formare un'adeguata intercapedine.

Le emissioni secondarie, invece, vengono captate tramite una cappa posta sopra il forno. La cappa può essere realizzata con aspirazione concentrata (Figura 1.5 a sinistra), prevedendo una configurazione con falde inclinate a 45° per garantire il convogliamento dei fumi a un unico condotto aspirante, oppure con aspirazione distribuita (Figura 1.5 a destra) su un'ampia superficie e collettori a prese multiple sul soffitto.

A causa della collocazione elevata della cappa in relazione alla sorgente di emissione, in particolare con significative correnti d'aria nel reparto, parte dei fumi può scavalcare la cappa e fuoriuscire all'esterno attraverso le aperture del fabbricato; la possibilità di evitare fuoriuscite di fumi non può essere garantita dal volume della cappa, insufficiente a trattenere al suo interno il volume di fumi sviluppato. La cappa deve quindi essere affiancata e integrata da sistemi di contenimento, destinati a trattenere le emissioni nell'attesa della loro completa estrazione.

I sistemi di contenimento sono di due tipi:

- incapsulamento del forno (dog-house). Questa soluzione prevede una struttura collocata a ridosso del forno. La dog-house racchiude la

sorgente dei fumi durante la fusione e riduce drasticamente il volume da controllare con l'aspirazione in questa fase. Durante lo spillaggio e l'affinazione la dog-house rimane sostanzialmente adeguata per aspirare le emissioni. Dovendo aprire la sezione superiore per la carica del rottame e per operazioni di rifacimento refrattario, la dog-house perde la capacità di contenere i fumi all'interno della segregazione.

- segregazione del reparto (hippo-house, elephant-house). Le segregazioni intercettano l'area di fusione in misura più o meno ampia: comportano la sigillatura delle aperture della copertura e implicano la necessità di mantenere chiuso il volume controllato, considerando che aperture, comprese quelle nelle posizioni elevate in corrispondenza al transito dei carriponte, penalizzano in misura rilevante la cattura degli inquinanti. Va quindi prevista la completa sigillatura del volume da controllare mediante aspirazione e l'installazione di pareti fisse e portelloni motorizzati, che consentano il transito dei materiali in ingresso e in uscita dal forno e del forno da avviare a rifacimento.

Quindi le emissioni di un forno elettrico vengano captate prima dell'invio in ciminiera. La ragione di tale pratica sta nel fatto che i fumi di acciaieria elettrica contengono una grande varietà di elementi, dai metalli pesanti a quelli leggeri, come zinco ed alluminio, dal cadmio all'arsenico. Questi elementi provengono dal rottame usato che non è soltanto acciaio zincato (di qui lo zinco e tutti i metalli pesanti o leggeri, lo zolfo, il silicio, il fosforo etc.), ma anche verniciato (di qui tutti gli elementi che entrano nella produzione delle vernici). Ci sono, inoltre, tutti quegli elementi che provengono anche dai materiali che solitamente accompagnano i rottami stessi, come plastiche ed altro, che non vengono separati dal rottame. Durante il processo fusorio, data la temperatura elevatissima che si realizza, passano in fase gassosa non solo i bassofondenti, ma anche il ferro stesso ed i metalli ad alta temperatura di fusione che vengono trascinati in fase vapore dalla forza di ebollizione della massa fusa. Tutti questi elementi si ritrovano, dunque, nei fumi che non possono essere inviati direttamente al camino, ma devono essere abbattuti, dando origine

Tabella 1.1 - Composizione tipica (%) delle polveri emesse dal forno fusorio

	<b>Acciai al carbonio (%)</b>	<b>Acciai inossidabili (%)</b>
Fe <sub>2</sub> O <sub>3</sub>	50,0	43,0
ZnO	18-35	2-3
Cr <sub>2</sub> O <sub>3</sub>	0,5	19,0
NiO	0,2	6,0
MnO	12,0	6,0
CaO	12,0	8,0
SiO <sub>2</sub>	9,0	7,0
MgO	8,0	6,0
Al <sub>2</sub> O <sub>3</sub>	1,0	0,5
Na <sub>2</sub> O	2,0	1,5
K <sub>2</sub> O	1,0	0,5
Ce	1,5	1,5
F	0,5	0,5
C	1,0	0,5
S	1,0	1,0
Pb	2-7	
Cd	0,03-0,1	
Cu	0,2-0,5	
As	0,06-0,2	

alle cosiddette polveri di acciaieria. Si calcola che la produzione di polveri in un forno elettrico oscilla tra 10 e 20 kg per tonnellata di acciaio prodotto. Le polveri, dunque, sono il vero problema delle acciaierie, in particolare quelle elettriche, e la loro gestione si presenta ancora alquanto problematica.

La Tabella 1.1 presenta una composizione tipica delle polveri di acciaieria elettrica.

Innanzitutto si osserva, in generale, che gli elementi sono presenti nelle polveri allo stato di ossidi in quanto la presenza di ossigeno nell'atmosfera del forno e l'elevata temperatura ne favoriscono l'ossidazione immediata.

Si deve notare l'elevato tenore di zinco (pari a circa 50.000-98.000 t/anno di Zn) e di piombo. La presenza di Zn e Pb non rende possibile l'uso delle polveri nemmeno come carica in altoforno, dove invece l'ossido di ferro può essere ridotto a ferro metallico e, quindi, fuso. Ciò in quanto lo Zn ed il Pb finirebbero col depositarsi sui refrattari degli altiforni, distruggendoli progressivamente. Tuttavia, questi quantitativi così elevati di Zn rendono

in qualche modo appetibili le polveri per il recupero di tale elemento al pari di quanto viene oggi fatto per le polveri provenienti dagli acciai speciali che, essendo ricche in cromo e nichel in particolare, vengono usate come materia prima secondaria.

Circa il piombo, si stima che nelle polveri di acciaieria prodotte in un anno in Italia ci siano almeno 7.000-10.000 t. Tanto piombo non può certamente provenire dagli acciai cosiddetti al piombo, i quali non potrebbero fornire più di 100 t di Pb. La ragione della presenza di piombo deve essere ricercata altrove. Si ritiene che il piombo contenuto nelle polveri di acciaieria provenga quasi tutto dalle batterie non ritirate dal COBAT (Consorzio Obbligatorio Batterie al Piombo Esauste e Rifiuti Piombosi, istituito con la legge 475/88) e vendute alle acciaierie elettriche assieme ai rottami delle automobili, dalle quali non sono state preventivamente tolte e consegnate al COBAT stesso.

A valle del sistema di abbattimento dei fumi, il contenuto di polveri si riduce notevolmente: dati comunitari, lo individuano nell'intervallo 5-10 g/tonnellata di acciaio prodotto.

I fumi inviati al camino trasportano, inoltre, i seguenti composti, rilevanti ai fini dell'inquinamento atmosferico:

- monossido di carbonio (CO e CO<sub>2</sub>);
- ossidi di azoto (NO<sub>x</sub>);
- composti organici volatili (VOC)

In Tabella 1.2 sono riportati dati indicativi del fattore di emissione (grammi di inquinante per tonnellata di acciaio liquido) misurato al camino:

Tabella 1.2 - Emissioni di CO, NO<sub>x</sub>, SO<sub>x</sub> e VOC

<b>Inquinante</b>	<b>Fattore di emissione [g/t]</b>
CO	300-900
NO <sub>x</sub>	100-150
SO <sub>x</sub>	10-30
VOC	100-200

### **1.3 - Lo stabilimento in esame<sup>5</sup>**

L'impianto in esame è fornito di due forni, realizzati in momenti successivi. Il progetto del Forno 1 prevedeva una elephant-house. La segregazione fu però completata soltanto 2 anni dopo l'avvio del forno; l'incompletezza del sistema di captazione secondario ha causato sensibili problemi di emissioni in atmosfera. Inoltre dal progetto iniziale il forno si discosta nella mancanza dell'impianto di preriscaldamento dei rottami (utile per la riduzione della quantità di fumi prodotti durante la fase di carica e di fusione).

Nel Forno 1 il colaggio è continuo. In questo caso la siviera è posta su una torretta girevole e l'acciaio viene colato in un contenitore (paniera) dal quale cola in continuo in lingottiere di rame raffreddate ad acqua.

Si stima che le polveri prodotte siano circa 22 Kg/ora (5% tra emissioni secondarie ed emissioni primarie non captate), che si disperdono nell'ambiente senza trattamenti, mentre dal camino, dopo la depurazione nei filtri a maniche, escono circa 9 Kg/ora di polveri (emissioni primarie captate emesse ed emissioni secondarie captate emesse). Il fattore di emissione complessivo è stimato pari a 0.24 Kg/t acciaio prodotto in linea (vedi Tabella 1.4).

Il Forno 2, invece, inizialmente fu dotato di una dog-house, che però presenta dei limiti nelle fasi di caricamento e durante lo spillaggio, cioè nelle fasi in cui la segregazione viene aperta. A vista di questi limiti, dopo alcuni anni di esercizio venne integrata l'aspirazione secondaria, installando una cappa nella parte alta del capannone con l'obiettivo di catturare i fumi non captati dalla dog-house.

---

<sup>5</sup> Si vedano [4] e [5]



Tabella 1.3 - Stime emissioni forno Forno 1

<b>Emissioni</b> 2286 Kg/h	<b>FORNO 1</b>	
<b>Emissioni primarie</b>	90% del tot e.	
e. p. captate	90% delle e.p.	
e.p.c.abbattute		99.6% delle e.p.c.
e.p.c.emesse		0.4% delle e.p.c.
e. p. non captate	10% delle e.p.	
<b>Emissioni secondarie</b>	10% del tot e.	
e.s. + e.p. nn c.	ca. 19% del tot e.	
e.s. captate	95% delle e.s+e.p.nc.	
e.s.c.abbattute		99.6% delle e.s.c.
e.s.c.emesse		0.4% delle e.s.c.
e. diffuse	5% delle e.s+e.p.nc.	

Nel Forno 2, lo spillaggio consiste nel rovesciamento dell'acciaio contenuto nel forno nella siviera, dove vengono aggiunte le ferroleghie ed i materiali necessari al bilanciamento chimico richiesto. Durante il 2003 nel forno vengono effettuate delle opere di adeguamento a cause delle elevate emissioni e sempre durante questo anno vengono completate alcune opere di sigillatura e tamponatura della nuova struttura per evitare la dispersione dei fumi catturati dalla segregazione. Le polveri prodotte si stimano in circa 13 Kg/ora (5% tra emissioni secondarie ed emissioni primarie non captate), che si disperdono nell'ambiente senza trattamenti mentre dal camino, dopo la depurazione nei filtri a maniche, escono circa 7 Kg/ora di polveri (emissioni primarie captate emesse ed emissioni secondarie captate emesse). Il fattore di emissione complessivo è stimato pari a 25 Kg/t acciaio prodotto in linea (vedi Tabella 1.3).

Il materiale che l'azienda utilizza per produrre l'acciaio è principalmente costituito da rottame e ghisa, acquistati sul mercato nazionale ed estero. A queste vanno ad aggiungersi i recuperi interni costituiti da spuntature, lingotti, colonne e radici di placche, parti di semilavorati scarti e colaticci. Tutto il materiale è suddiviso e depositato in aree prestabilite del parco rottami.

Tabella 1.4 - Stima emissioni del Forno 2

Emissioni 1400 Kg/h	<b>FORNO 2</b>	
<b>Emissioni primarie</b>	90% del tot e.	
e. p. captate	90% delle e.p.	
e.p.c.abbattute		99.5% delle e.p.c.
e.p.c.emesse		0.5% delle e.p.c.
e. p. non captate	10% delle e.p.	
<b>Emissioni secondarie</b>	10% del tot e.	
e.s. + e.p. nn c.	ca. 19% del tot e.	
e.s. captate	95% delle e.s+e.p.nc.	
e.s.c.abbattute		99.5% delle e.s.c.
e.s.c.emesse		0.5% delle e.s.c.
e. diffuse	5% delle e.s+e.p.nc.	

La preparazione della carica avviene attraverso il riempimento di ceste per un peso totale di 95 tonnellate. La carica del rottame all'interno delle ceste è eseguita mediante carroponi con magneti; nel caso di cariche destinate alla produzione di acciaio inox è utilizzato una benna *a polipo*.

L'aggiunta di additivi avviene tramite nastri trasportatori aerei e coperti.

### 1.4 - Il monitoraggio<sup>6</sup>

Considerata l'importanza quantitativa e qualitativa delle emissioni di un forno elettrico ad arco, si è ritenuto essenziale, da parte degli organi di controllo, disporre di un sistema di misura delle ricadute nell'intorno della sorgente.

Vengono qui indicate le problematiche che si era inteso affrontare nel predisporre un protocollo di verifica (Tabella 1.5):

- individuazione dei parametri da rilevare all'esterno dell'attività produttiva per valutare le ricadute secondo criteri di tossicità e di specificità rispetto alla sorgente, tenuta presente la disponibilità di tecniche analitiche per i livelli di inquinamento attesi. Nell'esperienza di

<sup>6</sup> Vedi [2]

riferimento, determinazione dei metalli nelle polveri inalabili: Zinco, Ferro, Manganese, Piombo e Cadmio. Con l'evoluzione delle tecniche analitiche disponibili, si sono successivamente considerate le ipotesi di determinare anche i metalli nelle PM10 e i microinquinanti organoclorurati;

- individuazione dei criteri di riferimento per la valutazione di accettabilità per parametri non normati. Nell'esperienza di riferimento sono stati scelti i valori guida WHO 1997 e poi 2000 per il Manganese e il Cadmio, i limiti di legge italiana per il piombo. Per Ferro e Zinco, considerata la limitata tossicità, i valori misurati sono utili ad individuare i contributi della sorgente specifica e le variazioni nel tempo delle emissioni in funzione dell'evoluzione impiantistica. Per il Piombo va considerato anche il contributo del traffico veicolare e la progressiva eliminazione di tale elemento dai carburanti;
- localizzazione dei punti di misura, tenute presenti le caratteristiche meteorologiche dell'area (venti prevalenti, intensità e direzione), la presenza di abitati e la disponibilità di supporti tecnici;
- scelta del numero dei punti di misura, considerata l'onerosità delle misure in funzione delle risorse disponibili;
- individuazione della durata delle campagne di misura: qualora le risorse disponibili non consentano il rilievo in continuo, si è ritenuto possibile individuare sulla base degli andamenti meteorologici e delle caratteristiche dei processi produttivi, periodi di misura più brevi, ma rappresentativi;
- individuazione della durata del singolo campione in funzione del processo tecnologico e della sensibilità del metodo analitico e delle prescrizioni di norma.

Tabella 1.5 - Criteri di valutazione della qualità dell'aria

Contaminante	Valore limite	Tipo di concentrazione limite
Ferro, Zinco		L'OMS non propone valori guida; i valori rilevati forniscono indicazioni in ordine alla provenienza delle polveri
Manganese	OMS 2000: 0.15 µg/m <sup>3</sup>	"media annuale dei valori medi sulle 24 ore"; limite proposto dall'OMS nel 2000
Piombo	0.5 µg/m <sup>3</sup>	"media annuale" valore da conseguire entro il 01/01/2005. Direttiva 1999/30/CE del 22/04/99, recepita dal D.M. 60/2002 – coincide con il valore limite OMS. I valori esistenti non dovrebbero essere peggiorati.
Cadmio	OMS 2000: 0.15 µg/m <sup>3</sup>	"media annuale dei valori medi sulle 24 ore" limite proposto dall'OMS nel 2000 (non viene indicato un limite di sicurezza in relazione alla cancerogenicità accertata per l'uomo - gruppo 1 IARC)

### **1.5 - Dati di letteratura**

Sono state reperite in letteratura alcune pubblicazioni inerenti la composizione dei fumi prodotti da acciaierie elettriche [7-10]. Le analisi chimiche sono state effettuate su campioni ottenuti dal quarto foro (per le emissioni primarie) e dalle cappe di aspirazione (per le emissioni secondarie).

In generale, nei campioni la qualità delle polveri captate indica un'elevata variabilità dei contributi, con riferimento a tutte le fasi di produzione, ad eccezione della fase di affinazione, caratterizzata da un'emissione più specifica (qui le emissioni dipendono dagli elementi aggiunti nel metallo fuso).

Questa ricerca di informazioni sulla composizione delle polveri emesse è orientata al riconoscimento, nei dati risultanti dal monitoraggio presso l'impianto, dei contributi imputabili alle diverse fasi della lavorazione.

Relativamente al quantitativo di polveri emesse, in letteratura si riporta un'elevata variabilità dei fattori di emissione associati ai diversi sistemi di captazione, riassunti in Tabella 1.6 [8].

I fattori di emissione calcolati sono relativi prevalentemente al sistema primario e dipendono in modo rilevante dalla qualità, dalla pezzatura e dallo sporcamento del rottame. Dalle analisi è possibile assumere che il

sistema primario sia in grado di captare il 90% delle emissioni totali ed è possibile stimare le emissioni diffuse come differenza fra l'emissione totale e l'emissione intercettata (vedi le Tabelle 1.6 e 1.7).

Il "controllo garantito del sistema secondario" è un indice sintetico del sistema secondario, che è definito come rapporto fra le emissioni secondarie captate e le emissioni secondarie attese e rappresenta in modo riassuntivo il livello di controllo delle emissioni raggiunto dal sistema secondario. Dai dati si evince una differenza evidente delle prestazioni a seconda del tipo di segregazione.

Dalle analisi chimiche effettuate, si evince anche che la composizione delle emissioni è diversa. Le emissioni primarie sono costituite principalmente da franklinite ( $ZnOFe_2O_3$ ) [6]. Le emissioni secondarie, invece, sono costituite generalmente da polveri di varia natura, come pure di altre sostanze come oli, policlorobifenili e diossine (entrambe inquinanti organici persistenti), oltre a residui organici incombusti di cui erano contaminati i rottami caricati.

Tabella 1.6 - Polveri captate durante un intero ciclo di lavoro e fattori di emissione (in Kg/t acciaio prodotto)

	<b>Forno parzialmente segregato</b>	<b>Forno totalmente segregato</b>
Polvere captata da quarto foro (sistema primario)	10.9 +- 5.9 Kg/t acciaio	7.2 +- 2.4 Kg/t acciaio
Polvere captata dalla cappa (sistema secondario)	0.2 +-0.1 Kg/t acciaio	0.6 +- 0.1 Kg/t acciaio
Fattore di emissione intercettato	11.1 +- 0.6 Kg/t acciaio	7.8 +- 2.5 Kg/t acciaio

Tabella 1.7 - Stima delle emissioni diffuse e prestazioni del sistema secondario

	<b>Forno parzialmente segregato</b>	<b>Forno totalmente segregato</b>
Fattore di emissione atteso	12.1 Kg/t acciaio	8.0 Kg/t acciaio
- di cui da sistema primario	10.9 Kg/t	7.2 Kg/t
- di cui da sistema secondario	1.2 Kg/t	0.8 Kg/t
Emissioni diffuse	1.0 Kg/t acciaio	0.2 Kg/t acciaio
Controllo garantito dal sistema secondario	17%	75%

## CAPITOLO 1 - L'acciaio: produzione ed emissioni

In Tabella 1.8 [6] vengono riportati le quantità dei metalli di maggiore interesse, differenziando le polveri del quarto foro da quelle aspirate dalla cappa.

Tabella 1.8 - Parametri di composizione riscontrabili in polveri da acciaieria elettrica

Parametro	Unità di misura	Sistema primario		Sistema secondario	
		Valore minimo	Valore massimo	Valore minimo	Valore massimo
Ferro	%	17	30	9.7	12.4
Manganese	%	0.3	4.5	0.2	4.1
Piombo	%	1.5	5.8	4.3	5.9
Zinco	%	15	33	21	38
Cadmio	mg/Kg	20	900	3800	879
Cromo totale	mg/Kg			490	1200
Cromo (III)	mg/Kg	500	4000		
Cromo (VI)	mg/Kg	<1	13500	<1	90
Nichel	mg/Kg	100	4000	110	185

# CAPITOLO 2

## I dati

### ***2.1 - Introduzione***

Al fine di organizzare la descrizione, i dataset utilizzati si possono distinguere in tre: le variabili inerenti alla produzione dell'acciaio, le variabili meteorologiche ed infine le concentrazioni dei metalli osservati dai vari campionatori.

Per utilizzare questi dati ci si è appoggiati all'utilizzo del software R [12].

Durante una prima analisi si sono riscontrati vari problemi nella forma originale dei dati, tratti da dataset non pensati per un futuro utilizzo da parte dei vari utenti (record mancanti, diversi codici identificativi dei dati mancanti, diversi formati per le date), non utilizzabili direttamente in R. Si è quindi proceduto ad una prima pulitura ed inquadramento temporale di tutte le serie storiche (inizio 1 gennaio 2001 e fine 31 dicembre 2006) per i vari dataset. Questo lavoro ha comportato la scrittura di varie funzioni, implementate in R.

Di seguito verranno esposte tutte le tipologie di dati e gli eventuali adattamenti di pre-analisi.

## 2.2 - I dati di produzione

I dati di produzione dell'acciaio a disposizione vanno dal 1 luglio 2003 fino al 15 aprile 2006. Questi rappresentano il numero di colate e i kilogrammi di acciaio spillato al giorno, suddivisi per forno.

Per motivi di riservatezza i dati sono stati normalizzati per i rispettivi valori massimi riscontrati; per ciascun forno, dunque, sia per lo spillato che per il numero di colate, si sono ottenuti indici di attività compresi nell'intervallo [0;1]. Inoltre, sempre per garantire la riservatezza, ai forni ci si riferisce come a *Forno1* e *Forno2*.

In prima analisi si è proceduto a un controllo del dataset per identificare eventuali dati mancanti o ripetizioni all'interno di questo. Si è quindi controllato l'andamento dei giorni in cui ciascun forno è rimasto fermo (per varie manutenzioni, periodi di ferie o altro). Principalmente ci sono fermi di uno/due giorni, con frequenza settimanale, a parte il periodo di ferragosto in cui l'azienda chiude. Salta all'occhio, però, un blocco della produzione di ben 70 giorni per uno dei due forni nel periodo tra settembre e dicembre del 2003, presumibilmente dovuta a un'opera di adeguamento strutturale straordinario per l'azienda stessa.

Dopo questa prima analisi si è proceduto all'analisi descrittiva.

E' stata creata la variabile *funzione* (*fz*), che è una variabile qualitativa, che rappresenta il tipo di funzionamento dello stabilimento:

$$\text{funzione} = \begin{cases} 0 & \text{non lavora nessun forno} \\ 1 & \text{lavora solo il Forno1} \\ 2 & \text{lavora solo il Forno2} \\ 3 & \text{lavorano entrambe i forni} \end{cases}$$

All'ARPA è stato comunicato che pressochè settimanalmente viene effettuato un blocco, per dei controlli ed eventuali manutenzioni di routine; ciò può effettivamente essere riscontrato dai dati.

Normalmente le manutenzioni avvengono di giovedì o venerdì e lo si può vedere dai box-plot delle Figure 2.1 e 2.2, che indicano gli andamenti della



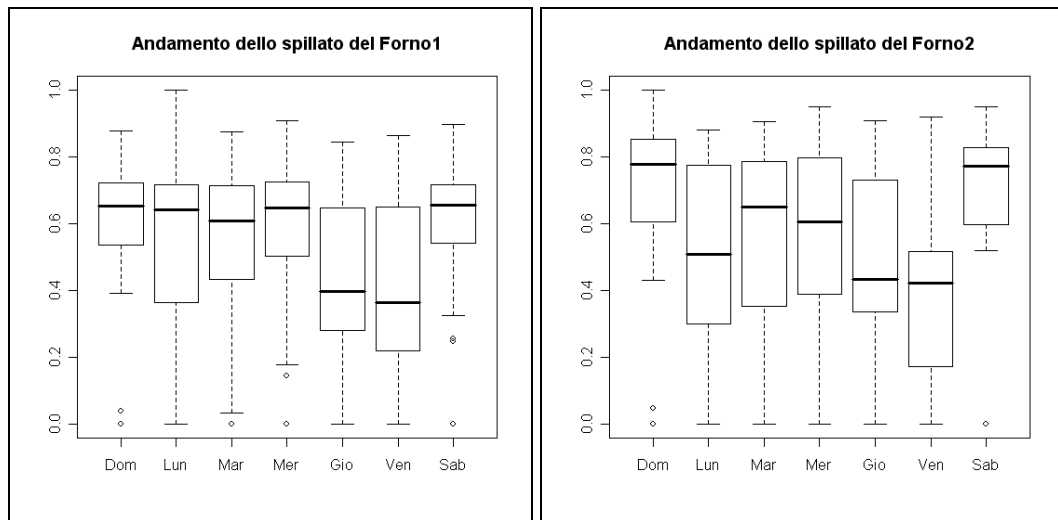


Figura 2.1 - Boxplot dello spillato del Forno1

Figura 2.2 - Boxplot dello spillato del Forno

produzione per giorno della settimana. Da questi si osserva anche che nel week-end l'azienda opera a pieno regime: molto probabilmente a causa dei costi minori dell'elettricità e di altri costi primari di fondamentale importanza per l'azienda.

La distribuzione dei quantitativi di acciaio spillato non appare essere gaussiana, come si può vedere nei valori di sintesi della variabile in Tabella 2.1 e dalle Figure 2.3 e 2.4.

Tabella 2.1 - Indici di sintesi delle variabili Spillato1 e Spillato2

Indice	Spillato1	Spillato2
<b>Min</b>	0.0000	0.0000
<b>1<sup>st</sup> Qu.</b>	0.3545	0.3400
<b>Median</b>	0.6104	0.6101
<b>Mean</b>	0.5082	0.5242
<b>3<sup>rd</sup> Qu.</b>	0.7116	0.8008
<b>Max</b>	1.0000	1.0000
<b>Asimmetria</b>	-0.3884	-0.26920
<b>Curtosi</b>	1.1353	0.89786

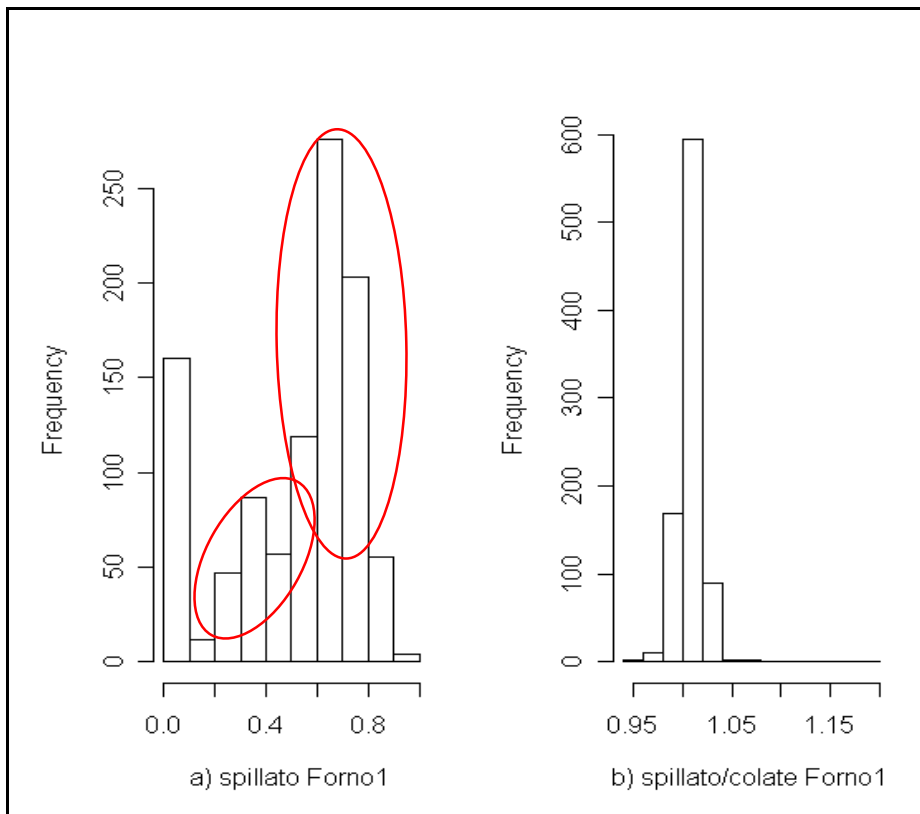


Figura 2.3 - Distribuzione dello spillato del Forno1

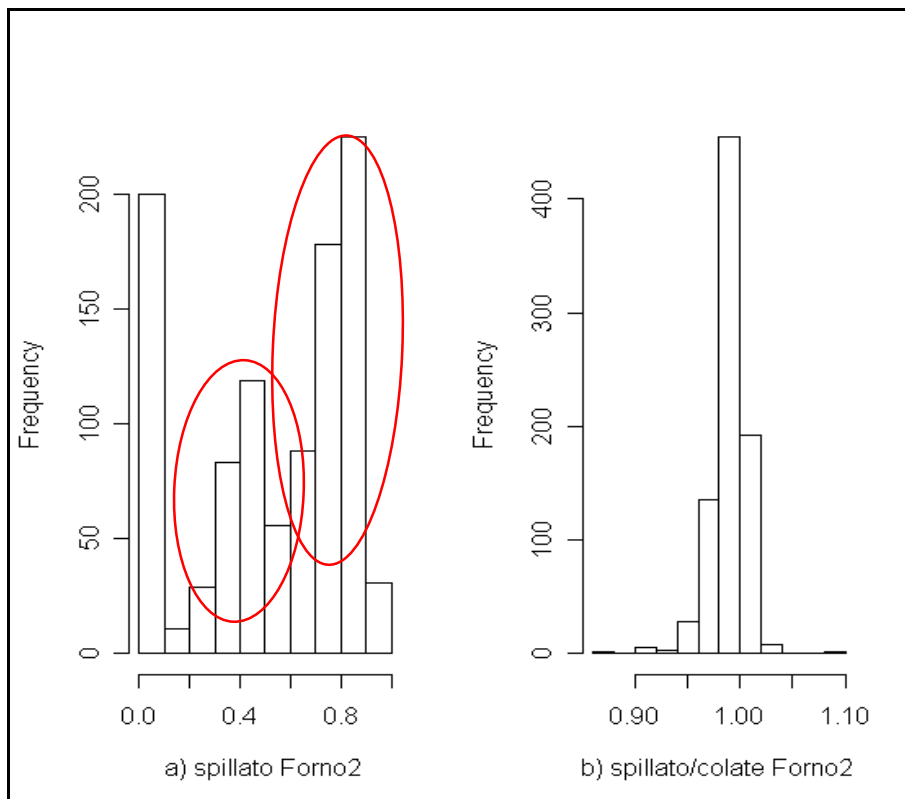


Figura 2.4 - Distribuzione dello spillato del Forno2

Osservando la Figura 2.3.a, lo spillato prodotto dal Forno1 sembra avere un andamento multimodale, cui si aggiunge il picco dei giorni di spegnimento. Ciò si riscontra anche se si effettua una distinzione tra i giorni in cui il forno lavora da solo ed i giorni in cui lavora insieme all'altro forno. Questo fenomeno in realtà è da imputare al fatto che i forni vengono spenti ogni settimana per un giorno per lavori di manutenzione ordinaria.

Può accadere che queste operazioni non si effettuino in una sola giornata, che lo spegnimento avvenga durante il giorno precedente quello previsto per la manutenzione ed il riavvio durante il giorno successivo: in corrispondenza di un giorno in cui la produzione risulta nulla, dunque, si hanno due giorni in cui la produzione risulta circa dimezzata rispetto al massimo regime. Quanto detto è confermato dall'andamento del rapporto tra la quantità di acciaio spillato e il numero di colate effettuate (Figura 2.3.b).

Per il Forno2 è riscontrabile lo stesso comportamento, osservabile in Figura 2.4.

Per entrambi i forni, la relazione tra la quantità di acciaio prodotto e il numero di colate è lineare ed è indipendente dalla modalità di funzionamento dello stabilimento.

È stata effettuata anche un'analisi sulle serie storiche degli spillati per controllare eventuali andamenti autoregressivi delle stesse.

Per lo *spillato1* si è giunti ad osservare che c'è la presenza di un comportamento autoregressivo del primo e del secondo ordine (non stagionale) accompagnato da un andamento autoregressivo stagionale del primo ordine (per stagione si intende la settimana).

Per lo *spillato2*, invece, si osserva solo un andamento autoregressivo del primo ordine a livello stagionale.

### **2.3 - I dati meteorologici**

I dataset dei dati atmosferici utilizzati sono due:

- i dati relativi al radiosondaggio di Campoformido, rappresentativo per tutto il Triveneto, cui si accede dal sito internet del National Climatic Data Center (NCDC);
- i dati meteorologici di una centralina meteorologica prossima al sito industriale, forniti dall'Osservatorio Meteorologico Regionale dell'ARPA (OsMeR).

Il radiosondaggio consiste nel lancio di un pallone aerostatico, munito di sensori meteorologici, destinato a salire indefinitamente in atmosfera fino a perdersi, trasmettendo i dati rilevati ad una stazione a terra.

L'utilizzo dei dati del radiosondaggio è stato il più problematico, soprattutto per la formattazione dei dati<sup>7</sup>. Si sono intesi utilizzare i dati corrispondenti al raggiungimento di una pressione pari a 8500 millibar<sup>8</sup> misurata dal pallone aerostatico. Le variabili estrapolate sono l'altezza in cui il radiosondaggio raggiunge la pressione indicata (*radio.height*), con rispettiva temperatura che si osserva a tale altezza (*radio.temp*) e la temperatura del punto di rugiada (*radio.dewpt*). Il lancio viene effettuato ogni giorno per un massimo di quattro lanci giornalieri: alle ore 5, alle 11, alle 17 ed alle 23 UTC (Coordinated Universal Time: nella nostra zona, un'ora in meno rispetto al Local Standard Time, LST, ora solare locale). Questo fa sì che ogni variabile si ripeta per quattro volte in un giorno. Nel caso in cui un lancio non venga effettuato si considera un dato non disponibile (codificato, in R, con la sigla NA).

Nell'ultimo anno, però, il numero di lanci è stato ridotto ai soli lanci delle ore 11 e delle ore 23 UTC. Per questo motivo nell'analisi sono state considerate solo le variabili di tali lanci.

Le variabili meteorologiche fornite dalla centralina dell'OsMeR sono:

- la temperatura, in gradi Celsius;

---

<sup>7</sup> Si veda in paragrafo A.1 dell'appendice

<sup>8</sup> La scelta è stata effettuata dal fatto che questa è una delle pressioni sempre misurate, al contrario di altre.

- la pioggia, in millimetri caduti;
- l'umidità relativa, in percentuale;
- la direzione istantanea del vento, a 10 metri dal suolo, espresso tra i 0° e i 360° (lo 0° corrisponde a vento proveniente da Nord e l'angolo cresce in senso orario);
- la velocità istantanea del vento a 10 metri dal suolo in metri al secondo;
- la pressione istantanea, in millibar;
- l'irradiazione solare, in kiloJoule su metro quadrato per ora.

Tutte queste variabili hanno una frequenza oraria, mentre per le variabili di produzione e i metalli osservati dai campionatori la frequenza è giornaliera. Nasce quindi il problema dell'utilizzo di tali variabili, anche perché non si è trovato alcun precedente studio in cui si trattasse un problema analogo a quello in esame, utilizzando delle variabili meteorologiche.

Si è proceduto utilizzando valori medi di tali variabili, suddividendo in medie diurne e notturne, in quanto è noto che la dispersione degli inquinanti in atmosfera obbedisce a meccanismi diversi in presenza od in assenza di radiazione solare. Il periodo diurno va dalle 8 alle 19, mentre quello notturno va dalle 20 alle 7 del giorno successivo. Questa scelta è in parte arbitraria, non considerando le differenze che si riscontrano nelle diverse stagioni.

Si è proceduto in prima battuta ad un'analisi descrittiva delle variabili meteorologiche disponibili, verificando anche se siano correlate tra di loro e se abbiano un andamento autoregressivo, in quanto verranno usate come variabili esplicative e si potrebbe creare un problema di multicollinearità.

Non si dispone di conoscenze approfondite sulle relazioni attese fra le concentrazioni dei metalli e le variabili meteorologiche: le relazioni emerse in fase di modellazione andranno interpretate alla luce di esperienze derivanti dall'applicazione di modelli numerici deterministici per la previsione della dispersione degli inquinanti.

Va peraltro osservato che, in tali modelli, molte delle variabili che alimentano gli algoritmi, e che sono dunque più direttamente connesse ai meccanismi di dispersione, non sono misurate ordinariamente (flusso di calore dal suolo, velocità di frizione), o non sono addirittura misurabili (altezza di rimescolamento, lunghezza di Monin-Obukhov).

Esse possono venire stimate a partire dalle osservabili misurate ordinariamente (vento e temperatura a diverse quote, copertura nuvolosa, umidità, radiazione solare) per mezzo di schemi parametrici estremamente complessi, ma inevitabilmente non accurati.

L'utilizzo, come variabili esplicative, di grandezze calcolate per mezzo degli schemi parametrici sopra citati, è stato abbandonato, per la poca significatività dimostrata; si ritiene opportuno che esso venga riconsiderato, dopo un miglioramento negli algoritmi utilizzati per la loro stima.

### 2.3.1 - Il vento

Una prima variabile, che si ipotizza molto importante per il trascinamento del metallo dalla fonte originale e per lo sviluppo di fenomeni turbolenti, è il vento.

La rappresentazione di tale variabile per mezzo di velocità e direzione istantanea, essendo quest'ultima una grandezza circolare, non risulta essere ottimale. Si è proceduto, quindi, alla costruzione di altre due variabili, combinazione delle precedenti, che rappresentano l'intensità del vento in direzione parallela ed ortogonale. Per direzione parallela si intende la direzione che va dall'impianto (in particolare, dal Forno1) al campionatore che di volta in volta viene preso in considerazione:

$$v_{1p} = \text{velocità} * \cos\left(\frac{(\text{direzione} - \text{angolo})}{180} \cdot \pi\right)$$
$$v_{1o} = \text{velocità} * \sin\left(\frac{(\text{direzione} - \text{angolo})}{180} \cdot \pi\right)$$

dove:

- per  $v1p$  è la velocità del vento parallela;
- per  $v1o$  è la velocità del vento ortogonale;
- *velocità* è la velocità istantanea del vento a 10 metri dal suolo;
- *direzione* è la direzione del vento a 10 metri dal suolo;
- *angolo* è l'angolo fra la congiungente Forno1-campionatore ed il Nord.

Effettuata questa trasformazione, vengono calcolate le medie.

Vengono di seguito riportate le distribuzioni di queste variabili, calcolate per la centralina A: in Tabella 2.2, gli indici sintetici delle quattro variabili, mentre in Figura 2.5 e 2.6 vengono rappresentati i rispettivi istogrammi.

Tabella 2.2 - Indici di distribuzione delle variabili vento

<b>Indice</b>	<b>v1p.d</b>	<b>v1o.d</b>	<b>v1p.n</b>	<b>v1o.n</b>
<b>Min</b>	-2.0250	-6.4930	-1.1160	-5.7880
<b>1<sup>st</sup> Qu</b>	-0.3219	-0.2488	0.5476	-0.8601
<b>Median</b>	0.3577	0.3355	0.8098	-0.4912
<b>Mean</b>	0.7341	0.3167	1.1330	-0.5314
<b>3<sup>rd</sup> Qu</b>	1.3670	0.9629	1.3520	-0.1731
<b>Max</b>	9.7410	6.8010	9.4310	4.0500
<b>Asimmetria</b>	1.2888	-0.3515	2.2437	-0.3781
<b>Curtosi</b>	5.0128	6.2271	10.7057	7.6828

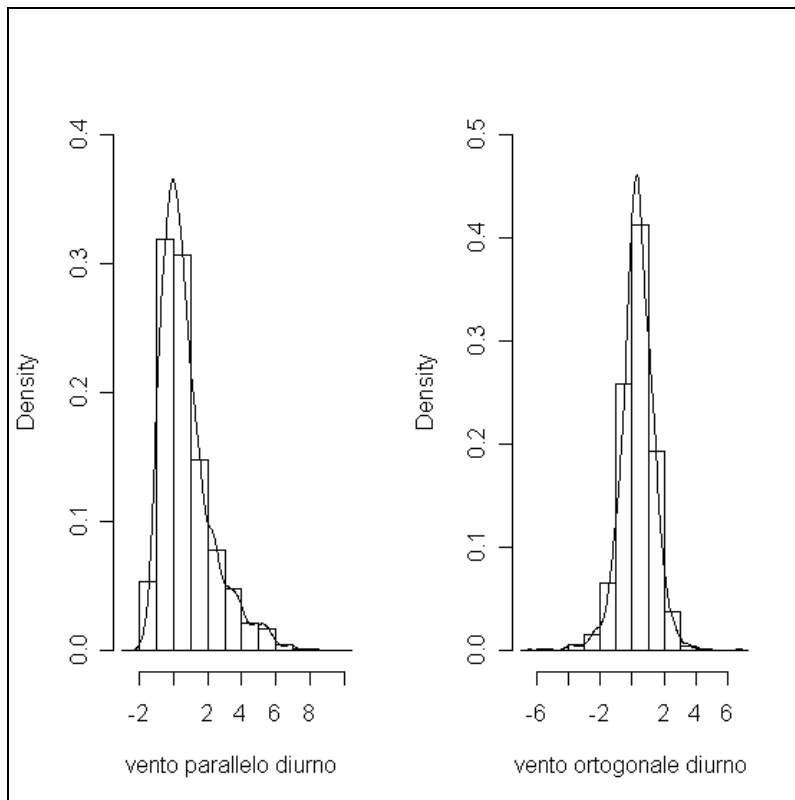


Figura 2.5 - Distribuzione del vento medio diurno

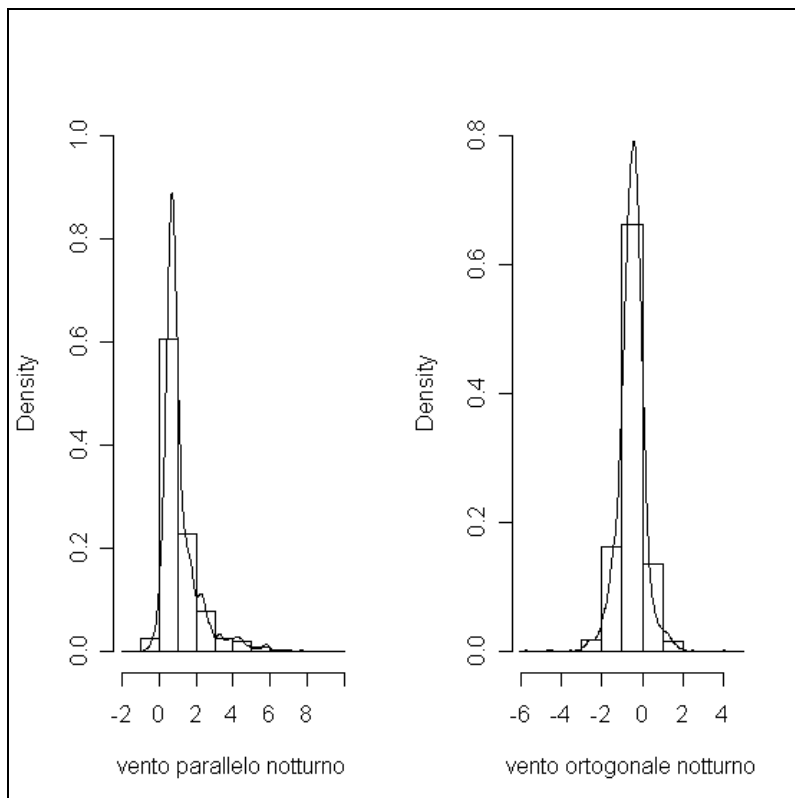


Figura 2.6 - Distribuzione del vento medio notturno



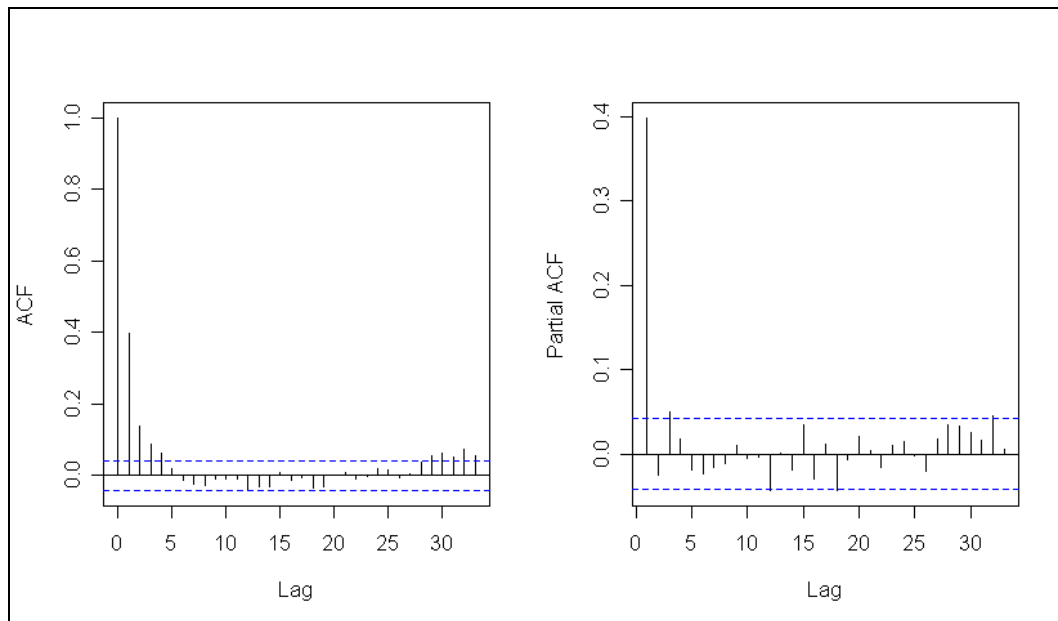


Figura 2.7 - Funzione di autocorrelazione e di autocorrelazione parziale della media notturna della velocità del vento parallelo

Si osserva che le variabili hanno delle code molto pesanti e che il vento parallelo è fortemente asimmetrico. Questa asimmetria è dovuta al fatto che la direzione Forno1/campionatore corrisponde a venti provenienti da Nord-Est, che è anche la direzione dei venti dominanti in quest'area: i valori molto alti sono dovuti prevalentemente alle giornate di bora.

Osservando attentamente le serie attraverso la funzione di autocorrelazione e la funzione di autocorrelazione parziale si evince che sia la media del vento parallelo che ortogonale, diurne e notturne, hanno un comportamento autoregressivo del primo ordine. Di seguito vengono illustrati, in Figura 2.7, le funzioni per la media notturna della velocità del vento parallelo.

### 2.3.2 - La temperatura

La temperatura, come le variabili meteorologiche che verranno illustrate successivamente, ha delle variazioni di tipo stagionale nell'anno, legate alle stagioni climatiche.

In Tabella 2.3 sono indicati i principali indici della temperatura media diurna e della temperatura media notturna.

Tabella 2.3 – Indici di distribuzione delle temperature medie

Indice	Temp. diurna	Temp. notturna
<b>Min</b>	-2.142	-6.117
<b>1<sup>st</sup> Qu.</b>	8.012	4.608
<b>Median</b>	15.130	10.800
<b>Mean</b>	15.420	10.690
<b>3<sup>rd</sup> Qu.</b>	22.430	16.790
<b>Max</b>	35.060	26.480
<b>Asimmetria</b>	0.1009	-0.0579
<b>Curtosi</b>	1.7223	1.7385

In Figura 2.8, a sinistra è rappresentata la distribuzione delle temperature medie diurne nei mesi autunnali e invernali, mentre a destra in quelli primaverili ed estivi.

I boxplot di Figura 2.9 (esempio con le temperature diurne) evidenziano l'andamento stagionale di questa variabile. Soprattutto si vede che tra il mese più caldo da quello più freddo c'è una distanza di ben 15°C.

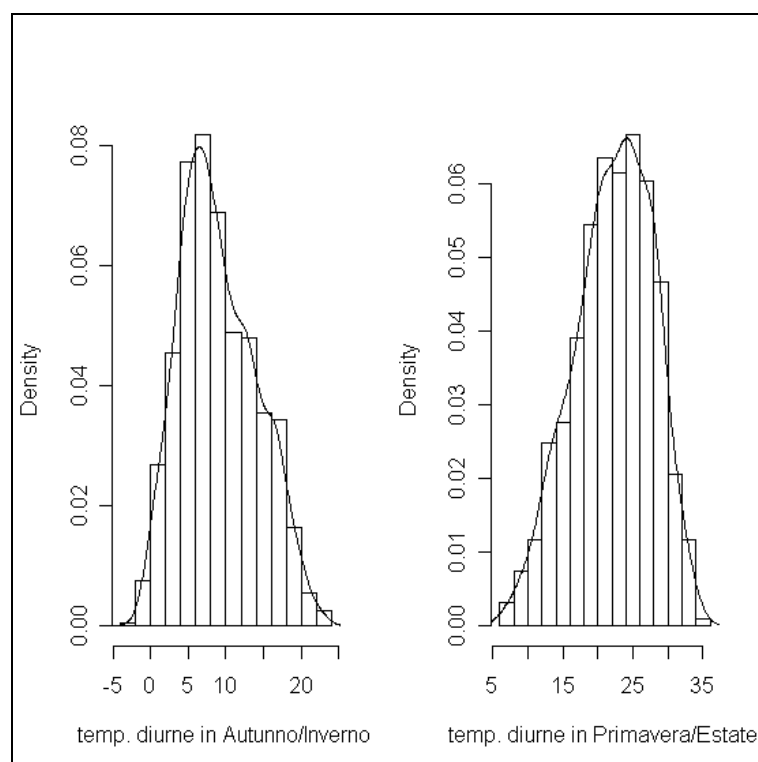


Figura 2.8 – Istogrammi delle temperature medie diurne suddivise tra mesi caldi e freddi

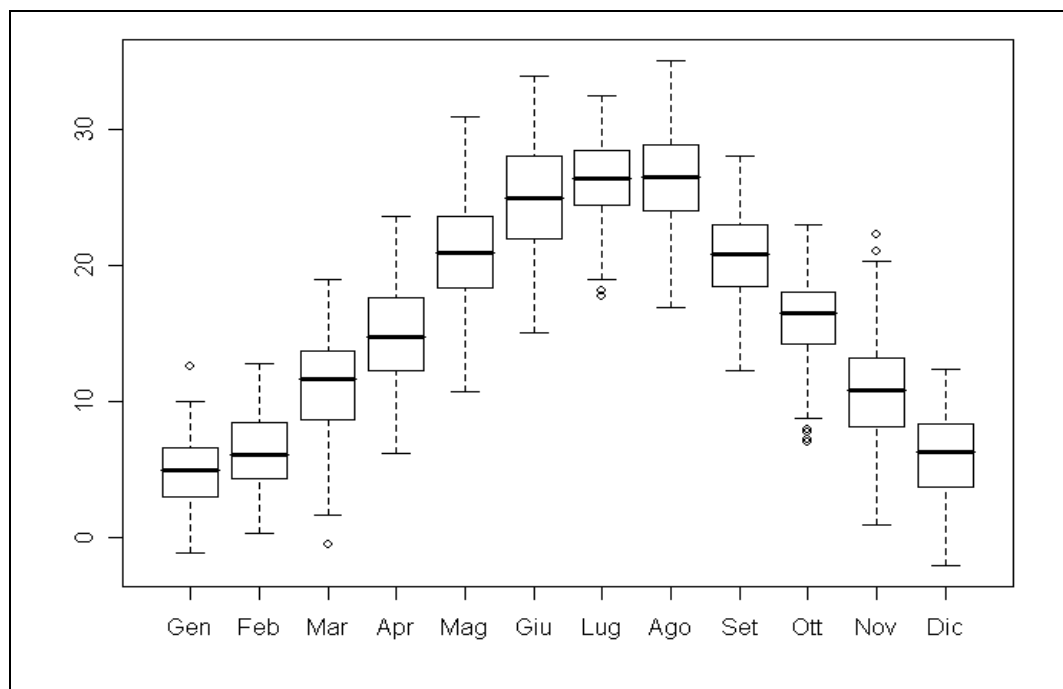


Figura 2.9 - Boxplot della temperatura diurna media distinta nei vari mesi

Analizzando la serie storica si vede che le due serie sono fortemente autocorrelate, a causa dell'effetto della stagionalità: in fase di modellazione si terrà conto di questa relazione (in Figura 2.10, esempio della temperatura notturna).

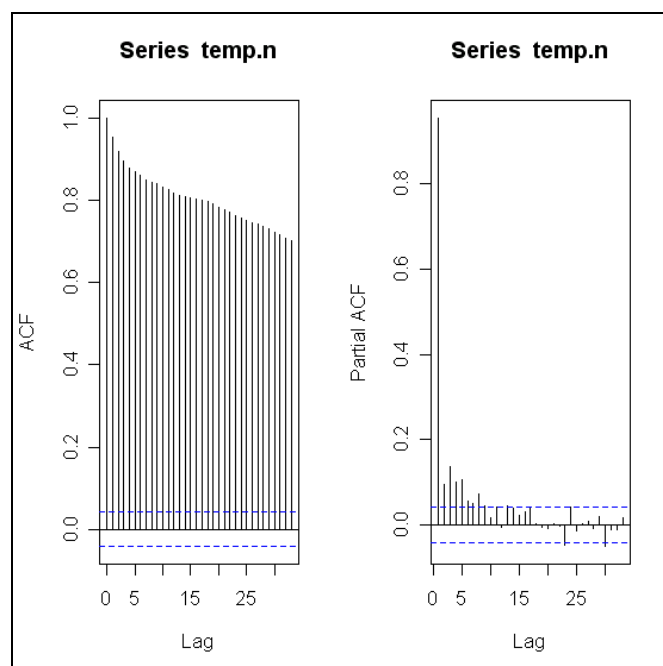


Figura 2.10 - Funzione di autocorrelazione, totale e parziale, della temperatura notturna media

### 2.3.3 - L'umidità relativa

La misura più utilizzata per descrivere l'umidità atmosferica è l'umidità relativa. Essa esprime il rapporto percentuale fra la quantità di vapore acqueo presente nell'aria e la quantità che, alla stessa temperatura, sarebbe necessaria perché il vapore condensi in microscopiche goccioline d'acqua. A questo punto si dice che l'aria è satura di vapore e, aggiungendo altro vapore, condensa in ulteriori goccioline.

La saturazione dipende fondamentalmente dalla temperatura dell'aria. Con basse temperature basta poco vapor acqueo perché si condensi una nuvola, mentre ne occorre molto di più man mano che la temperatura aumenta.

Quando l'umidità relativa dell'aria raggiunge il 100%, l'eccesso di vapore acqueo condensa in minuscole goccioline d'acqua e si assiste alla formazione di una nube.

Risulta, quindi, importante la relazione che gioca l'umidità con le altre variabili meteorologiche:

- si osserva una relazione positiva con la pioggia: l'umidità aumenta all'aumentare della pioggia;
- si osserva una relazione negativa con le radiazioni solari e la temperatura: all'aumentare di queste, l'umidità diminuisce.

Tabella 2.4 – Indici di distribuzione delle umidità medie

<b>Indice</b>	<b>Umidità diurna</b>	<b>Umidità notturna</b>
<b>Min</b>	12.17	20.08
<b>1<sup>st</sup> Qu.</b>	48.92	70.75
<b>Median</b>	60.67	82.25
<b>Mean</b>	62.15	79.75
<b>3<sup>rd</sup> Qu.</b>	75.56	91.75
<b>Max</b>	99.00	99.08
<b>Asimmetria</b>	0.0830	-0.7228
<b>Curtosi</b>	2.1105	2.8681

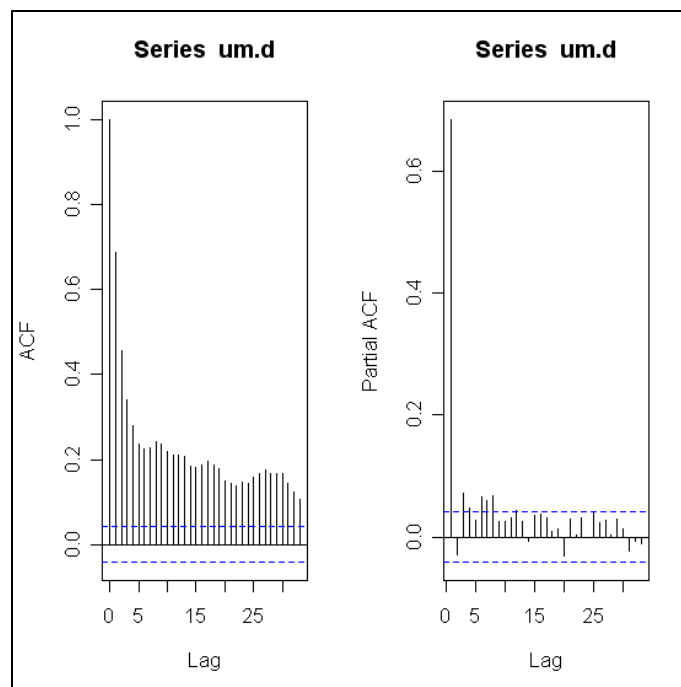


Figura 2.11 - Funzione di autocorrelazione e funzione di autocorrelazione parziale dell'umidità media diurna

L'umidità svolge un ruolo importante, ma molto complesso, nei processi di formazione e rimozione del particolato atmosferico.

Anche l'umidità è correlata con il valore osservato precedentemente e a causa della presenza di un trend. In Figura 2.11 si possono osservare le funzioni di autocorrelazione, totale e parziale, per la variabile diurna.

#### 2.3.4 - La pressione

La pressione varia durante il giorno perché dipende dalla variazione della temperatura della colonna d'aria. Il riscaldamento di questa, infatti, rende l'aria meno densa e quindi più leggera. Quindi durante le ore più calde del giorno, in seguito al riscaldamento del suolo e dell'aria negli strati più bassi, si ha un minimo di pressione. Di notte, al contrario, la colonna d'aria si raffredda e quindi si ha un aumento del suo peso, cioè della pressione.

La pressione può cambiare anche per il movimento delle masse d'aria: nella stessa colonna d'aria può entrare o uscire aria, per cui se quella che entra è maggiore di quella che esce si assiste ad un aumento di pressione, viceversa si avrà una sua diminuzione.

Usualmente la bassa pressione è associata a perturbazioni e piogge, spiegabile con considerazioni di dinamica atmosferica. Le zone di alta pressione, al contrario, tengono lontane le nubi in quanto queste zone sono caratterizzate da aria fredda che scende verso il suolo e che va incontro a compressione (ciò scongiura la condensazione e favorisce cielo sereno).

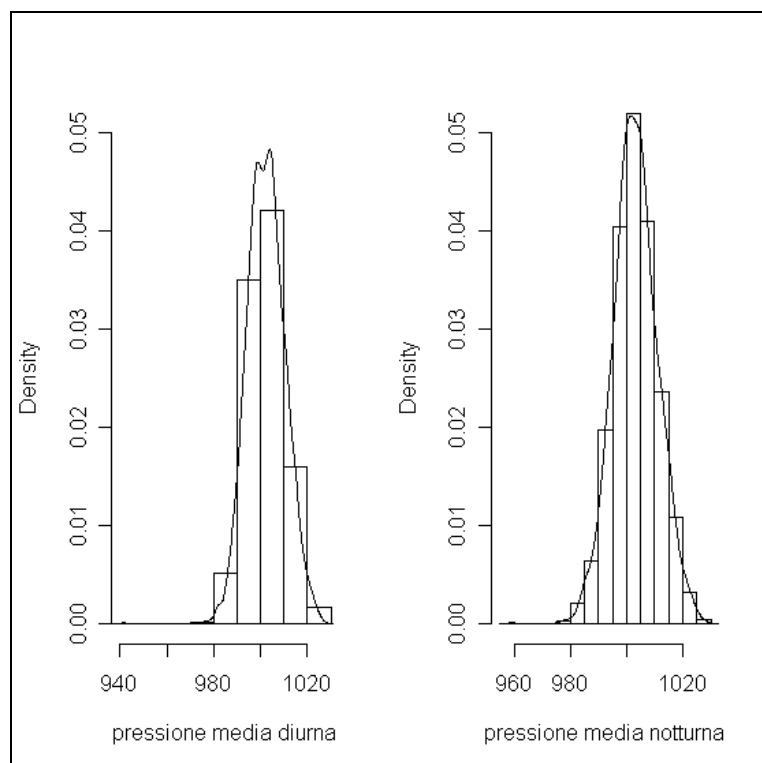


Figura 2.12 - Istogrammi della pressione media diurna e notturna

Tabella 2.5 – Indici di distribuzione delle pressioni medie

Indice	Pressione diurna	Pressione notturna
<b>Min</b>	941.5	959.1
<b>1<sup>st</sup> Qu.</b>	996.9	998.0
<b>Median</b>	1002.0	1003.0
<b>Mean</b>	1002.0	1003.0
<b>3<sup>rd</sup> Qu.</b>	1008.0	1008.0
<b>Max</b>	1026.0	1028.0
<b>Asimmetria</b>	-0.1123	-0.0173
<b>Curtosi</b>	3.9372	3.0560

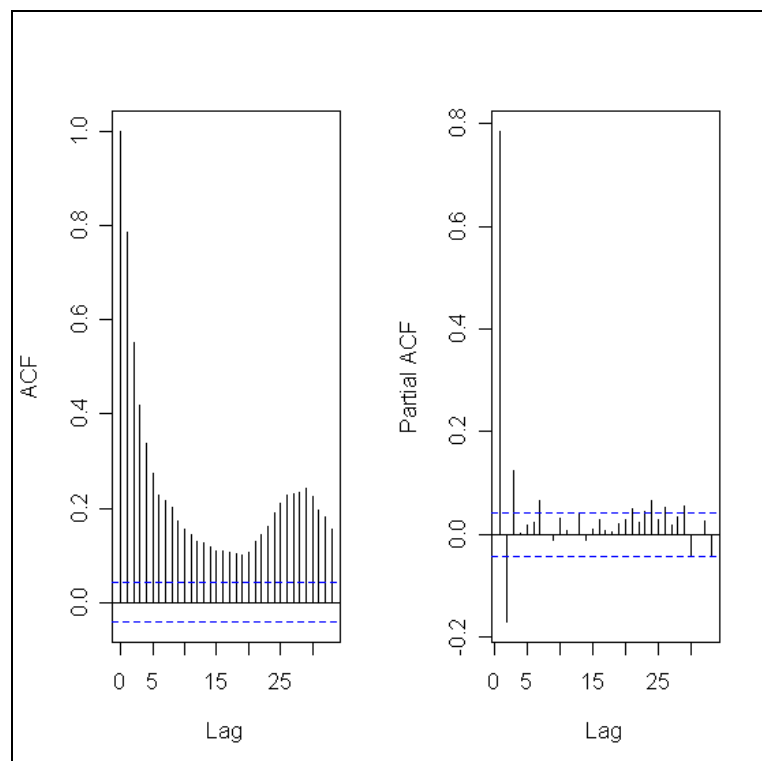


Figura 2.13 - Funzione di autocorrelazione e autocorrelazione parziale della pressione media notturna

Come per le altre variabili meteorologiche, sia la pressione diurna che notturna sono autocorrelate. In Figura 2.13 è rappresentata la pressione notturna. Nei modelli numerici per la dispersione degli inquinanti, la misura della pressione è utilizzata negli algoritmi che simulano i meccanismi di deposizione secca (distinta dalla deposizione umida, legata alle precipitazioni).

### 2.3.5 - La pioggia

Le precipitazioni sono un fenomeno che contribuisce a ripulire l'atmosfera, con particolare effetto sull'abbassamento dei livelli del particolato sospeso.

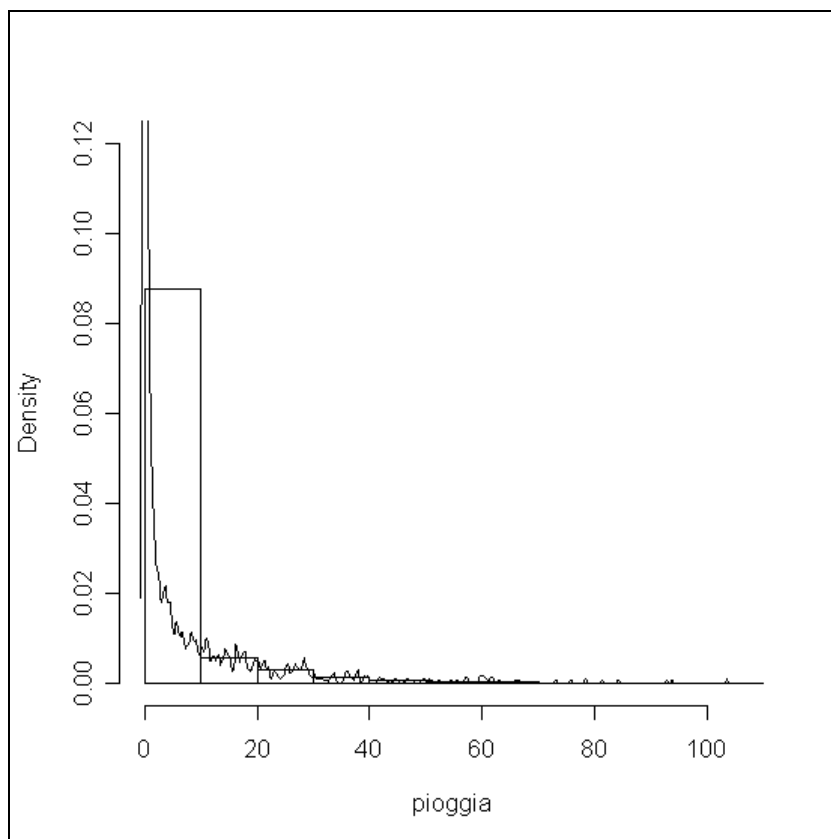


Figura 2.14 - Istogramma della pioggia

Tabella 2.6 – Indici di distribuzione delle precipitazioni giornaliere

<b>Indice</b>	<b>Pioggia giornaliera</b>
<b>Min</b>	0.000
<b>1<sup>st</sup> Qu.</b>	0.000
<b>Median</b>	0.000
<b>Mean</b>	3.946
<b>3<sup>rd</sup> Qu.</b>	1.400
<b>Max</b>	103.600
<b>Asimmetria</b>	3.7704
<b>Curtosi</b>	22.7151



Nel caso in esame, la pioggia è stata considerata come somma giornaliera. Per la maggior parte dell'anno non vi sono precipitazioni, quindi la variabile assume prevalentemente valore zero. Nei vari anni i giorni non piovosi sono stati 221 per il 2001, 184 per il 2002, 255 per il 2003, 198 per il 2004, 224 per il 2005 e 79 fino alla fine dell'aprile 2006. Il 2003, quindi, è stato l'anno meno piovoso.

Tuttavia, se si osserva la Figura 2.15, in cui si rappresenta il logaritmo della variabile quando questa è diversa da zero, si vede che il suddetto anno è quello che ha maggior variabilità e addirittura il valore maggiore osservato in tutto il dataset.

Osservando i comportamenti delle serie storiche (Figura 2.16) si può constatare che anch'essa è una serie che ha un comportamento di una serie autoregressiva del primo ordine.

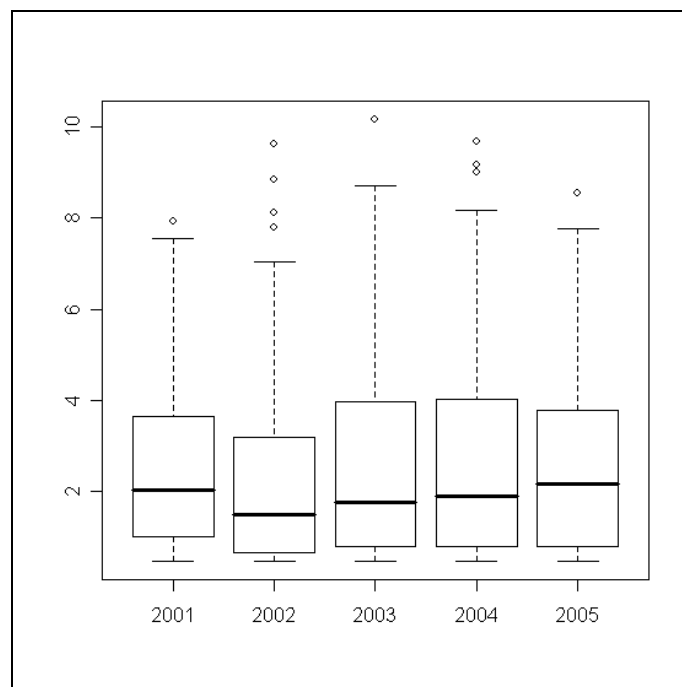


Figura 2.15 - Boxplot delle precipitazioni (in logaritmo) cadute nei vari anni

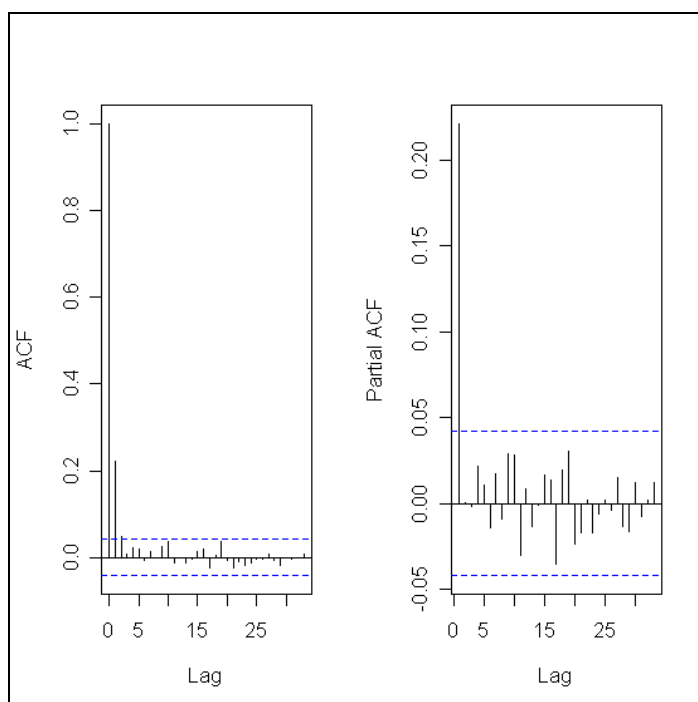


Figura 2.16 - Funzione di autocorrelazione e autocorrelazione parziale della pioggia giornaliera

### 2.3.6 - L'irraggiamento

Per quanto riguarda l'irraggiamento, la variabile a disposizione è il totale giornaliero della radiazione solare incidente sulla superficie (quindi la somma dei dati orari). Qui di seguito viene riportata una sintesi della variabile.

Tabella 2.7 – Indici di distribuzione delle radiazioni solari giornaliere

Indice	Radiazione solare giornaliera
<b>Min</b>	239
<b>1<sup>st</sup> Qu.</b>	5802
<b>Median</b>	10680
<b>Mean</b>	12500
<b>3<sup>rd</sup> Qu.</b>	19860
<b>Max</b>	30990
<b>Asimmetria</b>	0.33306
<b>Curtosi</b>	1.6821

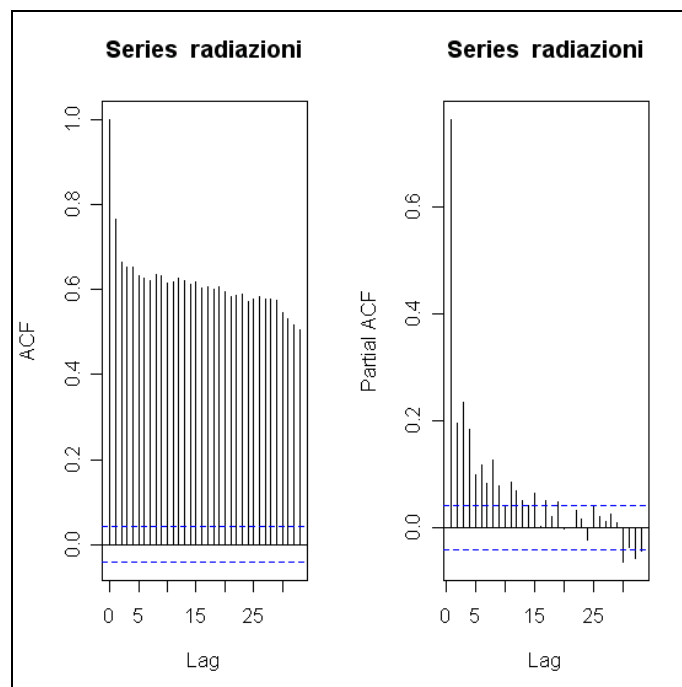


Figura 2.17 – Funzione di autocorrelazione, parziale e totale, dell'irraggiamento

Come si può vedere dalla Figura 2.17 la serie è autocorrelata.

### 2.3.7 - Le correlazioni tra le variabili meteorologiche

Nella fase di analisi e scelta delle variabili, soprattutto nei vari modelli lineari, è importante conoscere la correlazione tra le variabili esplicative; in caso di collinearità, infatti, si potrebbero creare dei problemi nella stima dei coefficienti dei modelli, come effetti di compensazione e aumento della varianza associata ai coefficienti stessi.

In Tabella 2.8, quindi, viene riportata la matrice di correlazione tra queste variabili. Essendo ben 18 le variabili, la matrice è stata spezzata in più tabelle.

di correlazione tra le variabili meteorologiche

temp.d	temp.n	um.d	um.n	press.d	press.n	Pioggia	radia-zioni	v1p.d	v1p.n	v1o.d	v1o.n	radio-height11	radio-temp11	radio-dewpt11	radio-height23	radio-temp23
1	.954	-.228	.126	-.236	-.219	-.009	.771	-.323	-.138	.290	-.006	.450	.878	.716	.449	.902
.954	1	-.027	.218	-.298	-.330	.082	.635	-.261	-.087	.260	.007	.368	.872	.810	.354	.900
-.228	-.027	1	.713	-.195	-.251	.386	-.604	-.003	-.094	-.066	.048	-.246	-.007	.281	-.257	-.027
.126	.218	.713	1	-.127	-.221	.248	-.167	-.317	-.427	.087	.087	.010	.265	.414	-.025	.285
-.236	-.298	-.195	-.127	1	.911	-.309	-.042	.020	-.037	-.042	.019	.601	-.171	-.361	.572	-.155
-.219	-.330	-.251	-.221	.911	1	-.302	-.017	.031	-.015	-.062	.027	.585	-.181	-.397	.638	-.185
-.009	.082	.386	.248	-.309	-.302	1	-.231	.146	.065	-.124	-.077	-.229	.066	.223	-.223	.059
.771	.635	-.604	-.167	-.042	-.017	-.231	1	-.347	-.106	.344	.026	.382	.563	.313	.397	.590
-.323	-.261	-.003	-.317	.020	.031	.146	-.347	1	.566	-.186	-.112	-.193	-.301	-.144	-.179	-.310
-.138	-.087	-.094	-.427	-.037	-.015	.065	-.106	.566	1	.006	.023	-.113	-.173	-.068	-.102	-.181
.290	.260	-.066	.087	-.042	-.062	-.124	.344	-.186	.006	1	.327	.118	.215	.187	.102	.209
-.006	.007	.048	.087	.019	.027	-.077	.026	-.112	.023	.327	1	.029	-.016	.022	.029	-.020
.450	.368	-.246	.010	.601	.585	-.229	.382	-.193	-.113	.118	.029	1	.510	.178	.923	.513
.878	.872	-.007	.265	-.171	-.181	.066	.563	-.301	-.173	.215	-.016	.510	1	.738	.463	.926
.716	.810	.281	.414	-.361	-.397	.223	.313	-.144	-.068	.187	.022	.178	.738	1	.149	.717
.449	.354	-.257	-.025	.572	.638	-.223	.397	-.179	-.102	.102	.029	.923	.463	.149	1	.492
.751	.838	.194	.419	-.308	-.377	.171	.397	-.210	-.090	.225	.019	.249	.728	.847	.195	1
.902	.900	-.025	.285	-.155	-.185	.059	.590	-.310	-.181	.209	-.020	.513	.926	.717	.492	.764

## **2.4 - I dati di concentrazione dei metalli**

Il particolato totale sospeso (PTS), nel quale viene determinata la presenza dei metalli di interesse, viene campionato da quattro centraline posizionate nelle vicinanze dell'acciaieria (Figura 2.18):

- la centralina A posizionata a circa 500 metri a valle dall'impianto nella direzione dei venti dominanti; i dati vanno dal 8 gennaio 2002 al 16 agosto 2006;
- la centralina B posizionata anch'essa a circa 500 metri dallo stabilimento ma nella direzione ortogonale a quella dei venti dominanti; i dati sono disponibili per il periodo che va dal 16 luglio 2002 al 18 agosto 2006;
- la centralina C posizionata a circa 750 metri dallo stabilimento, sempre sotto l'influsso dei venti dominanti ma situata più a sud rispetto alla centralina A; i dati sono disponibili dal 17 luglio 2002 al 16 agosto 2006;
- la centralina D, posizionata sulla direttrice della centralina A; essa dista dall'acciaieria, però, poco più di un kilometro; il periodo di riferimento è molto breve: dal 4 ottobre 2002 al 18 giugno 2004, perché la centralina è stata adibita da metà del 2004 al rilevamento del particolato di diametro aerodinamico inferiore ai 10  $\mu\text{m}$  (PM10).

I metalli che vengono campionati sono: il Manganese (Mn), il Ferro (Fe), il Nichel (Ni), lo Zinco (Zn), il Cadmio (Cd), il Piombo (Pb), il Cromo (Cr), il Vanadio (V) e l'Arsenico (As); le loro concentrazioni in aria sono espresse in  $\mu\text{g}/\text{m}^3$ .

È da notare che quest'ultimo metallo, l'Arsenico, si è cominciato a campionare soltanto dal febbraio 2004, così per la centralina D si possiedono 119 osservazioni su tale elemento.

Nella fase descrittiva dei dati, ci si è occupati di tutte e quattro le centraline, mentre nella fase di modellazione i dati sono quelli della centralina A, che è quella più vicina all'impianto e che subisce maggiormente l'effetto dei venti. Per questo motivo, ed anche per il fatto

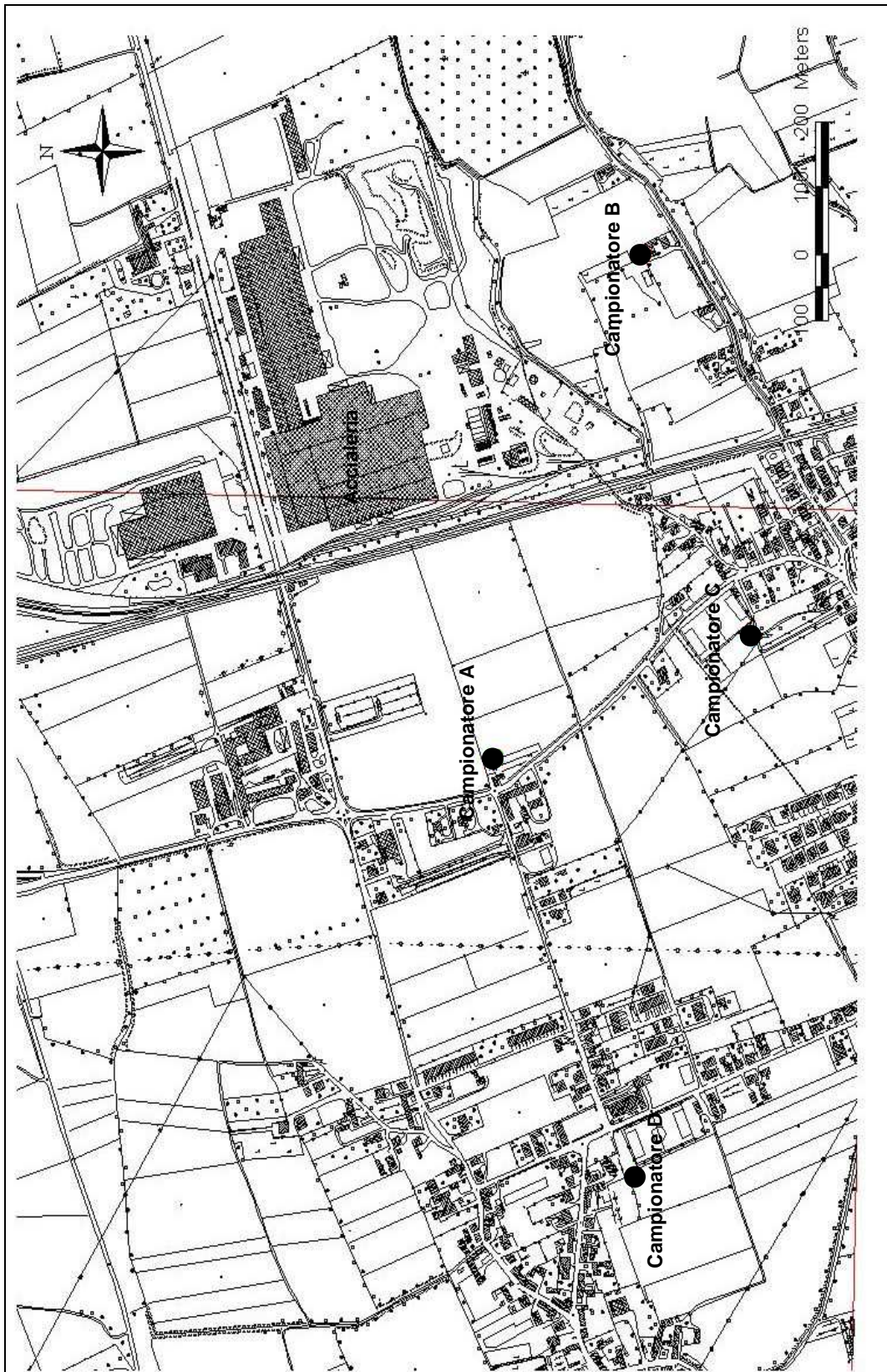


Figura 2.18 - Mappa delle posizioni delle centraline

che i metalli hanno mostrato uno stesso andamento in tutte le centraline, di seguito verranno inserite solo i grafici delle concentrazioni rilevate in A.

Nelle centraline, l'aria viene pompata dall'ambiente esterno, con un flusso controllato, attraverso dei filtri di nitrocellulosa, su cui rimane adeso il particolato atmosferico; ciascun filtro viene esposto per un giorno, dalle 00:05 alle 23:55. In seguito, i filtri vengono sottoposti ad un attacco acido; per la determinazione delle quantità dei metalli presenti, viene utilizzato uno spettrometro di massa (ICP-Massa). Il dato ottenuto viene ricondotto ad una concentrazione in aria-ambiente calcolando il volume di aria pompato attraverso il filtro durante la sua esposizione; più in particolare, note la temperatura e la pressione dell'aria, tale volume viene corretto nell'equivalente valore in condizioni standard di temperatura e pressione. La misurazione è affetta da un'incertezza che, per basse concentrazioni, può condurre a stime negative. Per questo motivo, sono state calcolate le soglie di rilevabilità del metodo per ogni metallo, al di sotto delle quali si può considerare una misurazione praticamente nulla.

Il procedimento per il calcolo delle soglie si basa sulla raccolta dei dati relativi ai metalli dei campioni cosiddetti "bianchi". Per "bianchi" si intende il filtro in estere di cellulosa pulito, non utilizzato per il campionamento, che viene trattato allo stesso modo dei filtri contenenti la polvere.

Per calcolare il limite di rilevabilità del metodo viene utilizzata la formula riportata dalla metodica EN 14902 del 2005.

Il test d'ipotesi che sta alla base è che sotto l'ipotesi  $H_0$  l'osservazione effettuata dal campionatore si distribuisca come la distribuzione dei filtri "bianchi", cioè si distribuisca come una normale  $N(0, \sigma^2)$ ; l'alternativa è che la distribuzione del campionatore sia maggiore di quella dei filtri bianchi (l'alternativa è quindi unilaterale destra).

In pratica, viene innanzi tutto calcolata la deviazione standard della media dei valori di massa del generico metallo ( $s_{Lba}$ ) misurati su filtri bianchi (in  $\mu\text{g}$  per filtro)

$$s_{Lba} = \sqrt{\frac{\sum_i (X_i - \bar{X})^2}{n-1}},$$

dove  $X_i$  è la massa dell'analita a (il metallo) rilevata nell'i-esimo filtro di laboratorio ed  $n$  è il numero di filtri.

Poi, con la seguente equazione, viene calcolato il limite di rilevabilità del metodo (sempre in  $\mu\text{g}$ ):

$$DL_{Ma} = t_{n-1,0.975} * s_{Lba}$$

in cui  $t_{n-1,0.95}$  è il valore della distribuzione t di Student con (n-1) gradi di libertà corrispondente al 95° percentile. Si effettua così un test con un  $\alpha$  pari al 5%. Dividendo, poi, per il volume campionato si ottiene il  $DL_{MCa}$ , che è il valore della soglia desiderata (in  $\mu\text{g}/\text{m}^3$ ).

I valori ottenuti sono i seguenti:

**Tabella 2.9 - Soglie rilevabilità degli elementi**

<b>Elemento</b>	<b>Soglia</b>
<b>Manganese</b>	0.001438
<b>Ferro</b>	0.109589
<b>Nichel</b>	0.004115
<b>Zinco</b>	0.049321
<b>Cadmio</b>	0.000014
<b>Piombo</b>	0.007892
<b>Cromo</b>	0.008122
<b>Vanadio</b>	0.001100
<b>Arsenico</b>	0.000975

I nove elementi chimici campionati si possono suddividere in tre gruppi a seconda della loro provenienza:

- Manganese e Ferro;
- Cadmio, Piombo e Zinco;
- Nichel, Cromo, Vanadio e Arsenico.

Di seguito vengono espone le analisi descrittive delle serie.



### 2.4.1 - Manganese e Ferro

La correlazione tra Manganese e Ferro è determinata dal fatto che sono metalli caratteristici di tutte le lavorazioni dell'acciaio.

Il Ferro viene emesse prevalentemente in fase fusoria e di affinazione, meno in in fase di carica e spillaggio. Il Manganese, invece, si osserva soprattutto in fase di spillaggio ma poco nella fase di carica e non lo si osserva in fase fusoria in quanto è un metallo bassofondente.

Il Ferro non ha rilevanza tossicologica, mentre il Manganese viene monitorato sia come tracciante delle lavorazioni siderurgiche che come fattore di rischio specifico per la rilevante tossicità.

E' importante evidenziare il fatto che mentre il Ferro presente nelle polveri atmosferiche in modo ubiquitario, avendo origine terrigena e dall'erosione dei manufatti, il Manganese è un metallo tipico della produzione di acciai e si riscontra nelle vicinanze di impianti destinati a tali attività.

In particolare i campionamenti presentano un distribuzioni non gaussiane (vedi Figura 2.19, 2.20 e Tabella 2.10, relative ai dati della centralina A). In entrambe i casi, la migliore trasformazione per renderle il più simili ad una distribuzione normale è quella logaritmica.

Per effettuare questa trasformazione, però, si è dovuto affrontare un problema molto importante: l'utilizzo dei dati pari a zero o addirittura negativi (molte volte in fase di calcolo del dato di concentrazione, se questo risulta negativo, viene posto pari a zero direttamente dall'addetto). L'unico metallo in cui non si incorre in questo problema è il Manganese: è l'unico ad avere tutte le osservazioni al di sopra della soglia di rilevabilità calcolata.

Per ovviare al problema, in fase di analisi descrittiva, prima di passare alla trasformazione logaritmica, si è scelto di effettuare una sostituzione: le misurazioni inferiori alla soglia di rilevabilità sono state sostituite con il valore della soglia stessa.

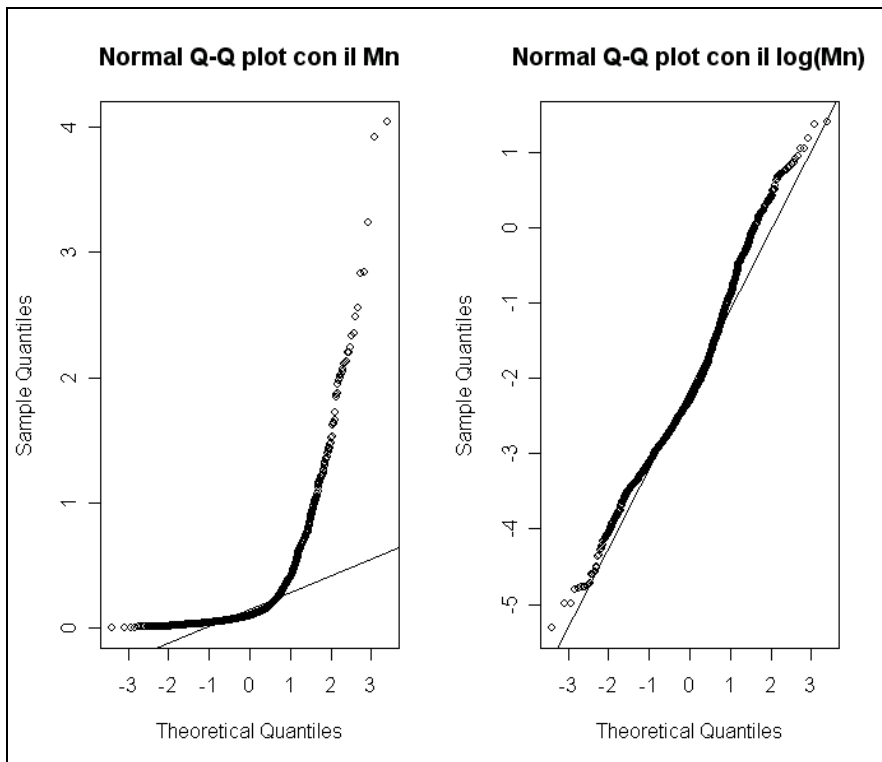


Figura 2.19 – Normal Q-Q plot del Manganese (Centralina A)

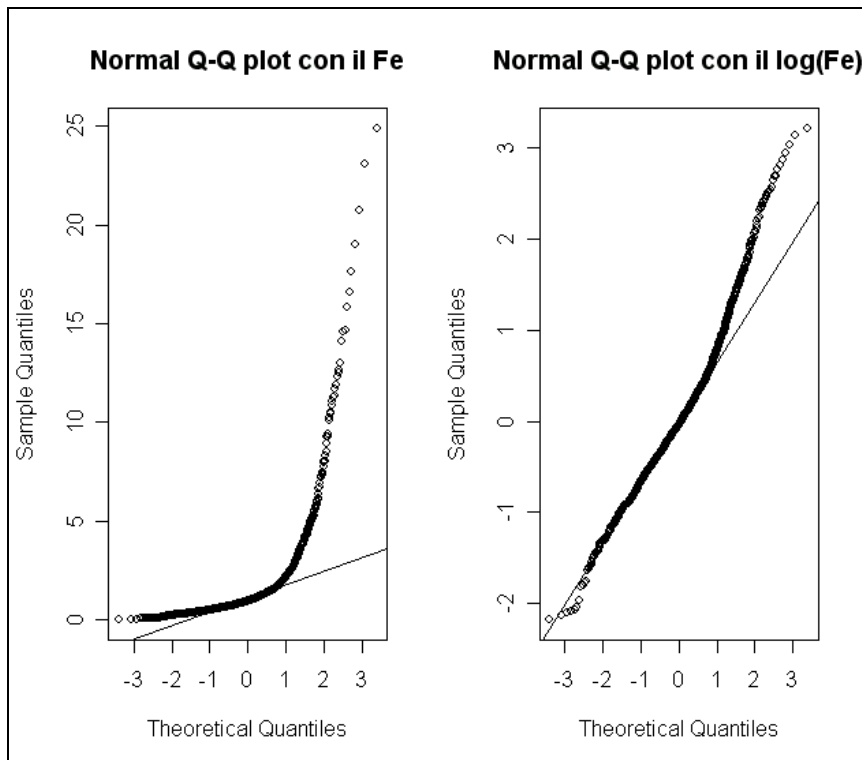


Figura 2.20 - Normal Q-Q plot del Ferro (Centralina A)

Tabella 2.10 - Indici delle distribuzioni di Mn e Fe (Centralina A)

Indice	Manganese (Mn)	log(Mn)	Ferro (Fe)	log(Fe)
Min	4.944e-03	-5.310	0.0000	-2.2110
1 <sup>st</sup> Qu.	5.772e-02	-2.852	0.6219	-0.4750
Median	1.006e-01	-2.297	0.9531	-0.0481
Mean	2.502e-01	-2.098	1.5590	0.0320
3 <sup>rd</sup> Qu.	2.401e-01	-1.427	1.5410	0.4323
Max	4.039e+00	1.396	24.8900	3.2140
Asimmetria	2.5619	0.3346	3.3561	0.3645
Curtosi	15.25435	2.0283	24.8871	2.7820

Sembra, inoltre, che ci sia un effetto di persistenza nell'aria dei due elementi chimici. Questo lo si può riscontrare nella Figura 2.21, la quale mostra che anche nei giorni in cui l'impianto produttivo non è in funzione si osservano dei valori, relativamente alti, che decrescono all'aumentare dei giorni di blocco.

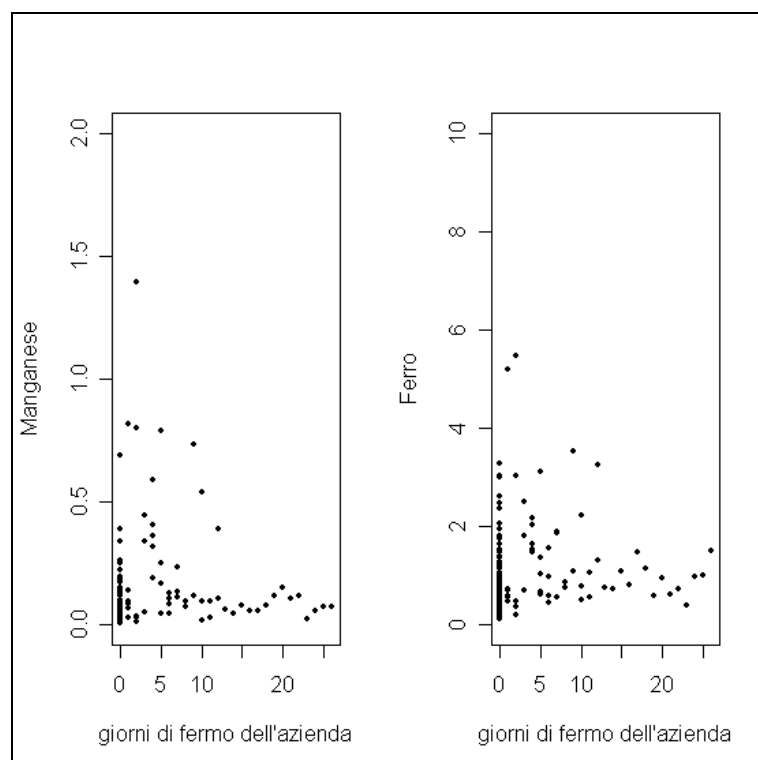


Figura 2.21 - Persistenza del Manganese e del Ferro nel tempo (Centralina A)

Si ipotizza che in zone non urbane (quindi aperte), il  $PM_{10}$  abbia una persistenza tra 1 e 3 giorni, mentre nelle zone urbane questa persistenza aumenti anche fino ad una settimana.

Per questo motivo si può ipotizzare che i valori osservati passati incidano sulle osservazioni odierne.

### 2.4.2 - Cadmio, Piombo e Zinco

Questi elementi sono frequentemente correlati tra loro in quanto il Cadmio risulta associato ai minerali di Zinco, mentre il Piombo è una impurezza dello Zinco.

Lo Zinco e il Piombo sono basso fondenti e perciò vengono emessi prevalentemente nella fase di carica e fusione (poco in affinazione e spillaggio) e la loro presenza dipende principalmente dalla qualità del rottame.

Quest'ultimo non ha rilevanza tossicologica e viene determinato unicamente come tracciante dell'attività delle acciaieria. Il Cadmio è un elemento dannoso per la salute a causa della sua rilevante tossicità.

Per quanto riguarda il Piombo è necessario ricordare che la sua presenza in atmosfera in un recente passato era dovuta soprattutto dalle emissioni del traffico veicolare; attualmente, nell'area di interesse, si ritiene legata prevalentemente alle emissioni dell'acciaieria (dall'anno 2002 è stata abolita la vendita di benzina che utilizzi tale metallo come antidetonante). Come specificato nel capitolo 1, è da ricordare che la provenienza di Piombo potrebbe derivare non solo dalla produzione di acciai speciali ma anche dalla fusione, nel rottame, di batterie scariche non ritirate.

Anche per le concentrazioni di questi elementi, attraverso una trasformazione logaritmica, si ottiene una distribuzione campanulare simile a quella gaussiana, a parte le code molto più pesanti e il problema legato alla soglia di rilevabilità (vedi Figure 2.22, 2.23 e 2.24 e Tabelle 2.11 e 2.12).

Tabella 2.11 - Indici distribuzione Cd (Centralina A)

Indice	Cadmio (Cd)	log(Cd)
Min	0.000e+00	-11.180
1 <sup>st</sup> Qu.	2.975e-04	-8.120
Median	6.055e-04	-7.409
Mean	1.395e-03	-7.401
3 <sup>rd</sup> Qu.	1.270e-03	-6.669
Max	3.100e-02	-3.474
Asimmetria	3.3668	-0.2177
Curtosi	24.9893	2.8317

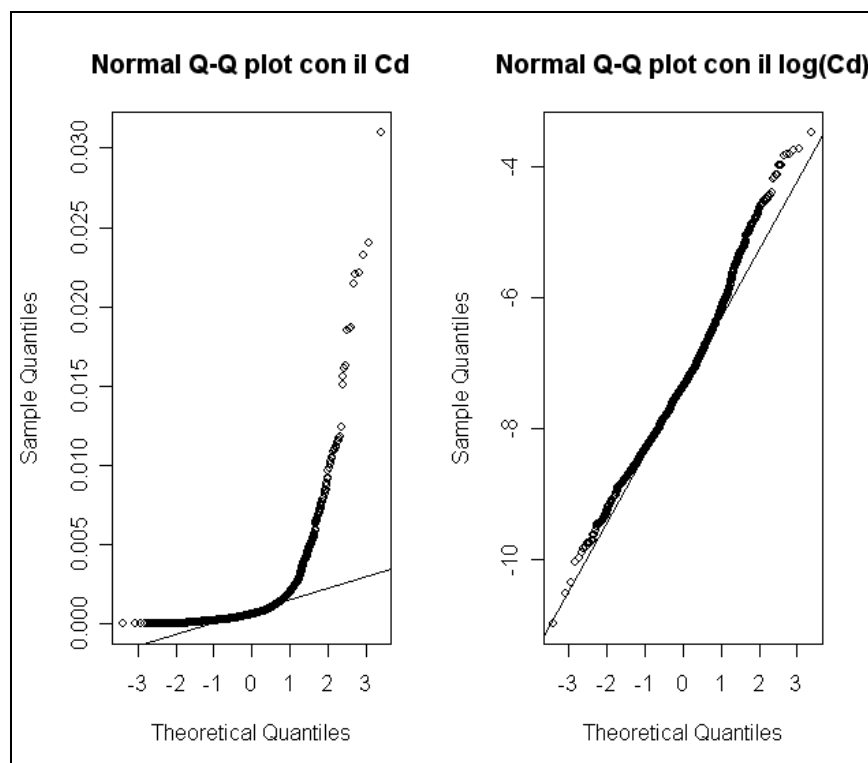


Figura 2.22 - Normal Q-Q plot del Cadmio (Centralina A)

Tabella 2.12 - Indici delle distribuzioni di Pb e Zn (Centralina A)

Indice	Piombo (Pb)	log(Pb)	Zinco (Zn)	log(Zn)
Min	-0.0016	-4.842	0.0000	-3.0090
1 <sup>st</sup> Qu.	0.0221	-3.812	0.1245	-2.0830
Median	0.0423	-3.162	0.2495	-1.3880
Mean	0.1363	-2.935	0.7486	-1.1800
3 <sup>rd</sup> Qu.	0.1131	-2.179	0.6665	-0.4058
Max	6.6420	1.893	13.7000	2.6170
Asimmetria	6.3966	0.4639	2.6776	0.4280
Curtosi	97.2579	2.0721	15.5679	1.8924

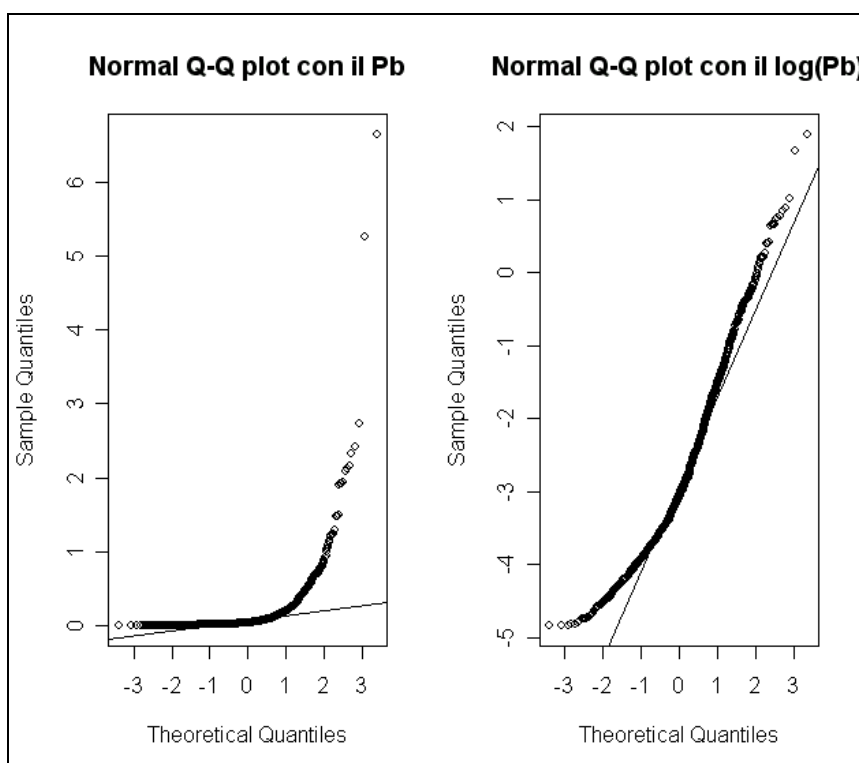


Figura 2.23 - Normal Q-Q plot del Piombo (Centralina A)

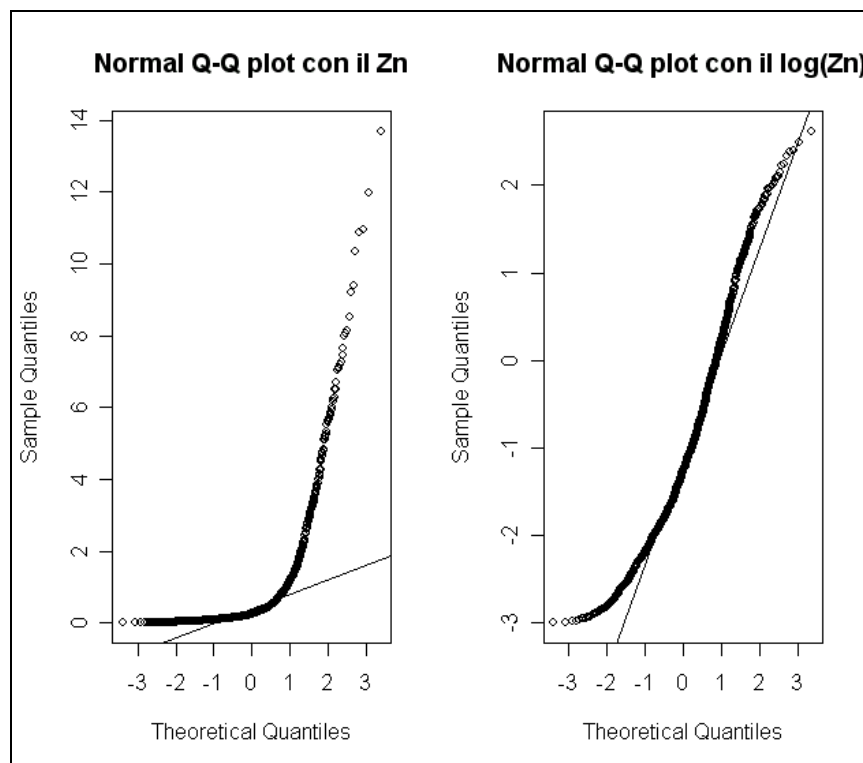


Figura 2.24 - Normal Q-Q plot dello Zinco (Centralina A)

Si è riscontrato l'effetto di persistenza del particolato nei giorni successivi alla chiusura dello stabilimento (nella Figura 2.25 l'esempio del Cadmio).

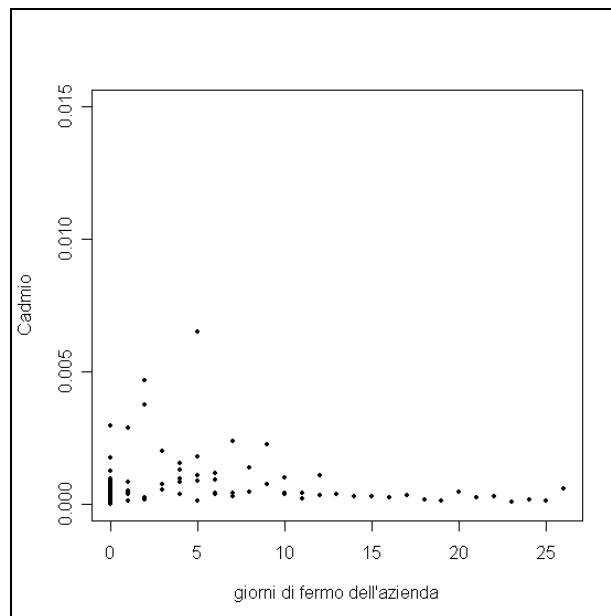


Figura 2.25 - Persistenza del Cadmio nel tempo

2.4.3 - Nichel, Cromo, Vanadio ed Arsenico

Si ritiene che la presenza di Nichel, Cromo e Vanadio nelle emissioni sia principalmente conseguenza della produzione di acciai speciali, oltre che della presenza di cromature e zincature nei rottami.

Il Nichel è indicato come potenziale cancerogeno, il Cromo specialmente nella forma esavalente risulta cancerogeno e mutageno anche se in altri stati di ossidazione (trivalente) risulta meno pericoloso.

Come per i metalli precedenti, la trasformazione logaritmica è risultata la migliore.

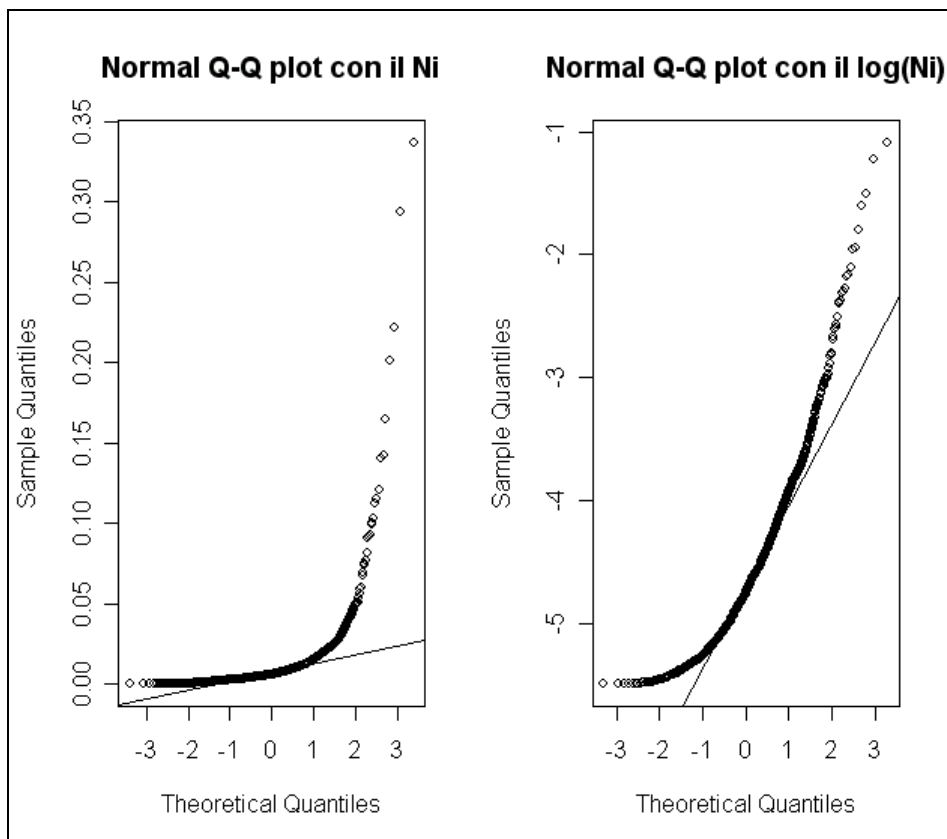


Figura 2.26 - Normal Q-Q plot del Nichel (Centralina A)



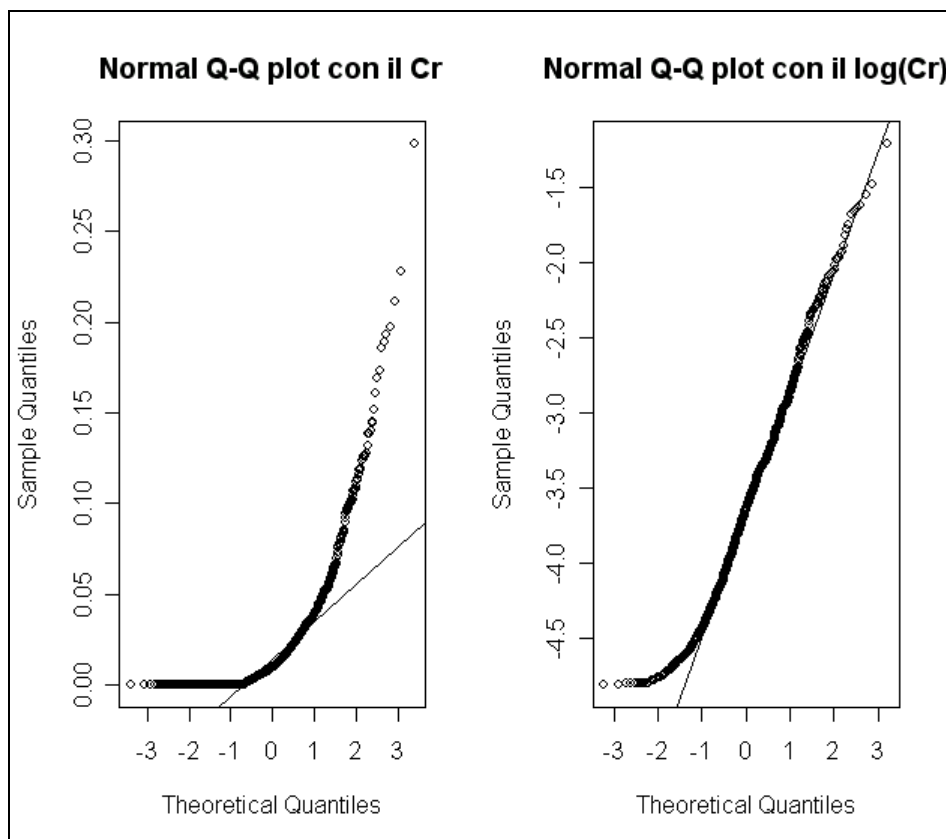


Figura 2.27 - Normal Q-Q plot del Cromo (Centralina A)

Tabella 2.13 - Indici delle distribuzioni di Ni e Cr (Centralina A)

Indice	Nichel (Ni)	log(Ni)	Cromo (Cr)	log(Cr)
<b>Min</b>	-0.000019	-5.493	-7.000e-05	-4.813
<b>1<sup>st</sup> Qu.</b>	0.003509	-5.493	4.215e-04	-4.813
<b>Median</b>	0.006067	-5.105	9.683e-03	-4.637
<b>Mean</b>	0.010570	-4.880	2.035e-02	-4.171
<b>3<sup>rd</sup> Qu.</b>	0.010850	-4.524	2.801e-02	-3.575
<b>Max</b>	0.337200	-1.087	2.985e-01	-1.209
<b>Asimmetria</b>	5.9490	1.0027	2.01874	0.7022
<b>Curtosi</b>	77.2634	3.7927	11.0839	2.0493

Tabella 2.14 - Indici di distribuzione di V e As (Centralina A)

Indice	Vanadio (V)	log(V)	Arsenico (As)	log(As)
<b>Min</b>	-0.000152	-6.812	-2.770e-04	-6.933
<b>1<sup>st</sup> Qu.</b>	0.001955	-6.237	1.059e-03	-6.850
<b>Median</b>	0.003254	-5.728	1.793e-03	-6.324
<b>Mean</b>	0.004123	-5.733	3.734e-03	-6.075
<b>3<sup>rd</sup> Qu.</b>	0.005329	-5.235	3.594e-03	-5.629
<b>Max</b>	0.041350	-3.186	4.816e-02	-3.033
<b>Asimmetria</b>	2.0802	0.1401	1.4738	0.4370
<b>Curtosi</b>	14.5314	1.7083	8.1726	1.3919

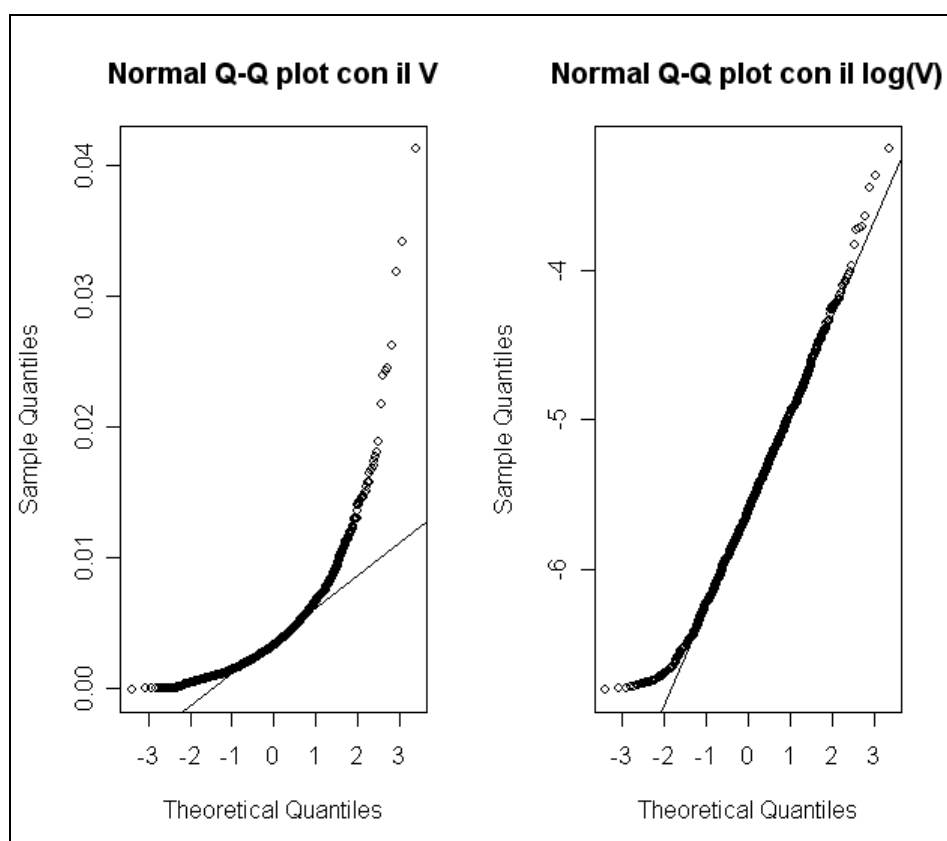


Figura 2.28 - Norma Q-Q plto del Vanadio (Centralina A)

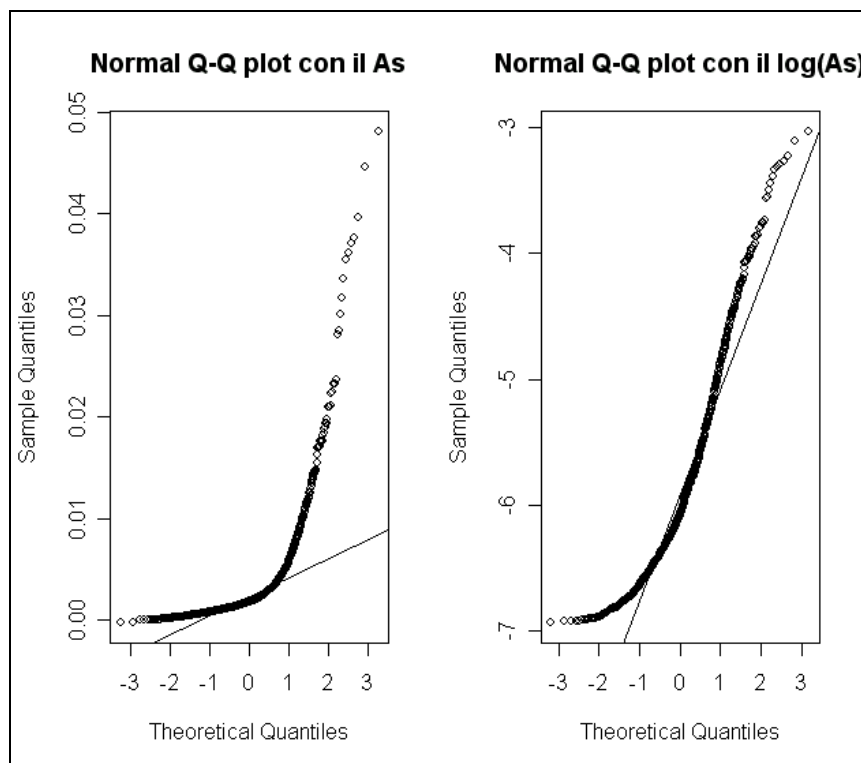


Figura 2.29 - Normal Q-Q plot dell'Arsenico (Centralina A)

Da studi fatti dall'ARPA FVG è risultato che i valori del Vanadio osservati nella centralina A sono equiparabili ai livelli dello stesso osservati nell'ambito urbano della città di Udine.

Osservando le figure inerenti al Cromo è facile notare che ben il 50% dei dati è al di sotto della soglia di rilevabilità. Infatti, di questo elemento si osservano spesso livelli veramente bassi, tali da creare dei problemi sulla stima dello stesso con il metodo di misura usato.

Infine, l'Arsenico si è cominciato ad osservare e campionare solo dal 2004: ciò fa sì che nelle analisi congiunte il dataset disponibile sostanzialmente si dimezzi.

Per i motivi appena elencati, questi tre metalli non verranno considerati nella fase di modellazione.

La persistenza dei metalli è riscontrata anche in quest'ultimo gruppo (nell'esempio di Figura 2.30 è rappresentato il Nichel).

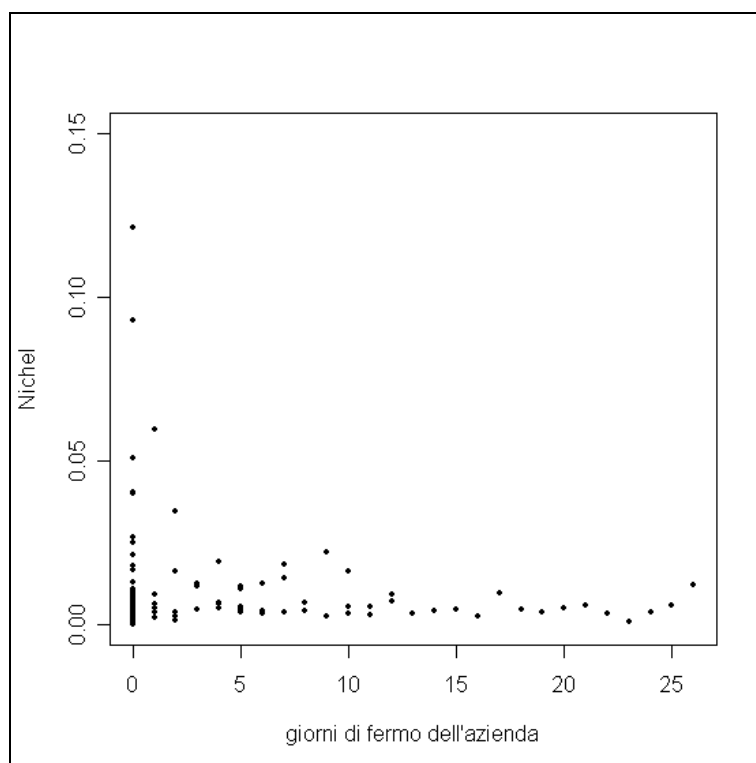


Figura 2.30 - Persistenza del Nichel nel tempo (Centralina A)

#### 2.4.4 - Correlazione tra gli elementi

Uno degli obiettivi che l'ARPA FVG vorrebbe raggiungere è la riduzione del numero di elementi chimici da analizzare. Questo porterebbe ad una sensibile riduzione dei costi e dei carichi di lavoro legati a questo monitoraggio, liberando risorse da impegnare presso altre situazioni potenzialmente critiche.

Individuare le relazioni tra i metalli, quindi, è fondamentale.

Qui di seguito, in Figura 2.31, viene riportata la matrice di correlazione in versione grafica dei metalli campionati in A. La matrice dei 9 elementi è rappresentata da un quadrato di 9x9 quadrati più piccoli, ognuno dei quali è una tonalità di grigio e rappresenta la correlazione. La correlazione pari a 1 è rappresentata dal nero, mentre la correlazione pari a 0 è identificata dal bianco; per i valori intermedi si utilizza la scala dei grigi.

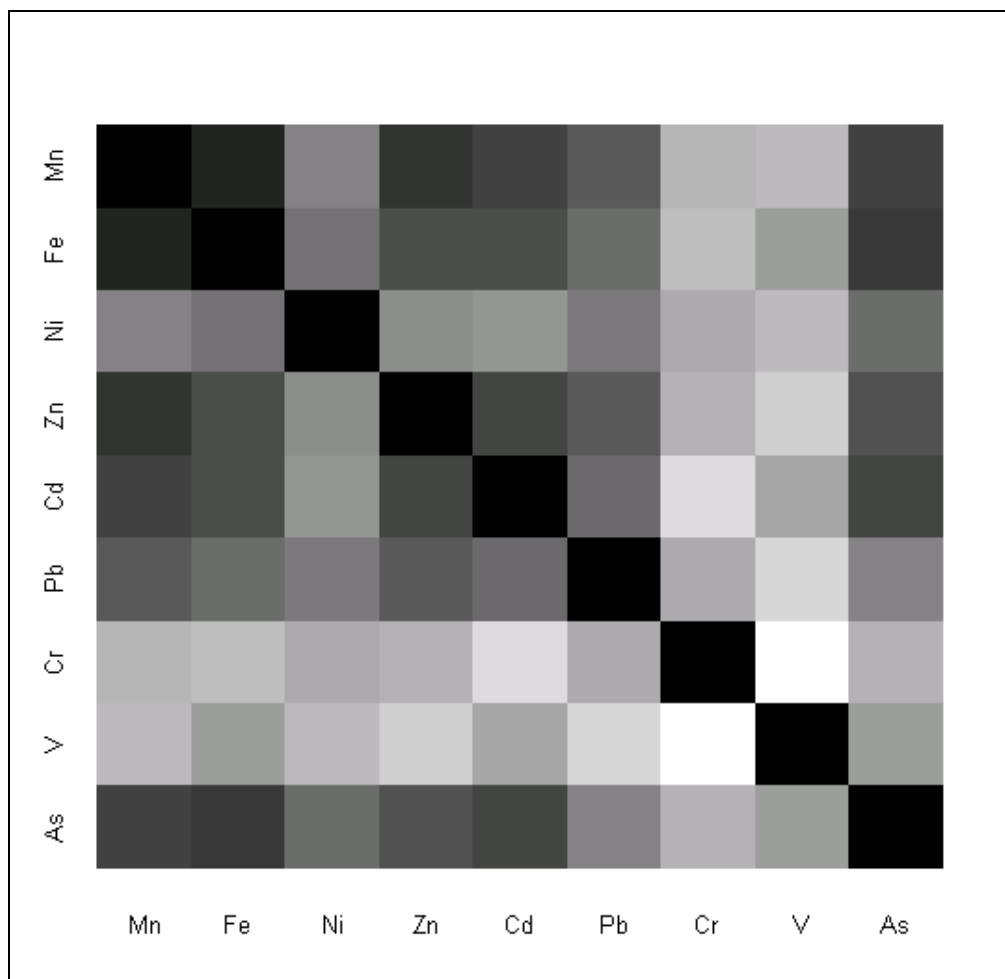


Figura 2.31 - Matrice di correlazione dei metalli osservati nella Centralina A

È da aggiungere una osservazione importante su queste correlazioni. Queste appena rappresentate sono relative ai dati della centralina A, che è quella più vicina allo stabilimento ed è sotto l'influenza dei venti dominanti. Inoltre la matrice viene generata a partire dall'intero set tutti di dati, indipendentemente dalle condizioni di funzionamento dello stabilimento.

In Figura 2.31 si possono vedere dei gruppi (si esclude dai commenti l'Arsenico a causa dell'esigua numerosità di tali dati). Innanzitutto si vede che il Cromo, il Vanadio e il Nichel, i quali correlano poco con gli altri metalli. Un primo gruppo è dato dal Manganese e Ferro con una correlazione pari a 0.866. Il secondo gruppo è rappresentato dallo Zinco, Cadmio e Piombo. Questa suddivisione è la stessa che viene effettuata

dalla dalla letteratura per distinguere i metalli a seconda della loro provenienza.

Se però si estraggono solamente i dati in cui l'acciaieria non lavora (Figura 2.32), questa matrice assume dei valori del tutto simili alla matrice di correlazione calcolata per la centralina B (spostata in direzione Sud-Est) nel caso generale.

Si osserva che tutte le correlazioni diminuiscono e l'unica correlazione davvero importante rimane quella tra Ferro e Manganese.

Questa similitudine mette in luce il fatto che molto probabilmente la centralina B non osserva tutti gli effetti prodotti dall'acciaieria e questo è molto probabilmente dovuto alla posizione della centralina: infatti, è posizionata ortogonalmente rispetto alla direzione prevalente dei venti.

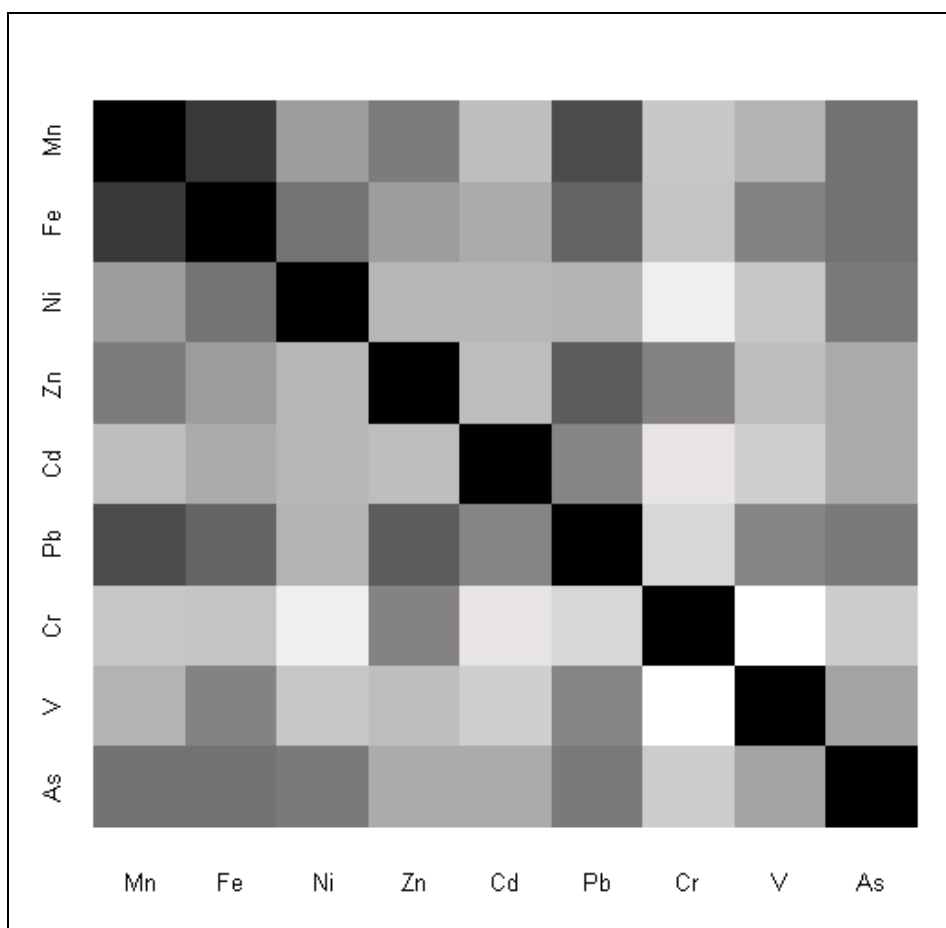


Figura 2.32 - Correlazioni tra i metalli osservati presso la Centralina A, quando l'acciaieria non produce

## CAPITOLO 3

### I modelli

#### ***3.1 - L'analisi delle componenti principali***

Prima di procedere con lo sviluppo di modelli lineari, si è inteso effettuare un'analisi delle componenti principali (PCA) [9,10].

Stabilito che, nel nostro problema, si intendono spiegare le misure di concentrazione osservate dei vari metalli, con la PCA si punta ad una riduzione del numero di tali variabili, mediante la costruzione di variabili derivate, ottenute da combinazioni lineari delle originarie. Con ciò, si intende:

- semplificare il lavoro di costruzione dei modelli esplicativi
- raggruppare le variabili di interesse, analogamente a quanto emerso dall'analisi della matrice di correlazione

I risultati, qui sotto riportati ed affiancati dai grafici utilizzati per la scelta del numero di componenti, sono stati del tutto insoddisfacenti.

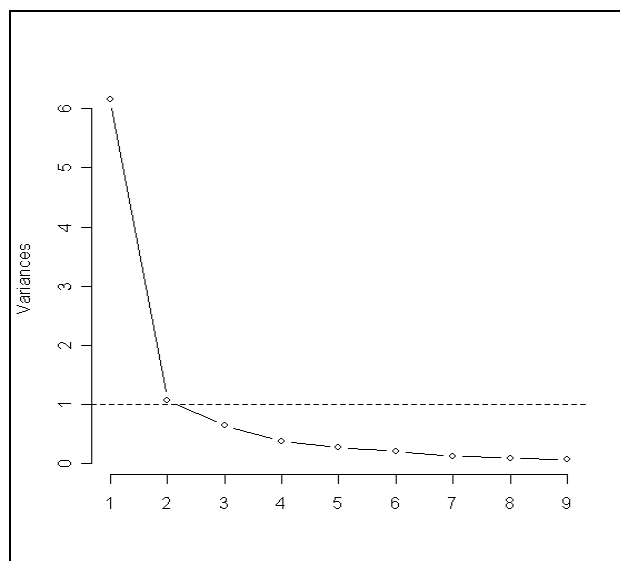


Figura 3.1 - Plot degli autovalori calcolati dalla PCA

Importance of components:

	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9
Standard deviation	2.38	0.966	0.848	0.721	0.658	0.507	0.452	0.426	0.282
Proportion of Variance	0.63	0.104	0.080	0.058	0.048	0.029	0.023	0.020	0.009
Cumulative Proportion	0.63	0.734	0.814	0.872	0.920	0.948	0.971	0.991	1.000

Con le prime tre componenti si riesce a spiegare già più dell'80% della varianza totale e per decidere il numero di componenti si è usato il criterio di Kaiser: si tengono in considerazione solo i fattori con autovalori maggiori di 1.

La matrice di rotazione è stata riportata fino fino alla quinta componente, in quanto i fattori rimanenti spiegano soltanto l'8% della varianza totale.

	PC1	PC2	PC3	PC4	PC5
Mn	-0.3854355	0.05218066	-0.24018714	0.11781428	-0.07530782
Fe	-0.3891817	-0.06414849	-0.03772721	-0.01228840	-0.20488472
Ni	-0.3094894	0.04710385	0.47312109	-0.64526042	-0.34441537
Zn	-0.3649553	0.08450513	-0.37839972	0.07624856	0.02506251
Cd	-0.3594270	-0.18630548	-0.26238004	0.19859671	0.05791431
Pb	-0.3235929	0.22426948	-0.08529943	-0.44381254	0.72735841
Cr	-0.2245760	0.64677395	0.50523803	0.50956434	0.08755358
V	-0.2263985	-0.68620061	0.49228995	0.24334115	0.35568172
As	-0.3679814	-0.09801996	-0.01107898	0.09226580	-0.40807041



Come si può osservare dai valori forniti, il criterio adottato conduce alla scelta di una sola componente, che risulta essere una media ponderata dei metalli, come è tipico dell'analisi delle componenti principali.

Poiché questa soluzione non si è ritenuta adeguata ai fini della semplificazione del problema si è proceduto direttamente all'analisi dei modelli lineari.

### **3.2 - La censura dei dati**

Nel capitolo 2, è stato mostrato che tutti e nove i metalli osservati da tutte e quattro le centraline assumono una distribuzione che si avvicina a una normale, dopo una trasformazione logaritmica.

Come detto nel capitolo precedente, in fase di analisi si considerano solo i dati della centralina A.

Il problema di questa trasformazione è che è definita solo per valori maggiori di zero. Qui, invece, per effetto delle incertezze legate alla misura, ci sono delle osservazioni sia nulle che negative.

La creazione di un modello, quindi, deve tenere in considerazione queste problematiche: per utilizzare la trasformazione logaritmica, si è dovuto effettuare una censura sui dati.

L'utilizzo dei dati censurati non è sempre condotto con facilità, soprattutto nella scelta del limite, in questo caso inferiore. Per fortuna, per la scelta di tale limite le normative vigenti ci vengono incontro (vedi capitolo 2.4): è la stessa normativa europea che ci descrive come calcolare la soglia di rilevabilità del campionario.

La censura è stata effettuata su tali soglie; ciò evita anche che si tenti di modellare dati determinati per lo più dall'incertezza strumentale.

Attuata la censura, si è cercato di creare un modello nel complesso semplice ma che allo stesso tempo mettesse in evidenza la censura stessa.

A tal fine, si è proceduto alla creazione di due variabili per ogni metallo:

$$X = \begin{cases} 0 & \text{osservazione} < \text{soglia} \\ 1 & \text{altrimenti} \end{cases}$$

$$Z = \begin{cases} \text{NA} & X = 0 \\ \text{osservazione} & X = 1 \end{cases}$$

Il modello finale, quindi, si basa sulla stima di due modelli:

- uno binomiale, per la stima di X
- uno lineare, per la stima del log(Z)

È da ricordare che l'Arsenico, il Vanadio e il Cromo non sono stati analizzati per i motivi esposti nel capitolo 2.

In alcuni modelli, inoltre, si è anche inserito un'ulteriore variabile (*wd* – week day) che rappresenta i giorni della settimana. L'inserimento di *wd* è importante, in quanto si osserva, soprattutto per il Manganese e il Ferro, una diminuzione di tali metalli durante il sabato e la domenica.

Ciò si può vedere dalla Figura 3.2 in cui sono raffigurati i boxplot distinti per i giorni della settimana per il Manganese e il Ferro.

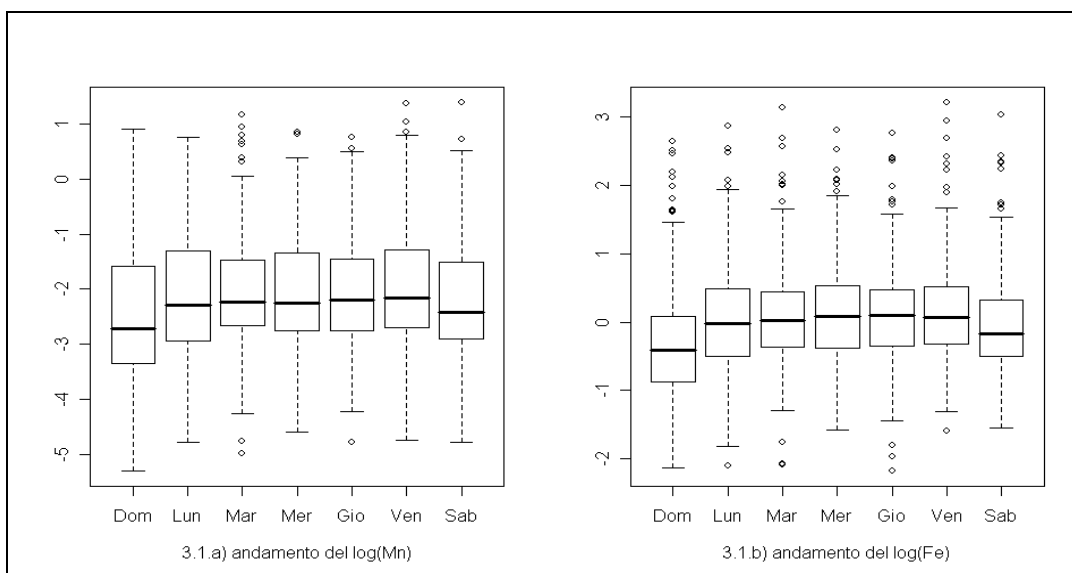


Figura 3.2 - Andamento del Mn e del Fe suddivisi per i giorni della settimana

È importante dire che questo è in contrasto con l'andamento settimanale delle quantità di acciaio spillato, che mediamente hanno valori più alti nel week-end.

Relativamente alle variabili di produzione, va ricordato che *spillato1* e *spillato2* sono altamente correlate con *funzione* e quindi il loro utilizzo in simultanea nei modelli crea problemi di collinearità.

L'elenco completo delle variabili utilizzate è visibile in Appendice A.2.

### 3.3 - I modelli binomiali

Il primo modello ad essere stimato è quello binomiale, dopo aver creato la variabile X, precedentemente definita, associata a ciascun metallo.

La distribuzione di tale variabile è sotto riportata nelle Tabelle 3.1. Dalle queste tabelle si possono dare dei primi commenti.

Innanzitutto è da mettere in luce il fatto che il Manganese ha tutte le osservazioni al di sopra della soglia.

Il Cromo ha quasi il 50% dei dati al di sotto della soglia: questo conferma quanto emerge dai colloqui con i tecnici dell'ARPA FVG e cioè che i livelli del Cromo sono così bassi che le apparecchiature fanno difficoltà a misurarlo.

Tabella 3.1 – Distribuzione delle X a seconda del metallo

<b>X_Mn</b>	<b>X_Fe</b>	<b>X_Ni</b>
0   1	0   1	0   1
0   1462	13   1449	458   1004
<b>X_Cd</b>	<b>X_Zn</b>	<b>X_Pb</b>
0   1	0   1	0   1
50   1412	91   1371	82   1380
<b>x_Cr</b>	<b>X_V</b>	<b>X_As</b>
0   1	0   1	0   1
678   784	161   1300	192   670

Anche il Nichel ha un buon numero di osservazioni sotto la soglia: in questo caso la numerosità di dati al di sotto della soglia arriva al 31%.

Per gli altri metalli, il numero è veramente esiguo.

Ciò fa sì che la stima dei modelli binomiali risulti molto difficoltosa, se non critica. L'unico metallo per cui non c'è stato tale problema è proprio il Nichel. Per questo motivo, vengono riportati solamente i risultati del modello riguardante il Nichel.

In questo modello, si è riscontrato che in questo caso le variabili esplicative significative sono sostanzialmente quelle scelte per il modello lineare del  $\log(Z_{Ni})$ .

Il modello per lo Zinco che è stato stimato è il seguente:

$$[x_{Ni}]_t = \text{probit} \{ \beta_0 + \beta_1(wd_t) + \beta_2(fz_t) + \beta_3 \cdot v1p.d_t + \beta_4 \cdot v1o.d_t + \beta_5 \cdot um.d_t + \beta_6 \cdot \text{press}.d_t + \beta_7 \cdot \text{spillato}2_t + \beta_8(fz_t) \cdot v1p.d_t \}$$

dove:

- $\beta_1(wd_t) = \beta_{1.1} \cdot I(wd_t = \text{"Lun"}) + \beta_{1.2} \cdot I(wd_t = \text{"Mar"}) + \dots + \beta_{1.6} \cdot I(wd_t = \text{"Sab"})$ , il valore base della variabile è "Dom" e  $I(\cdot)$  è la funzione indicatrice;
- $\beta_2(fz_t) = \beta_{2.1} \cdot I(fz_t = 1) + \beta_{2.2} \cdot I(fz_t = 2) + \beta_{2.3} \cdot I(fz_t = 3)$  e il valore base è 0;
- $\beta_{10}(fz_t) \cdot v1p.d_t = \beta_{10.1} \cdot I(fz_t = 1) \cdot v1p.d_t + \dots + \beta_{10.3} \cdot I(fz_t = 3) \cdot v1p.d_t$  e rappresenta l'interazione tra la variabile funzione e la variabile  $v1p.d$ , il valore base per  $fz$  rimane sempre 0;

Le stime di questo modello sono le seguenti:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	-41.960264	10.271954	-4.085	4.41e-05
wdGio	1.926695	0.302987	6.359	2.03e-10
wdLun	1.153571	0.291214	3.961	7.46e-05
wdMar	2.064151	0.308735	6.686	2.30e-11
wdMer	1.811546	0.299369	6.051	1.44e-09
wdSab	1.033179	0.287822	3.590	0.000331
wdVen	1.847906	0.300128	6.157	7.41e-10
fz1	0.149559	0.474649	0.315	0.752691
fz2	0.465010	0.666831	0.697	0.485588
fz3	0.645700	0.234899	2.749	0.005981
v1p.d	0.051280	0.134394	0.382	0.702786
v1o.d	0.169101	0.077473	2.183	0.029058

um.d	0.018024	0.004513	3.994	6.51e-05
press.d	0.039249	0.010217	3.841	0.000122
fz1:v1p.d	0.620440	0.374228	1.658	0.097333
fz2:v1p.d	0.999831	0.678354	1.474	0.140506
fz3:v1p.d	0.470858	0.151396	3.110	0.001870

Qui sopra sono indicate le stime e lo standard error dei vari coefficienti, il valore della Student sotto l'ipotesi nulla che la stima sia pari a 0 e il corrispondente p-value. Inoltre per "fz1:v1p.d" si intende il coefficiente di interazione tra la variabili v1p.d e fz, quando quest'ultima assume valore 1.

Nel modello esposto si vede che tra le variabili selezionate c'è sia la variabile *giorno della settimana* che *funzione*. Quest'ultima è l'unica variabile a indicare un eventuale convolgimento dell'attività siderurgica nella modellazione: qui risulta significativa solamente la differenza tra quando l'impianto opera con entrambe i forni. Tra le variabili meteorologiche sono state scelte il vento diurno, l'umidità diurna e la pressione diurna.

### **3.4 - I modelli lineari con variabili contemporanee**

Il passo successivo alla ricerca di un modello binomiale per la variabile X è stato quello di cercare di stimare un modello lineare per le variabili Z.

Da questa fase in poi, si è sempre dovuto tener conto delle correlazioni tra le variabili esplicative, per cercare di ovviare al problema di multicollinearità.

In prima battuta si è proceduto all'inserimento progressivo nei modelli delle variabili esplicative, sia meteorologiche che di produzione, aggiungendo ad ogni passo la variabile che più migliorava la prestazione del modello, in termini di varianza spiegata (metodo *forward*).

Con questo metodo, le variabili di produzione venivano inserite immediatamente, mentre la scelta delle variabili meteorologiche nei modelli avveniva in modo disordinato.

Per questo motivo, la stima di tali modelli è stata rifatta inserendo, in una fase iniziale, le sole variabili meteorologiche, con il metodo *forward* sopra introdotto; solo dopo si sono valutati i modelli con l'inclusione tutte le combinazioni possibili degli indicatori di produzione (essendo 3 variabili: *spillato1*, *spillato2* e *funzione*, le combinazioni possibili sono solo 8).

Così facendo si è prodotto un risultato interessante: si possono distinguere tre gruppi di metalli, a seconda delle esplicative scelte. Questi gruppi sono proprio quelli illustrati nel capitolo 2, dove si differenziano i metalli a seconda della loro provenienza. Inoltre, lo stesso raggruppamento viene riprodotta molto bene anche dalla matrice di correlazione di questi elementi.

Di seguito, quindi, verranno esposti i modelli suddivisi per ogni gruppo evidenziatosi.

All'interno di ogni gruppo, sono stati stimati altri modelli, in cui viene inserito un elemento del gruppo come variabile esplicativa.

### 3.4.1 - Manganese e Ferro

I due modelli che adesso vengono esposti differiscono tra loro di una sola variabile: la *pioggia*, che è stata inserita solo per il Manganese. Le altre variabili considerate sono *funzione*, la temperatura diurna, l'umidità diurna, il vento diurno e notturno. Inoltre, c'è un'interazione tra il vento diurno parallelo e la variabile funzione, cioè a seconda del tipo del valore assunto dalla variabile funzione, il coefficiente per il vento diurno cambia.

Il modello per il Manganese che è stato stimato è:

$$\log([Mn]_t) = \beta_0 + \beta_1(wd_t) + \beta_2(fz_t) + \beta_3 \cdot v1p \cdot d_t + \beta_4 \cdot v1o \cdot d_t + \beta_5 \cdot v1p \cdot n_t + \beta_6 \cdot v1o \cdot n_t + \beta_7 \cdot temp \cdot n_t + \beta_8 \cdot um \cdot d_t + \beta_9 \cdot pioggia_t + \beta_{10}(fz_t) \cdot v1p \cdot d_t + \varepsilon_t$$

Le stime di tale modello sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.512168	0.157491	-15.951	< 2e-16
wdGio	0.401763	0.099315	4.045	5.67e-05
wdLun	0.360703	0.100163	3.601	0.000334

wdMar	0.468803	0.099323	4.720	2.74e-06
wdMer	0.452602	0.099418	4.552	6.03e-06
wdSab	0.321328	0.099940	3.215	0.001350
wdVen	0.546418	0.098758	5.533	4.13e-08
fz1	0.577755	0.160836	3.592	0.000346
fz2	0.417558	0.229438	1.820	0.069104
fz3	0.759009	0.086064	8.819	< 2e-16
v1p.d	0.060757	0.052068	1.167	0.243572
v1o.d	0.106253	0.027545	3.857	0.000123
v1p.n	0.029616	0.030104	0.984	0.325491
v1o.n	0.160787	0.042057	3.823	0.000141
temp.n	-0.011886	0.003734	-3.183	0.001507
um.d	-0.011849	0.001626	-7.289	6.85e-13
pioggia	0.007364	0.002578	2.857	0.004375
fz1:v1p.d	0.263150	0.081532	3.228	0.001294
fz2:v1p.d	0.430857	0.118194	3.645	0.000282
fz3:v1p.d	0.348251	0.052807	6.595	7.26e-11

A questo modello è stato calcolato un  $R^2$  pari a 0.523.

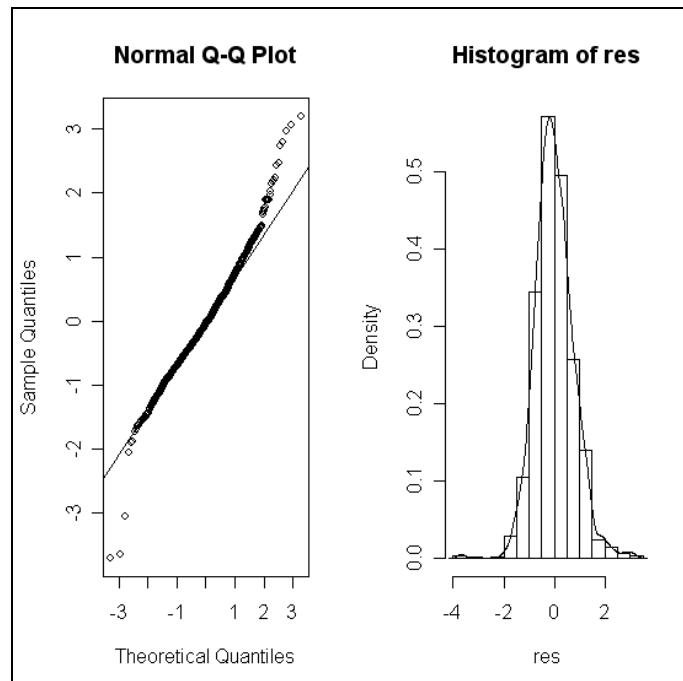


Figura 3.3 - Andamento dei residui del modello del Mn

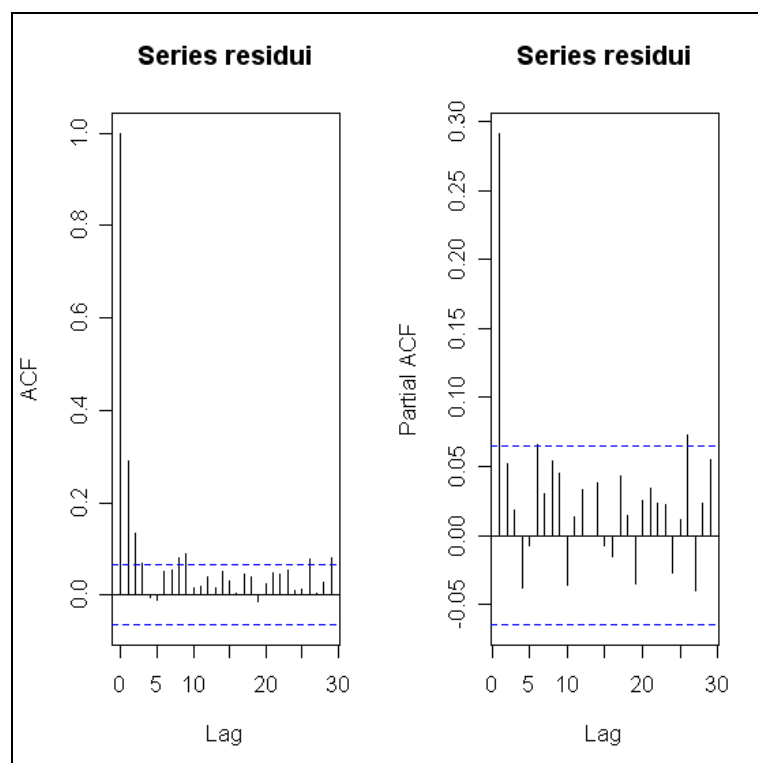


Figura 3.4 - Funzioni di autocorrelazione dei residui del Mn

I residui del modello sono asimmetrici (l'indice è pari a 0.2357) e la curtosi è pari a 4.9220. I valori di questi indici indicano che la distribuzione dei residui sia non gaussiana. La non normalità dei residui è dovuta essenzialmente alle code della loro distribuzione (Figura 3.3): questo, molto plausibilmente, è dovuto al fatto che il modello lineare fatica a spiegare i valori estremi del Manganese.

I residui del modello, inoltre, sono autocorrelati, mostrando le caratteristiche di una serie autoregressiva del primo ordine (Figura 3.4).

Il modello del Ferro, invece, è il seguente:

$$\log([\text{Fe}]_t) = \beta_0 + \beta_1(\text{wd}_t) + \beta_2(\text{fz}_t) + \beta_3 \cdot \text{v1p.d}_t + \beta_4 \cdot \text{v1o.d}_t + \beta_5 \cdot \text{v1p.n}_t + \beta_6 \cdot \text{v1o.n}_t + \beta_7 \cdot \text{temp.n}_t + \beta_8 \cdot \text{um.d}_t + \beta_9(\text{fz}_t) \cdot \text{v1p.d}_t + \varepsilon_t$$

Le stime dei beta soli risultano i seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.125063	0.121548	-1.029	0.303794
wdGio	0.448775	0.079101	5.673	1.89e-08



wdLun	0.376462	0.079854	4.714	2.81e-06
wdMar	0.485934	0.079088	6.144	1.21e-09
wdMer	0.478680	0.079208	6.043	2.21e-09
wdSab	0.310612	0.079662	3.899	0.000104
wdVen	0.523072	0.078772	6.640	5.42e-11
fz1	0.240620	0.127742	1.884	0.059938
fz2	0.180877	0.181979	0.994	0.320517
fz3	0.309943	0.068515	4.524	6.89e-06
v1p.d	-0.080314	0.041464	-1.937	0.053062
v1o.d	0.046418	0.021766	2.133	0.033230
v1p.n	0.027901	0.023837	1.171	0.242107
v1o.n	0.128317	0.033329	3.850	0.000127
temp.n	-0.013666	0.002926	-4.670	3.47e-06
um.d	-0.007302	0.001173	-6.225	7.38e-10
fz1:v1p.d	0.241010	0.064877	3.715	0.000216
fz2:v1p.d	0.380546	0.093974	4.049	5.58e-05
fz3:v1p.d	0.321723	0.042212	7.622	6.38e-14

In questo caso,  $R^2$  è stimato a circa 0.409. Come per il Manganese, anche i residui di questo modello non hanno una distribuzione gaussiana: l'indice di asimmetria è pari a 0.2199 e quello di curtosi è pari a 6.1918 (Figura 3.5).

I residui sono autocorrelati e sembra che il loro andamento sia di tipo autoregressivo del primo o al massimo secondo ordine (Figura 3.6).

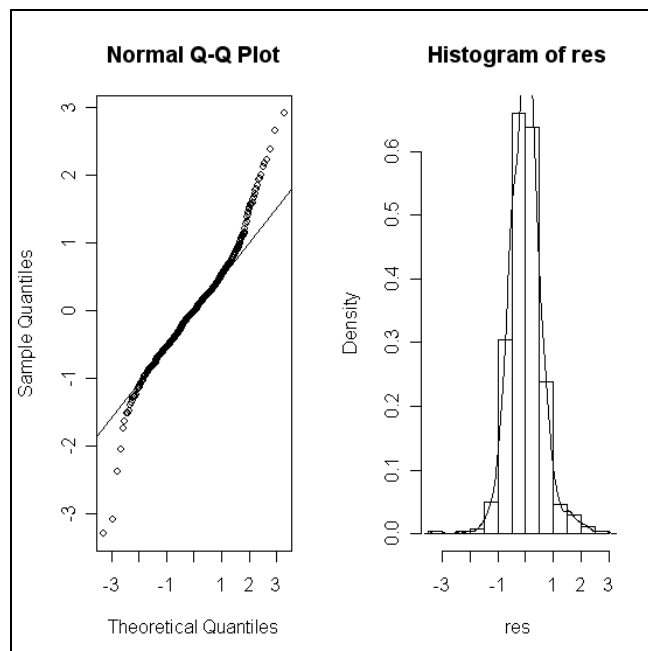


Figura 3.5 – Andamento dei residui nel modello del Fe

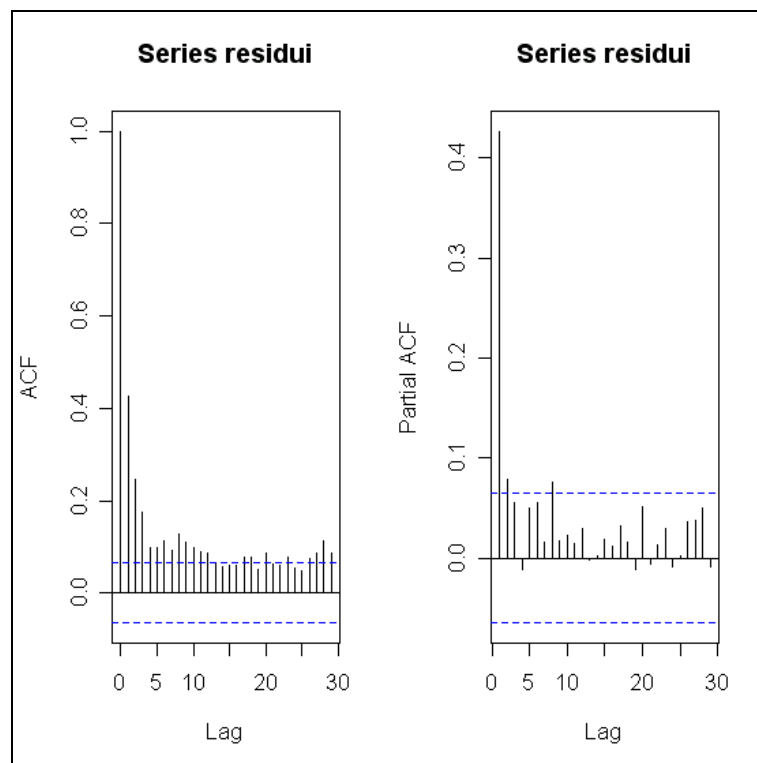


Figura 3.6 - Funzioni di autocorrelazione del modello Fe

All'interno dello stesso gruppo è importante capire anche la relazione che esiste tra gli stessi metalli. Per questo motivo si è considerato un modello tra i due metalli, in cui la variabile esplicativa è il Manganese, mentre quella risposta è il Ferro.

$$\log([\text{Fe}]_t) = \beta_0 + \beta_1 \cdot \log([\text{Mn}]_t) + \varepsilon_t$$

Il risultato di stima è il seguente:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.395481	0.021114	66.09	<2e-16
log(z_Mn)	0.647569	0.008999	71.96	<2e-16

In questo modello  $R^2$  è molto alto: la varianza spiegata è il 0.782 della varianza totale.

In questo caso i residui hanno un indice di simmetria molto più vicino allo zero (è pari a 0.0336) e le code della sua distribuzione sono meno pesanti (curtosi pari a 3.6362).

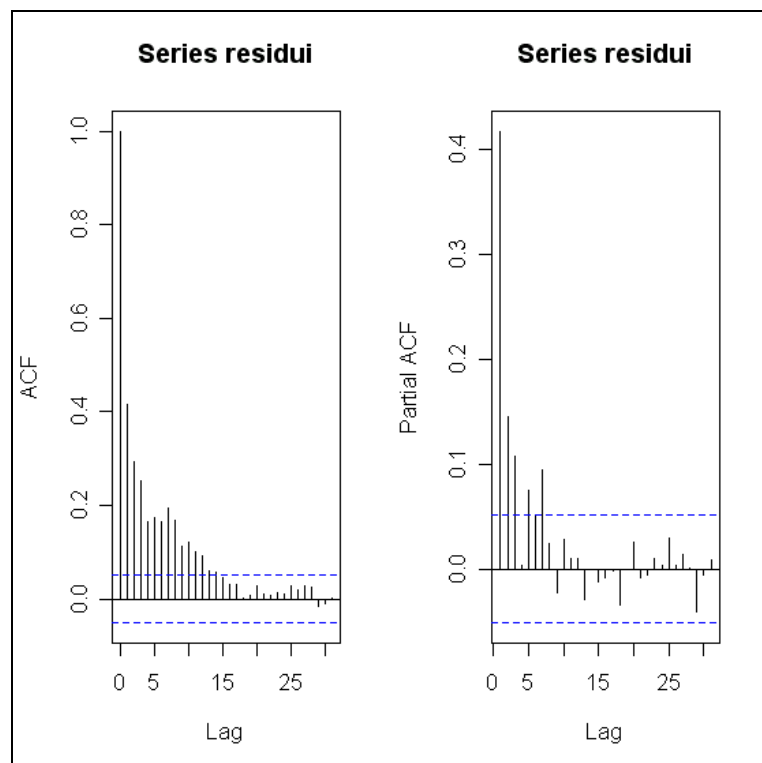


Figura 3.7 - Funzioni di autocorrelazione dei residui del modello Fe2

Nonostante questi siano dei risultati migliori rispetto ai due modelli illustrati precedentemente, l'autocorrelazione dei residui rimane (Figura 3.7).

### 3.4.2 - Nichel

Il Nichel è il metallo, tra quelli modellati, meno correlato con gli altri. Coerentemente si è riscontrato in fase di stima dei modelli, che la scelta delle variabili esplicative differiva di molto rispetto agli altri gruppi.

In questo caso le variabili scelte sono solamente il vento diurno e la pressione diurna:

$$\log([\text{Ni}]_t) = \beta_0 + \beta_1 \cdot v1p.d_t + \beta_2 \cdot v1o.d_t + \beta_3 \cdot \text{press}.d_t + \varepsilon_t$$

I valori dei coefficienti stimati sono riportati nella tabella seguente.

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-15.457609	2.456352	-6.293	4.76e-10
vlp.d	0.175218	0.012259	14.293	< 2e-16
vlo.d	0.007091	0.019369	0.366	0.714
press.d	0.010659	0.002451	4.349	1.52e-05

Come si può vedere, nel modello non ci sono le variabili di produzione. Questo è l'unico metallo in cui il modello rigetta tali variabili.

La distribuzione degli errori è decisamente asimmetrica (l'indice è pari a 1.4503) e ha delle code pesanti (curtosi pari a 6.9745). Ciò si può vedere dalla Figura 3.8. Come per il gruppo precedente gli errori sono comunque correlati (Figura 3.9).

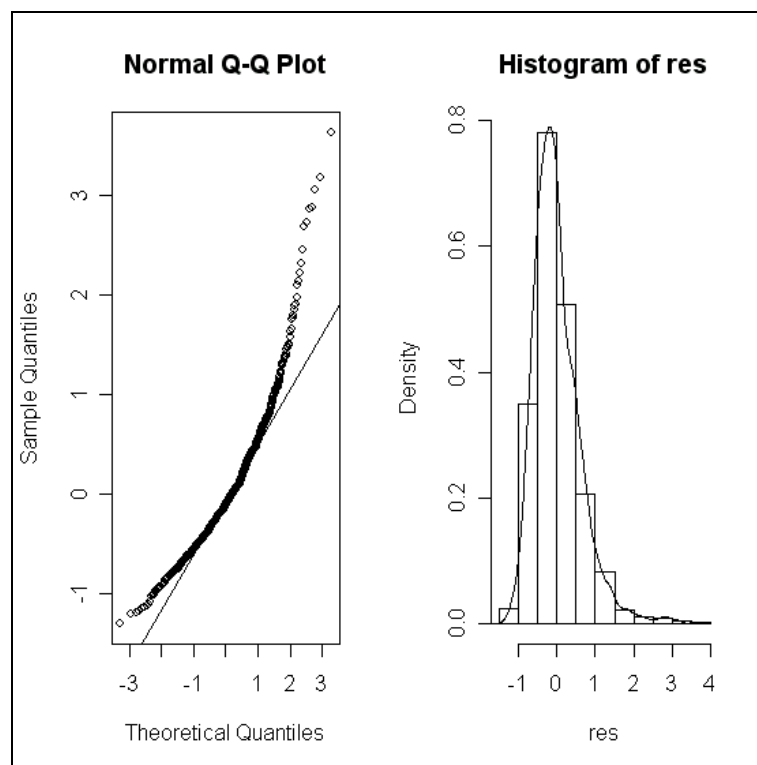


Figura 3.8 - Normal Q-Q plot e istogramma dei residui del modello del Nichel

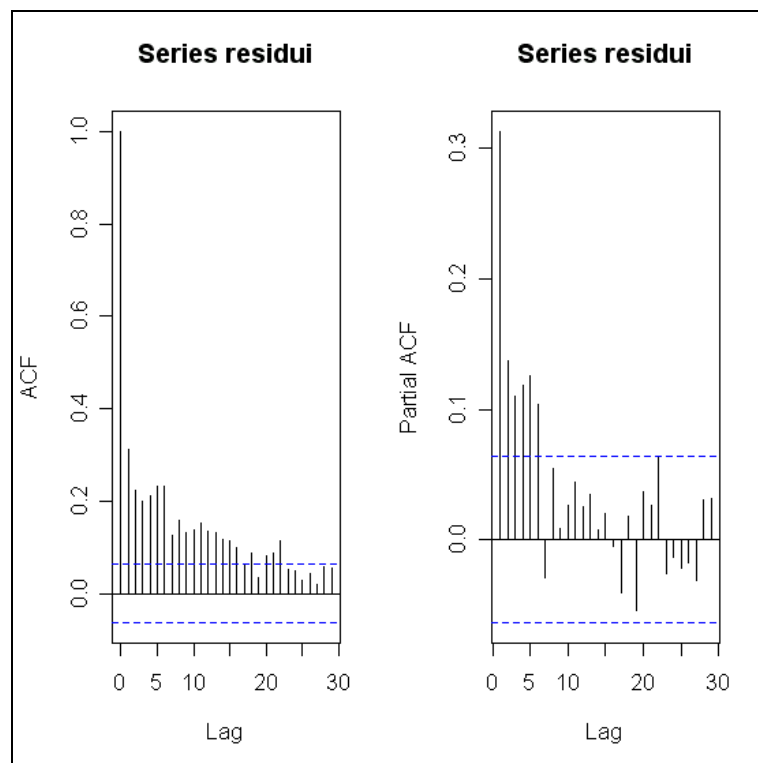


Figura 3.9 – Funzioni di autocorrelazione del modello Ni

### 3.4.3 - Cadmio, Piombo e Zinco

Il Cadmio, il Piombo e lo Zinco sono modellati da un gruppo di variabili che è comune; per ognuno di loro si prevede poi un'ulteriore variabile esplicativa, che è distintiva rispetto agli altri. Le variabili scelte in comune, tra quelle meteorologiche, sono il vento diurno e notturno e la temperatura notturna; fra quelle di produzione, oltre alla variabile *funzione*, è stata inserita anche lo *spillato1*. Come detto, questa scelta comporta un problema di collinearità e ciò si è visto nell'effetto di compensazione delle stime dei coefficienti di tali variabili e in un aumento del loro errore standard.

Le variabili distintive, invece, sono la pressione diurna per il Cadmio, l'umidità diurna per il Piombo e l'umidità notturna per la Zinco.

Di seguito sono riportati i risultati dei modelli.

Il primo è quello del Cadmio.

$$\log([\text{Cd}]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2(\text{fz}_t) + \beta_3 \cdot \text{v1p.d}_t + \beta_4 \cdot \text{v1o.d}_t + \beta_5 \cdot \text{v1p.n}_t + \beta_6 \cdot \text{v1o.n}_t + \beta_7 \cdot \text{press.d}_t + \beta_8 \cdot \text{temp.n}_t + \beta_{10}(\text{fz}_t) \cdot \text{v1p.d}_t + \varepsilon_t$$

e le stime sono:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-21.288961	3.920073	-5.431	7.29e-08
spillato1	0.736535	0.180408	4.083	4.87e-05
fz1	-0.036298	0.218079	-0.166	0.867845
fz2	0.369532	0.248124	1.489	0.136770
fz3	0.460863	0.143229	3.218	0.001340
v1p.d	-0.017904	0.059121	-0.303	0.762084
v1o.d	0.136954	0.029970	4.570	5.59e-06
v1p.n	0.029729	0.032358	0.919	0.358474
v1o.n	0.186683	0.045938	4.064	5.27e-05
press.d	0.013284	0.003899	3.407	0.000687
temp.n	-0.018520	0.004422	-4.188	3.11e-05
fz1:v1p.d	0.275241	0.089824	3.064	0.002250
fz2:v1p.d	0.473323	0.128175	3.693	0.000236
fz3:v1p.d	0.360268	0.060489	5.956	3.76e-09

Il modello ha un  $R^2$  pari a 0.45.

Come per gli altri metalli, i residui non hanno una distribuzione gaussiana, dovuta soprattutto al fatto che il modello non riesce a stimare in modo buono i valori estremi. Inoltre, dalle funzioni di autocorrelazione risultano essere di tipo autoregressivo del primo ordine. (Figure 3.10 e 3.11).

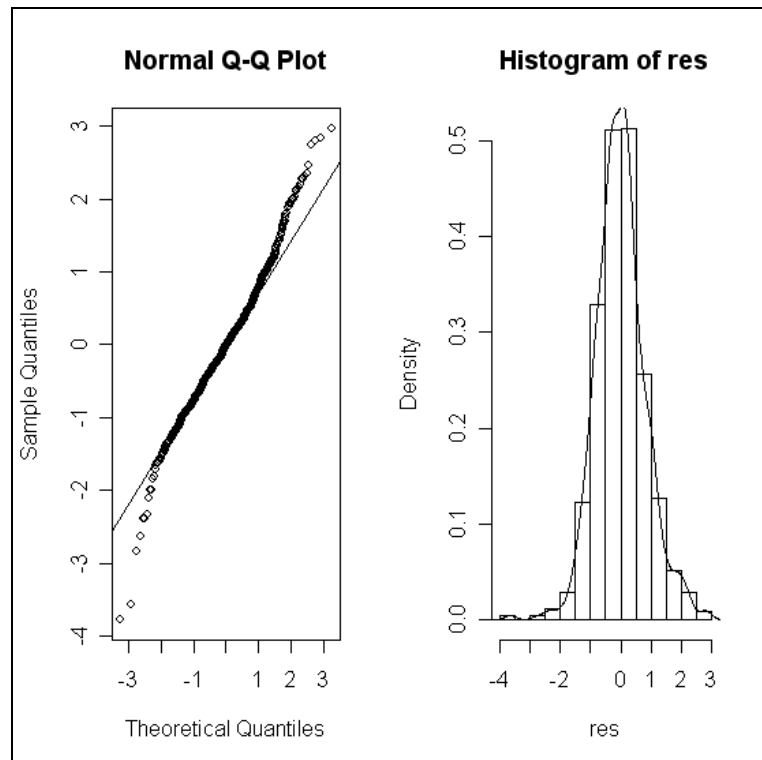


Figura 3.10 - Normal Q-Q plot e istogramma dei residui del modello Cd

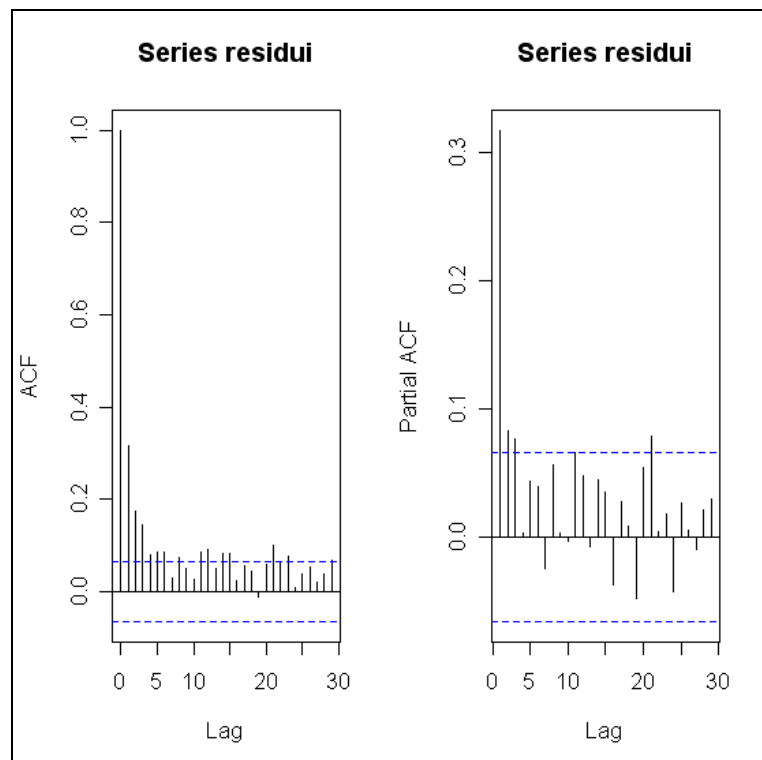


Figura 3.11 - Funzioni di autocorrelazione dei residui del modello Cd

Di seguito sono riportati il modello del Piombo e quello dello Zinco.

Il modello del Piombo risulta:

$$\log([\text{Pb}]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2(\text{fz}_t) + \beta_3 \cdot \text{v1p.d}_t + \beta_4 \cdot \text{v1o.d}_t + \beta_5 \cdot \text{v1p.n}_t + \beta_6 \cdot \text{v1o.n}_t + \beta_7 \cdot \text{temp.n}_t + \beta_8 \cdot \text{um.d}_t + \beta_9(\text{fz}_t) \cdot \text{v1p.d}_t + \varepsilon_t$$

Di tale modello i valori dei coefficienti  $\beta_i$  stimati sono i seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.281548	0.159262	-20.605	< 2e-16
spillato1	0.470040	0.190491	2.468	0.013794
fz1	0.148520	0.232215	0.640	0.522612
fz2	0.485342	0.264001	1.838	0.066338
fz3	0.530496	0.152217	3.485	0.000516
v1p.d	0.006522	0.061308	0.106	0.915300
v1o.d	0.096238	0.031662	3.040	0.002439
v1p.n	0.068621	0.034514	1.988	0.047097
v1o.n	0.182507	0.048487	3.764	0.000178
temp.n	-0.013880	0.004261	-3.257	0.001168
um.d	-0.006731	0.001708	-3.940	8.81e-05
fz1:v1p.d	0.307357	0.094554	3.251	0.001195
fz2:v1p.d	0.499464	0.135747	3.679	0.000248
fz3:v1p.d	0.406980	0.062730	6.488	1.45e-10

La varianza spiegata dal modello è quasi la metà di quella totale: l'indice  $R^2$  si aggira sul 45%.



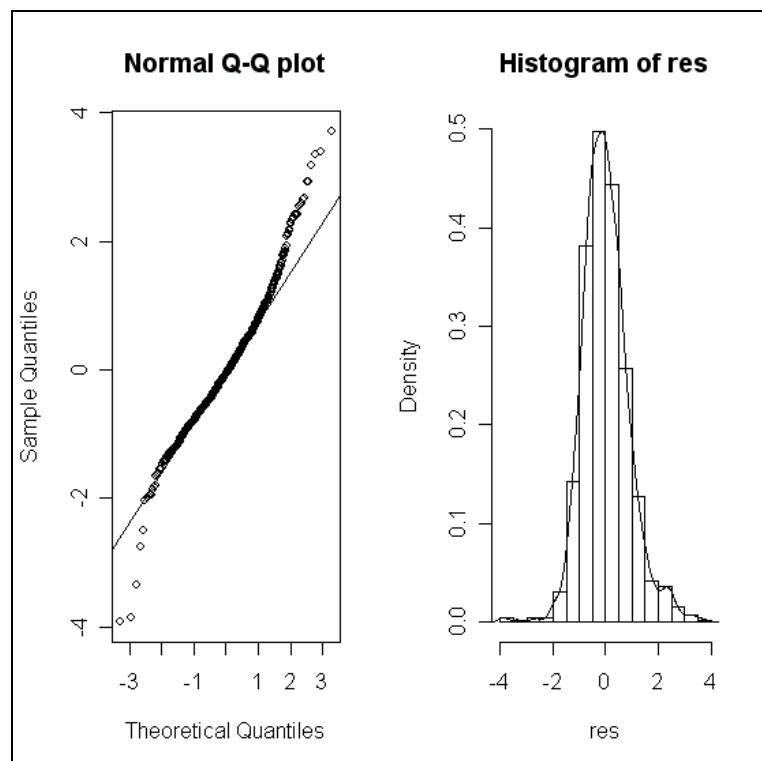


Figura 3.12 – Normal Q-Q plot e istogramma dei residui del modello Pb

Il modello dello Zinco, invece, è:

$$\log([Zn]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2(fz_t) + \beta_3 \cdot v1p.d_t + \beta_4 \cdot v1o.d_t + \beta_5 \cdot v1p.n_t + \beta_6 \cdot v1o.n_t + \beta_7 \cdot \text{temp}.n_t + \beta_8 \cdot \text{um}.n_t + \beta_9(fz_t) \cdot v1p.d_t + \varepsilon_t$$

Le stime sono risultate:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.841577	0.214938	-3.915	9.75e-05
spillato1	0.358098	0.180000	1.989	0.04697
fz1	0.414454	0.227969	1.818	0.06941
fz2	0.446169	0.249838	1.786	0.07448
fz3	0.395802	0.151621	2.610	0.00920
v1p.d	0.000539	0.060926	0.009	0.99294
v1o.d	0.153294	0.029729	5.156	3.14e-07
v1o.n	0.146880	0.045468	3.230	0.00128
v1p.n	0.017015	0.034338	0.496	0.62035
temp.n	-0.021796	0.004127	-5.282	1.63e-07
um.n	-0.011545	0.002260	-5.110	3.99e-07
fz1:v1p.d	0.190051	0.090832	2.092	0.03671
fz2:v1p.d	0.507688	0.127781	3.973	7.70e-05
fz3:v1p.d	0.405913	0.061906	6.557	9.53e-11

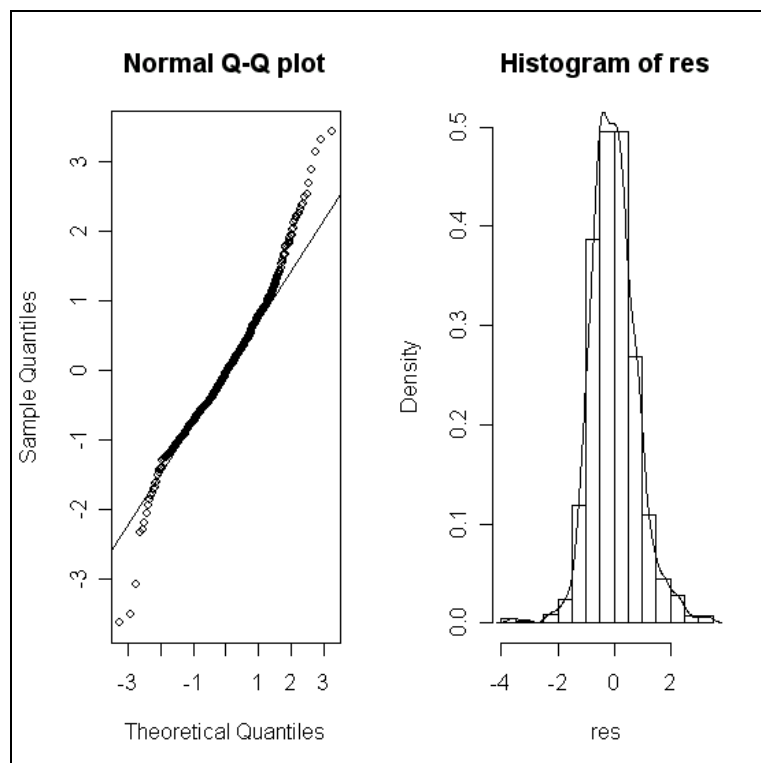


Figura 3.13 – Normal Q-Q plot e istogramma dei residui del modello Zn

L'indice  $R^2$  di questo modello è leggermente più alto rispetto agli altri due modelli: qui è pari a circa 0.495

I residui di entrambe i modelli sono asimmetrici (l'indice è pari a 0.4197 per il modello del Piombo e pari a 0.3621 per quello dello Zinco) ed hanno delle distribuzioni con delle code pesanti (il loro indice di curtosi è di circa 4.85).

Le loro distribuzioni hanno un andamento autoregressivo del primo ordine (si veda le Figure 3.14 e 3.15)

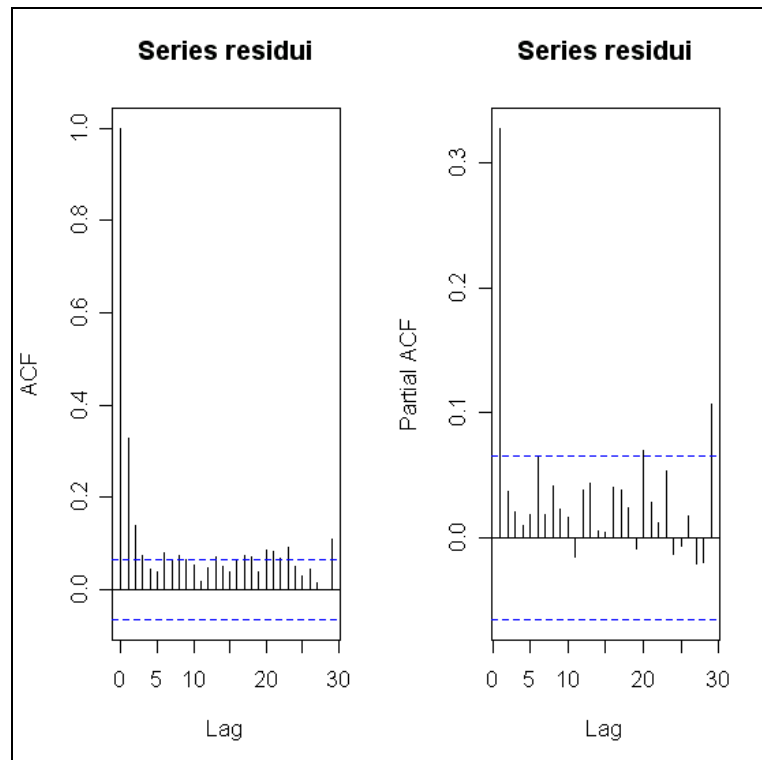


Figura 3.14 - Funzioni di autocorrelazioni dei residui del modello del Pb

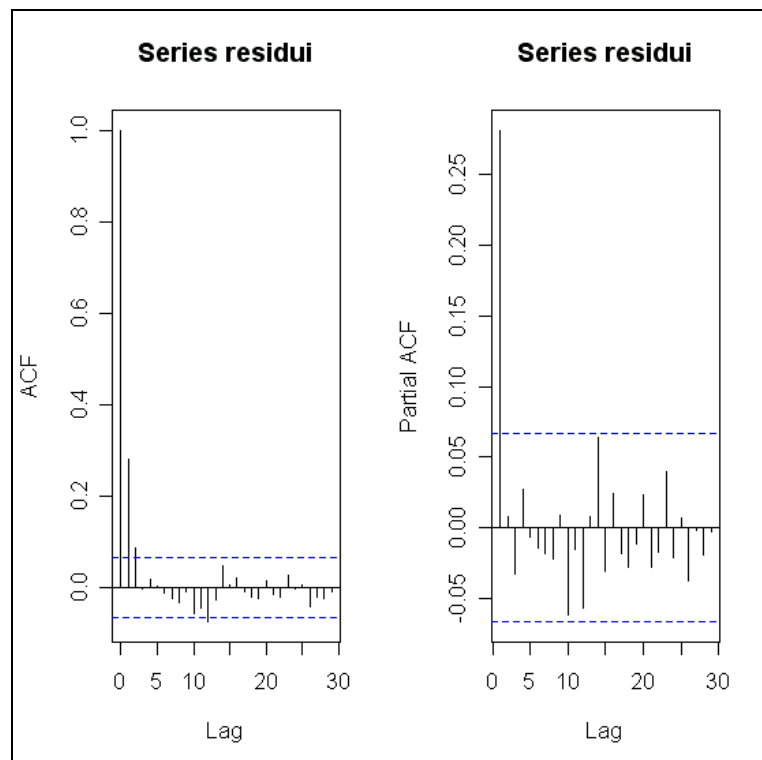


Figura 3.15 - Funzioni di autocorrelazione dei residui del modello Zn

Per questo gruppo, infine, sono stati stimati dei modelli all'interno del gruppo. In questo caso il metallo che è stato scelto come variabile esplicativa è il Cadmio. La varianza spiegata di questi modelli è decisamente superiore a quelli appena presentati; questo è soprattutto dato dal fatto che il Cadmio è altamente correlato con entrambi i restanti metalli (con lo Zinco ha una correlazione pari a 0.76, mentre con il Piombo tale indice è pari a 0.64).

Per il Piombo, le altre variabili inserite a spiegare il metallo sono il vento diurno, l'umidità diurna e la pressione diurna.

$$\log([\text{Pb}]_t) = \beta_0 + \beta_1 \cdot \log([\text{Cd}]_t) + \beta_2 \cdot \text{vlp.d}_t + \beta_3 \cdot \text{v1o.d}_t + \beta_4 \cdot \text{um.d}_t + \beta_5 \cdot \text{press.d}_t + \varepsilon_t$$

Le stime sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11.6122188	2.1184335	5.482	5.09e-08
log(z_Cd)	0.8249665	0.0175906	46.898	< 2e-16
vlp.d	0.1400043	0.0119215	11.744	< 2e-16
v1o.d	0.0144894	0.0155756	0.930	0.352
um.d	-0.0049133	0.0009135	-5.379	8.94e-08
press.d	-0.0082392	0.0020764	-3.968	7.66e-05

Con queste variabili esplicativa l'indice  $R^2$  è pari a 0.764.

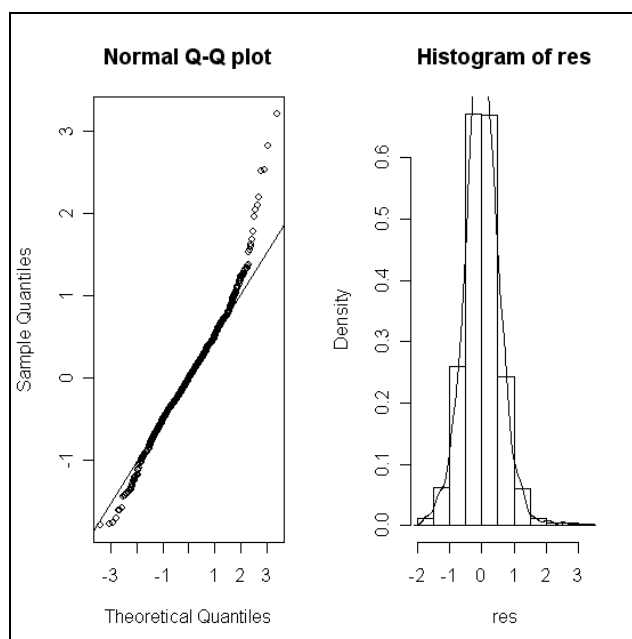


Figura 3.16 – Normal Q-Q plot dei residui del secondo modello Pb

Per lo Zinco, invece, è stato inserito il vento diurno, l'umidità notturna e la temperatura notturna:

$$\log([Zn]_t) = \beta_0 + \beta_1 \cdot \log([Cd]_t) + \beta_2 \cdot v1p.d_t + \beta_3 \cdot v1o.d_t + \beta_4 \cdot um.n_t + \beta_5 \cdot temp.n_t + \varepsilon_t$$

Si sono ottenute le seguenti stime:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.308309	0.182786	23.570	< 2e-16
log(z_Cd)	0.695540	0.020360	34.162	< 2e-16
v1p.d	0.168778	0.014285	11.815	< 2e-16
v1o.d	0.060945	0.018597	3.277	0.001077
um.n	-0.004978	0.001377	-3.616	0.000312
temp.n	-0.007714	0.002810	-2.745	0.006130

Rispetto al modello con il Piombo, questo ha un adattamento leggermente peggiore: qui  $R^2$  è 0.681.

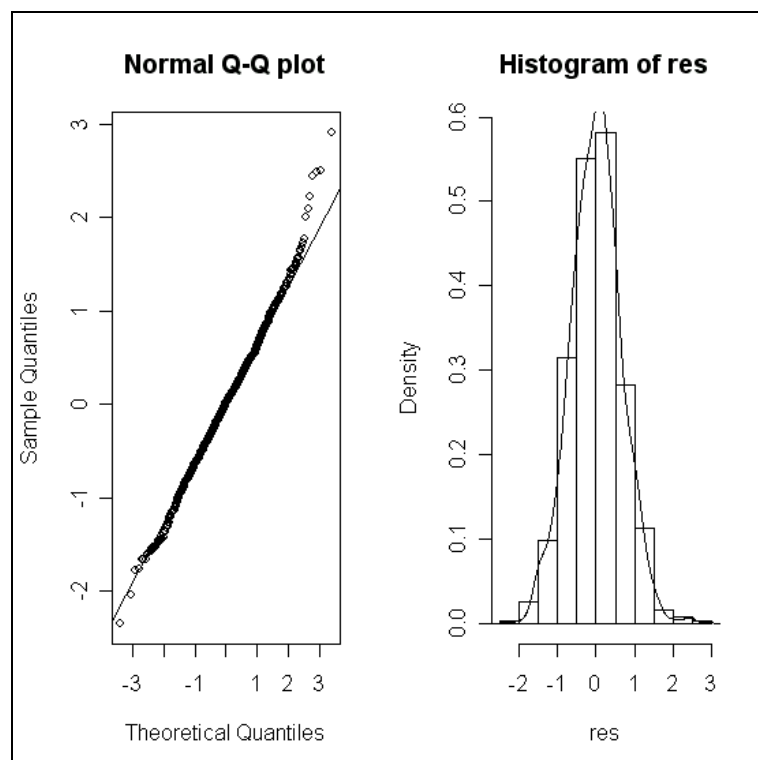


Figura 3.17 – Normal Q-Q plot dei residui del secondo modello Zn

Nonostante questo miglioramento, i residui rimangono autocorrelati; le loro distribuzioni sono sempre asimmetriche e presentano curtosi maggiore di 3 (Figure 3.18 e 3.19).

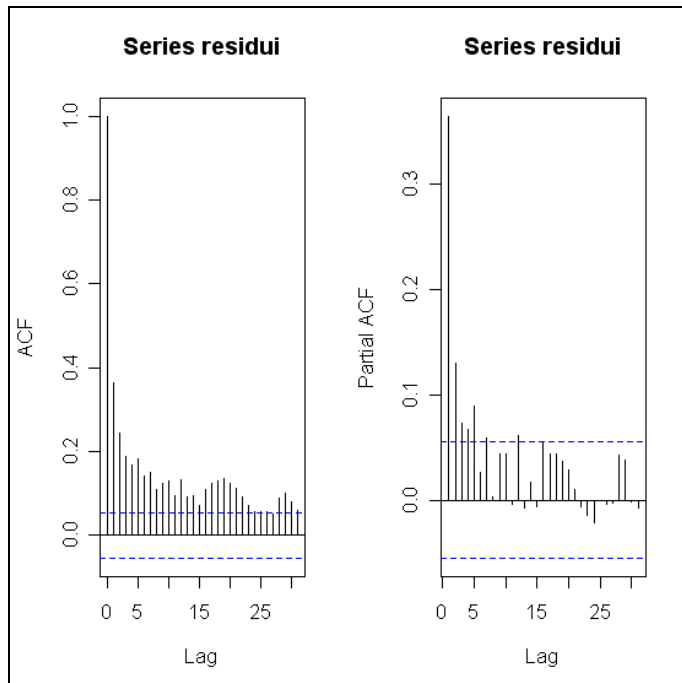


Figura 3.18 - Funzioni di autocorrelazione dei residui del secondo modello Pb

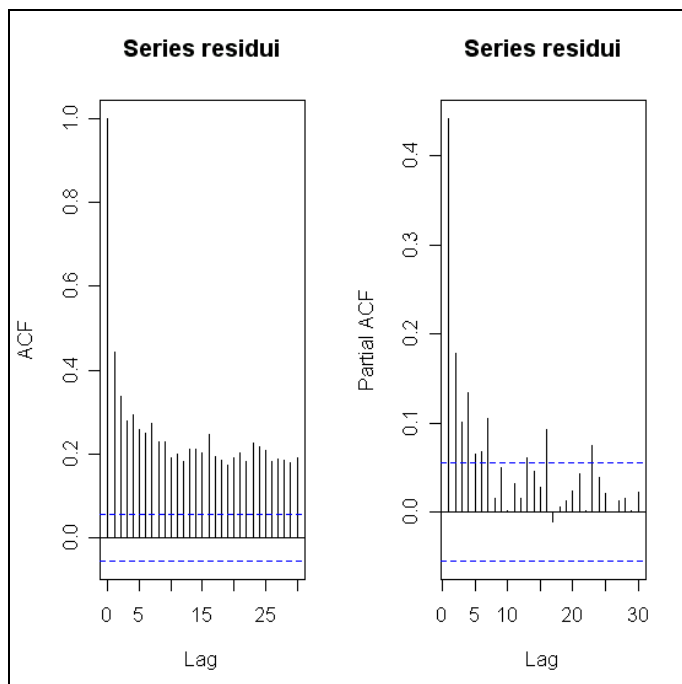


Figura 3.19 - Funzioni di autocorrelazione dei residui del secondo modello Zn

### **3.5 - Modelli con variabili esplicative ritardate**

I modelli appena rappresentati hanno tutti un problema: gli errori sono correlati.

Questo problema può provenire anche dal fatto che i metalli stessi hanno un andamento autoregressivo che generalmente è del primo ordine.

Si è cercato, quindi, di pervenire ad una soluzione utilizzando sempre un modello lineare. Tra le varie opzioni a disposizione, quella che è stata applicata è l'utilizzo delle variabili esplicative ritardate di un periodo.

Si ricorda che le medie notturne di un certo giorno al tempo  $t$  vengono calcolate considerando i dati che vanno dalle ore 20 del giorno  $t$  alle ore 7 del giorno al tempo  $t+1$ . Per questo motivo ci si può aspettare che i modelli preferiscano le variabili notturne del giorno prima, rispetto a quelle odierne proprio perché le medie notturne del tempo  $t-1$  vengono calcolate considerando un maggior numero di osservazioni che appartengono al tempo  $t$ .

A parte il vento, che non è autocorrelato, le variabili meteorologiche ritardate possono creare dei problemi di collinearità nel modello stesso e quindi in caso di utilizzo di ritardate nel modello si è dovuto tenere conto di tale effetto.

Anche qui si è proceduto all'inserimento prima delle variabili meteo con un metodo forward e solo dopo si sono valutati tutte le possibilità dei di inserimento delle variabili di produzione.

Alla fine, anche gli errori di questi secondi modelli sono autocorrelati anche se l'inserimento di variabili ritardate ha apportato dei miglioramenti.

Per tali motivi si ritenuto il caso tenere in considerazione questi ultimi modelli e passare poi alla stima della struttura di correlazione dei vari errori.

Inoltre anche in questo caso la distinzione tra i tre gruppi è rimasta.

Vengono qui riportate le stime di tali modelli prima e dopo la stima della struttura di autocorrelazione dell'errore.

### 3.5.1 - Nichel

Il Nichel continua a volere poche variabili esplicative: in questo caso insieme al vento diurno si è inserita la pressione notturna del giorno prima. L'errore viene modellato solo nella parte non stagionale. Anche se nella funzione di autocorrelazione (Acf) rimangono dei valori che superano la soglia, possiamo considerarli dei valori non influenti. Possiamo dire ciò in quanto riducendo la varianza non spiegata nel modello, le bande di confidenza si sono ristrette e così valori che prima si consideravano nulli adesso sono significativamente diversi da zero. I valori di queste correlazioni di rado superano (e di poco), però, un valore pari 0.1. Quindi queste correlazioni si possono considerare non rilevanti.

Inizialmente il modello per il Nichel è stato stimato come:

$$\log([Ni]_t) = \beta_0 + \beta_1 \cdot v1p.d_t + \beta_2 \cdot v1o.d_t + \beta_3 \cdot press.n_{t-1} + \varepsilon_t$$

Le stime sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-15.738281	2.492465	-6.314	4.17e-10
v1p.d	0.177855	0.012238	14.534	< 2e-16
v1o.d	0.006316	0.019360	0.326	0.744
press.n1	0.010924	0.002484	4.399	1.21e-05

L'indice  $R^2$  è pari a 0.1990 e se si confronta con il modello precedente, questo non si è modificato. Dopo aver effettuato questa stima si passa ai residui, che in questo caso sono stimati come un ARMA(2,1). Le stime del modello finale sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
ar1	1.10781	0.05158	21.4787	0.0000
ar2	-0.13554	0.04444	-3.0496	0.0024
ma1	-0.87050	0.03380	-25.7511	0.0000
(Intercept)	-13.66698	2.98984	-4.5711	0.0000
v1p.d	0.16669	0.01189	14.0190	0.0000
v1o.d	0.02497	0.01802	1.3856	0.1662
press.n1	0.00885	0.00298	2.9705	0.0030



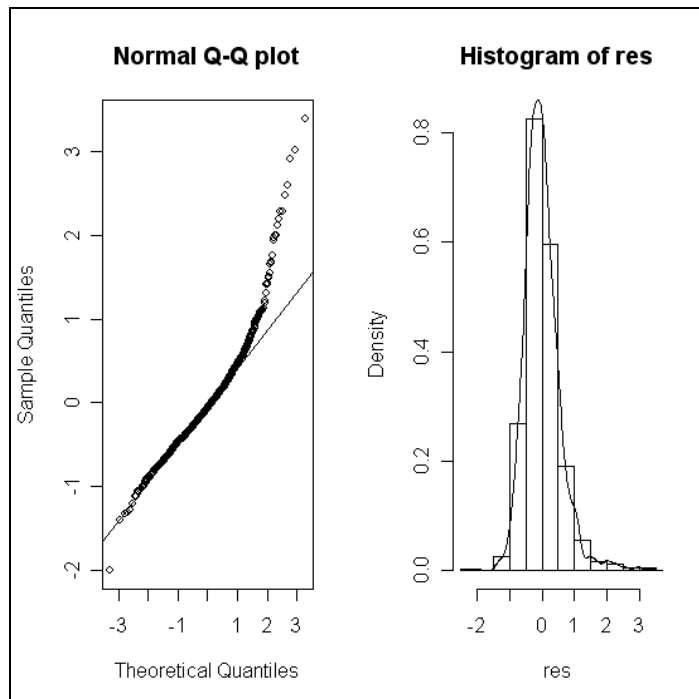


Figura 3.20 -Normal Q-Q plot e istogramma dei residui del modello Ni

I residui del modello finale rimangono asimmetrici (l'indice è pari a 0.57) ma la curtosi si è abbassata 3.26 (si veda Figura 3.20).

In Figura 3.21 sono raffigurate le funzioni di autocorrelazione dei residui del modello finale.

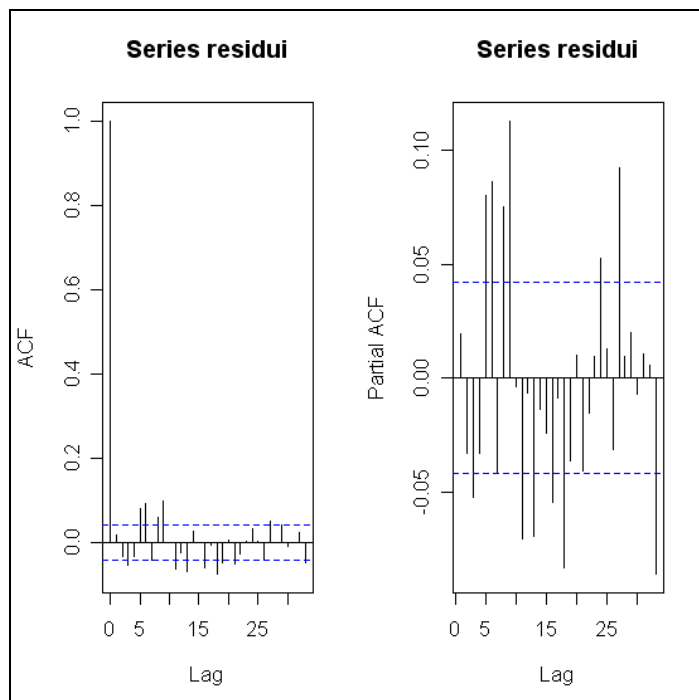


Figura 3.21 - Funzioni di autocorrelazione del modello Ni finale

### 3.5.2 - Manganese e Ferro

I modelli per il Manganese e Ferro hanno molte variabili in comune: i due spillati, i giorni della settimana, il vento diurno e notturno odierni, il vento notturno del giotno precedente e l'umidità notturna del giorno prima. In quello del Ferro è stata inserita la pressione notturna del giorno prima, mentre in quello del Manganese vi è una interazione in più, che è quella tra la variabile funzione e il vento notturno parallelo.

In questi modelli finali, si è effettuata una scelta molto importante sulle variabili di produzione: la variabile funzione viene utilizzata solo per le interazioni e al suo posto si usano i due spillati. Questo perché dopo aver costruito i vari modelli (con le interazioni), confrontando il modello che utilizza *funzione* e quello che usa tutte e due i *spillati*, si vede che questi differiscono di poco e quindi si è preferito il secondo modello.

Il modello del Manganese è stato stimato all'inizio come segue:

$$\log([Mn]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2 \cdot \text{spillato2}_t + \beta_3(wd_t) + \beta_4 \cdot v1p.d_t + \beta_4 \cdot v1o.d_t + \beta_6 \cdot v1p.n_t + \beta_7 \cdot v1o.n_t + \beta_8 \cdot v1p.n_{t-1} + \beta_9 \cdot v1o.n_{t-1} + \beta_{10} \cdot um.n_{t-1} + \beta_{11}(fz_t) \cdot v1p.d_t + \beta_{12}(fz_t) \cdot v1p.n_t + \varepsilon_t$$

Le stime del modello sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-2.346318	0.214109	-10.959	< 2e-16
spillato1	0.669297	0.142706	4.690	3.16e-06
spillato2	0.088749	0.122893	0.722	0.470380
wdGio	0.502398	0.098852	5.082	4.54e-07
wdLun	0.417088	0.098065	4.253	2.33e-05
wdMar	0.491144	0.096914	5.068	4.89e-07
wdMer	0.491141	0.096976	5.065	4.97e-07
wdSab	0.302350	0.096867	3.121	0.001858
wdVen	0.638655	0.098809	6.463	1.68e-10
v1p.d	0.069275	0.056375	1.229	0.219458
v1o.d	0.070194	0.026824	2.617	0.009024
v1p.n	-0.195498	0.068768	-2.843	0.004573
v1o.n	0.113262	0.040740	2.780	0.005548
v1p.n1	0.154409	0.031738	4.865	1.35e-06
v1o.n1	0.163779	0.041121	3.983	7.36e-05

um.n1	-0.010708	0.001941	-5.516	4.54e-08
fz1:v1p.d	0.062284	0.106366	0.586	0.558317
fz2:v1p.d	0.344670	0.125014	2.757	0.005951
fz3:v1p.d	0.261751	0.058576	4.469	8.89e-06
fz1:v1p.n	0.322696	0.140991	2.289	0.022325
fz2:v1p.n	0.383502	0.190186	2.016	0.044051
fz3:v1p.n	0.268585	0.072921	3.683	0.000244

L'indice  $R^2$  è migliorato di oltre 3 punti percentuali, arrivando a 0.5517.

I residui, successivamente, sono stati modellati con una parte autoregressiva del secondo ordine (non stagionale) affiancata da una parte autoregressiva del primo ordine stagionale (il ritardo della stagione è il 7, cioè la settimana).

Inoltre nel modello è stata inserita la variabile  $wd$  solo quando questa assume valore "Sabato" e "Domenica"; in quanto nella fase modellazione precedente si è riscontrato che solo in questi giorni c'è una differenza significativa rispetto agli altri giorni.

Alla fine le stime sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.31531	0.03441	9.1629	0.0000
ar2	0.06848	0.03503	1.9546	0.0509
sar1	0.09477	0.03718	2.5490	0.0110
(Intercept)	-1.98650	0.21414	-9.2768	0.0000
$wd="Sab"$	-0.23316	0.07356	-3.1697	0.0016
$wd="Dom"$	-0.49764	0.07481	-6.6518	0.0000
spillato1	0.49364	0.13614	3.6259	0.0003
spillato2	0.15621	0.12716	1.2284	0.2196
v1p.d	0.06600	0.05190	1.2718	0.2038
v1o.d	0.06430	0.02521	2.5509	0.0109
v1p.n	-0.14030	0.06266	-2.2391	0.0254
v1o.n	0.13923	0.03783	3.6803	0.0002
v1p.n1	0.17452	0.02973	5.8704	0.0000
v1o.n1	0.18380	0.03843	4.7831	0.0000
um.n1	-0.00823	0.00221	-3.7261	0.0002
fz1_v1p.d	0.07231	0.09285	0.7787	0.4364
fz2_v1p.d	0.41202	0.11309	3.6432	0.0003
fz3_v1p.d	0.25468	0.05409	4.7084	0.0000
fz1_v1p.n	0.32018	0.12765	2.5083	0.0123
fz2_v1p.n	0.15459	0.19081	0.8102	0.4180
fz3_v1p.n	0.21365	0.06653	3.2115	0.0014

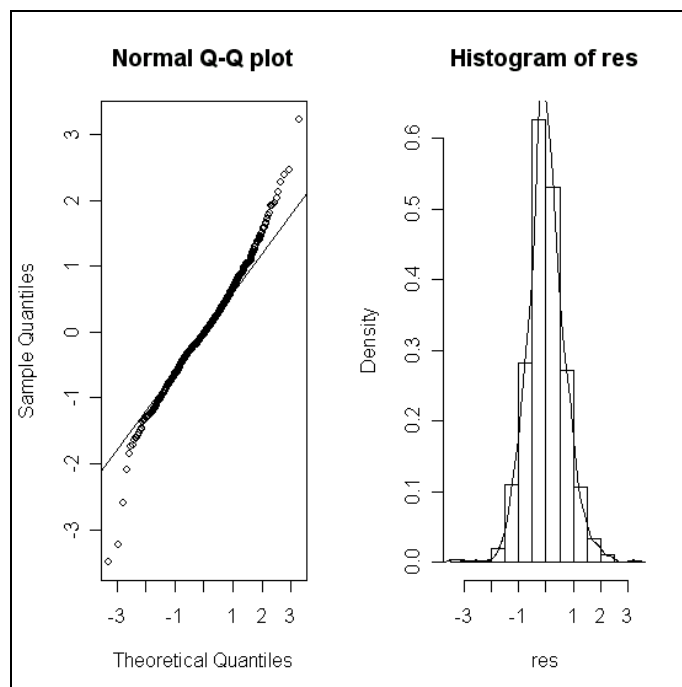


Figura 3.22- Normal Q-Q plot e istogramma dei residui del modello Mn

I residui di questo modello sono asimmetrici e con un indice di curtosi minore di 3 (si osservi Figura 3.22). I residui del modello finale, inoltre, adesso sono autocorrelati, ma con valori veramente bassi (Figura 3.23).

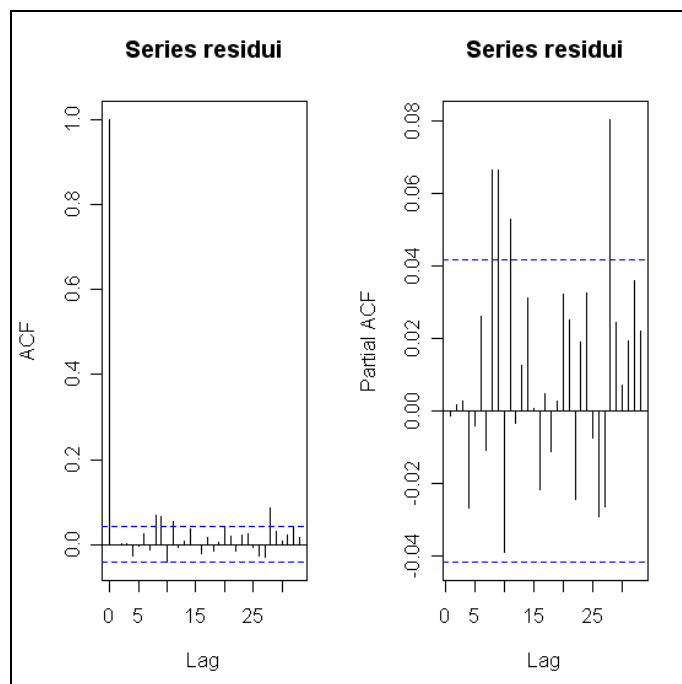


Figura 3.23 - Funzioni di autocorrelazione dei residui del modello Mn

Il modello del Ferro è:

$$\log([\text{Fe}]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2 \cdot \text{spillato2}_t + \beta_3(\text{wd}_t) + \beta_4 \cdot \text{v1p.d}_t + \beta_5 \cdot \text{v1o.d}_t + \beta_6 \cdot \text{v1p.n}_t + \beta_7 \cdot \text{v1o.n}_t + \beta_8 \cdot \text{v1p.n}_{t-1} + \beta_9 \cdot \text{v1o.n}_{t-1} + \beta_{10} \cdot \text{press.n}_{t-1} + \beta_{11} \cdot \text{um.n}_{t-1} + \beta_{12}(\text{fz}_t) \cdot \text{v1p.d}_t + \varepsilon_t$$

I valori dei  $\beta_i$  sono:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-10.442460	2.593144	-4.027	6.13e-05
wdGio	0.514414	0.078679	6.538	1.05e-10
wdLun	0.406974	0.078729	5.169	2.90e-07
wdMar	0.503279	0.077554	6.489	1.43e-10
wdMer	0.498014	0.077825	6.399	2.52e-10
wdSab	0.286260	0.077884	3.675	0.000252
wdVen	0.572553	0.078896	7.257	8.58e-13
spillato1	0.445764	0.098826	4.511	7.33e-06
spillato2	0.046697	0.086968	0.537	0.591440
v1p.d	-0.099278	0.040226	-2.468	0.013774
v1o.d	0.025165	0.021513	1.170	0.242409
v1p.n	0.031302	0.023263	1.346	0.178791
v1o.n	0.104200	0.032722	3.184	0.001501
v1p.n1	0.043606	0.025580	1.705	0.088596
v1o.n1	0.122159	0.032943	3.708	0.000222
press.n1	0.010405	0.002546	4.087	4.76e-05
um.n1	-0.008737	0.001614	-5.412	8.01e-08
fz1:v1p.d	0.194269	0.056823	3.419	0.000657
fz2:v1p.d	0.428349	0.082757	5.176	2.80e-07
fz3:v1p.d	0.315837	0.039849	7.926	6.73e-15

Anche in questo caso si è registrato un miglioramento della varianza spiegata, anche se minore rispetto al Manganese. Con questo modello il rapporto tra varianza spiegata e varianza totale è pari al 0.4365.

I residui di tale modello, poi, sono stati modellati con un AR(2) per la parte non stagionale e con un AR(1) per la parte stagionale (la frequenza della stagionalità è pari a 7 giorni). Le stime finali sono:

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.43267	0.03431	12.6106	0.0000
ar2	0.07850	0.03524	2.2279	0.0261
sar1	0.09642	0.03759	2.5651	0.0105

(Intercept)	-10.90739	3.38043	-3.2266	0.0013
wd="Sab"	-0.20697	0.05385	-3.8433	0.0001
wd="Dom"	-0.45314	0.05519	-8.2098	0.0000
spillato1	0.31427	0.09598	3.2743	0.0011
spillato2	0.05696	0.09250	0.6158	0.5382
vlp.d	-0.05708	0.03576	-1.5960	0.1108
vlo.d	0.03508	0.01849	1.8975	0.0581
vlp.n	0.04493	0.01977	2.2722	0.0233
vlo.n	0.12439	0.02819	4.4129	0.0000
vlp.n1	0.07796	0.02239	3.4829	0.0005
vlo.n1	0.12403	0.02854	4.3455	0.0000
um.n1	-0.00572	0.00180	-3.1723	0.0016
press.n1	0.01113	0.00332	3.3524	0.0008
fz1_vlp.d	0.17929	0.04987	3.5953	0.0003
fz2_vlp.d	0.38139	0.06858	5.5612	0.0000
fz3_vlp.d	0.26292	0.03645	7.2124	0.0000

La distribuzione dei residui del modello finale è asimmetrica con delle code poco pesanti (curtosi minore di 3, Figura 3.24). Le autocorrelazioni sono significativamente diverse da zero, ma tali valori sono così bassi (minori di 0.1) da non essere considerati come correlazioni importanti.

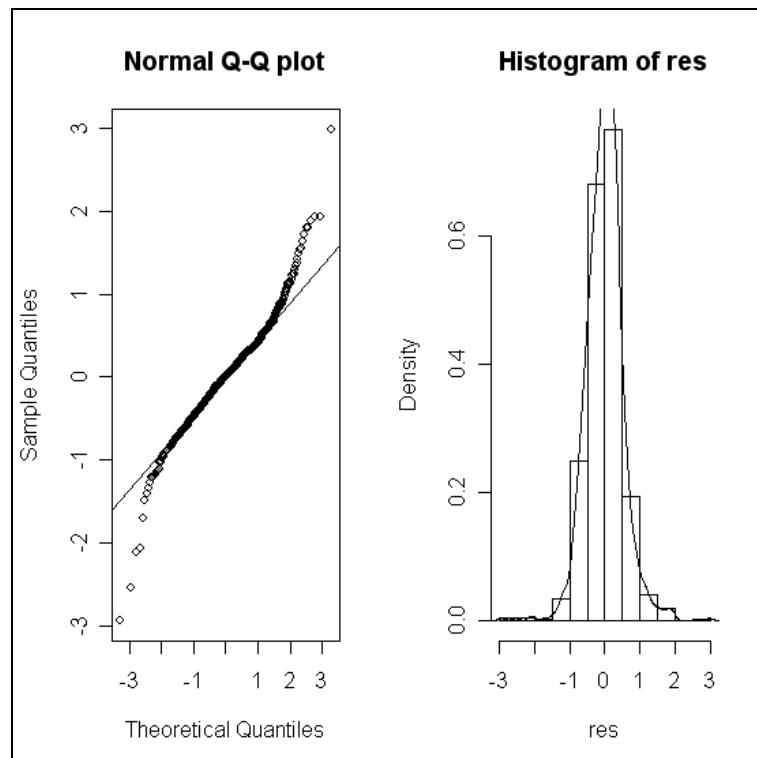


Figura 3.24 - Normal Q-Q plot dei residui del modello Fe

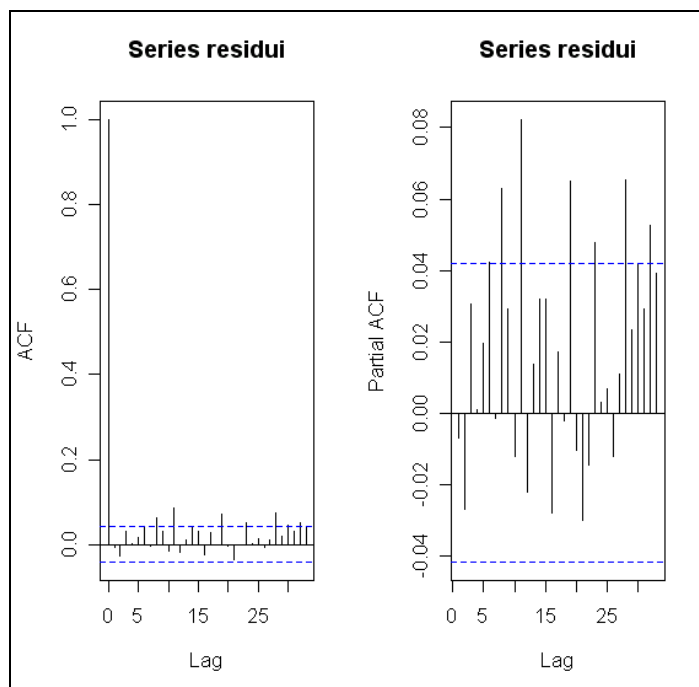


Figura 3.25 - Funzioni di autocorrelazione dei residui del modello Fe

Infine, si è ripreso in considerazione il modello per il Ferro avente il Manganese come variabile esplicativa. Rispetto al modello esposto nel paragrafo precedente, in questo oltre al metallo si è inserito il vento notturno e la pressione notturna (entrambe del giorno precedente):

$$\log([\text{Fe}]_t) = \beta_0 + \beta_1 \cdot \log([\text{Mn}]_t) + \beta_2 \cdot v1p.n_{t-1} + \beta_3 \cdot v1o.n_{t-1} + \beta_4 \cdot \text{press}.n_{t-1} + \varepsilon_t$$

Le stime che si sono ottenute sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-6.523784	1.243675	-5.246	1.81e-07
log(z_Mn)	0.666091	0.010498	63.446	< 2e-16
v1o.n1	0.022218	0.014966	1.485	0.138
v1p.n1	-0.046106	0.010329	-4.464	8.72e-06
press.n1	0.008005	0.001237	6.473	1.35e-10

Anche qui si verifica un aumento del  $R^2$ , arrivando a 0.80 (l'aumento è pari a 0.02).

Dopo aver inserito nel modello le variabili esplicative, si è passati a stimare la parte dei residui. I residui sono stati modellati come un ARMA(1,1) per la parte non stagionale e come un AR(1) per la parte

stagionale (la frequenza è settimanale). In conclusione le stime del modello sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.80051	0.04151	19.2841	0.0000
ma1	-0.51092	0.06147	-8.3110	0.0000
sar1	0.09668	0.03123	3.0962	0.0020
(Intercept)	-6.79473	1.53753	-4.4193	0.0000
log(Mn)	0.64351	0.01016	63.3311	0.0000
vlp.n1	-0.03395	0.00945	-3.5907	0.0003
vlo.n1	0.01947	0.01340	1.4529	0.1465
press.n1	0.00821	0.00153	5.3648	0.0000

Dalla Figura 3.26 si può osservare che i residui non hanno una distribuzione gaussiana. I valori delle autocorrelazioni, anche qui, seppur significativi non sono così alti da essere considerati importanti (Figura 3.27).

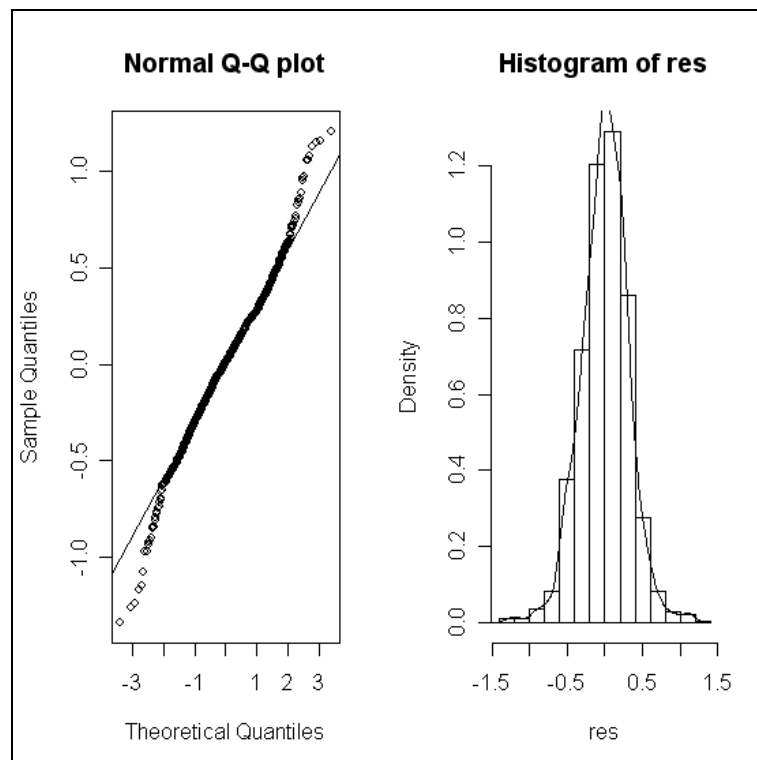


Figura 3.26 - Normal Q-Q plot dei residui del secondo modello Fe



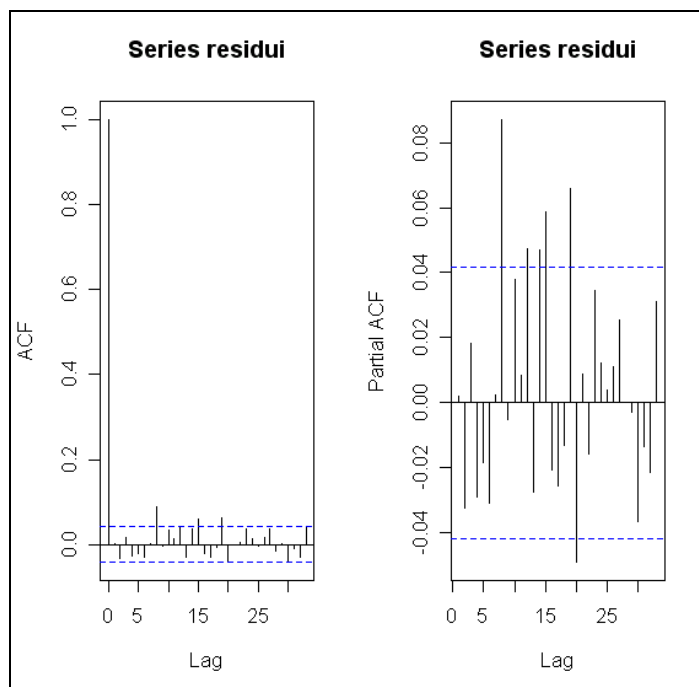


Figura 3.27 - Funzioni di autocorrelazione dei residui del secondo modello Fe

### 3.5.3 - Cadmio, Piombo e Zinco

Come nel paragrafo precedente, anche qui i modelli dei tre metalli differiscono di una variabile che li identifica. Le variabili in comune sono gli spillati, la temperatura notturna, i venti diurni e notturni odierni e il vento notturno del giorno prima. Il Cadmio si differenzia per la pressione diurna del giorno precedente, il Piombo per l'umidità diurna del giorno prima mentre lo Zinco è stata scelta l'umidità notturna del giorno corrente.

Il Cadmio è stato stimato come segue:

$$\log([Cd]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2 \cdot \text{spillato2}_t + \beta_3 \cdot \text{v1p.d}_t + \beta_4 \cdot \text{v1o.d}_t + \beta_5 \cdot \text{v1p.n}_t + \beta_6 \cdot \text{v1o.n}_t + \beta_7 \cdot \text{v1p.n}_{t-1} + \beta_8 \cdot \text{v1o.n}_{t-1} + \beta_{10} \cdot \text{temp.n}_t + \beta_{11} \cdot \text{press.d}_{t-1} + \beta_{12} \cdot (\text{fz}_t) \cdot \text{v1p.d}_t + \varepsilon_t$$

E le stime sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-22.537307	3.770064	-5.978	3.30e-09
spillato1	0.920796	0.134401	6.851	1.39e-11

spillato2	0.283399	0.116874	2.425	0.015519
vlp.d	-0.078153	0.058345	-1.339	0.180762
vlo.d	0.090662	0.030526	2.970	0.003060
vlp.n	0.021472	0.031786	0.676	0.499533
vlo.n	0.155246	0.045451	3.416	0.000666
vlp.nl	0.121845	0.032084	3.798	0.000156
vlo.nl	0.218461	0.045879	4.762	2.25e-06
temp.n	-0.015083	0.004312	-3.498	0.000493
press.dl	0.014628	0.003750	3.901	0.000103
fz1:vlp.d	0.208746	0.079582	2.623	0.008869
fz2:vlp.d	0.524197	0.113658	4.612	4.59e-06
fz3:vlp.d	0.382417	0.057839	6.612	6.64e-11

Con l'inserimento di queste variabili, l'indice  $R^2$  passa da 0.45 a 0.47.

I residui dei modelli del Cadmio sono stati modellati, poi, come un SARIMA (1,0,0)x(1,0,1) con frequenza stagionale pari a 3.

Come risultati finali, si è ottenuto:

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.33788	0.03375	10.0113	0.0000
sar1	0.89146	0.07864	11.3363	0.0000
sma1	-0.82779	0.09936	-8.3316	0.0000
(Intercept)	-22.63751	4.57295	-4.9503	0.0000
spillato1	0.67788	0.13729	4.9375	0.0000
spillato2	0.20473	0.12563	1.6297	0.1035
vlp.d	-0.06145	0.05370	-1.1445	0.2527
vlo.d	0.09829	0.02777	3.5399	0.0004
vlp.n	0.02999	0.02883	1.0402	0.2985
vlo.n	0.16937	0.04190	4.0421	0.0001
vlp.nl	0.13893	0.02953	4.7055	0.0000
vlo.nl	0.20920	0.04256	4.9153	0.0000
temp.n	-0.01651	0.00680	-2.4266	0.0154
press.dl	0.01488	0.00456	3.2667	0.0011
fz1_vlp.d	0.17482	0.07298	2.3954	0.0168
fz2_vlp.d	0.49912	0.10194	4.8961	0.0000
fz3_vlp.d	0.35191	0.05441	6.4676	0.0000

La distribuzione dei residui, rappresentata in Figura 3.28, ha delle code poco pesanti (curtosi minore di 3) ed è asimmetrica. Dalle funzioni di autocorrelazione si vede che queste hanno dei valori veramente bassi per essere considerate (Figura 3.29).

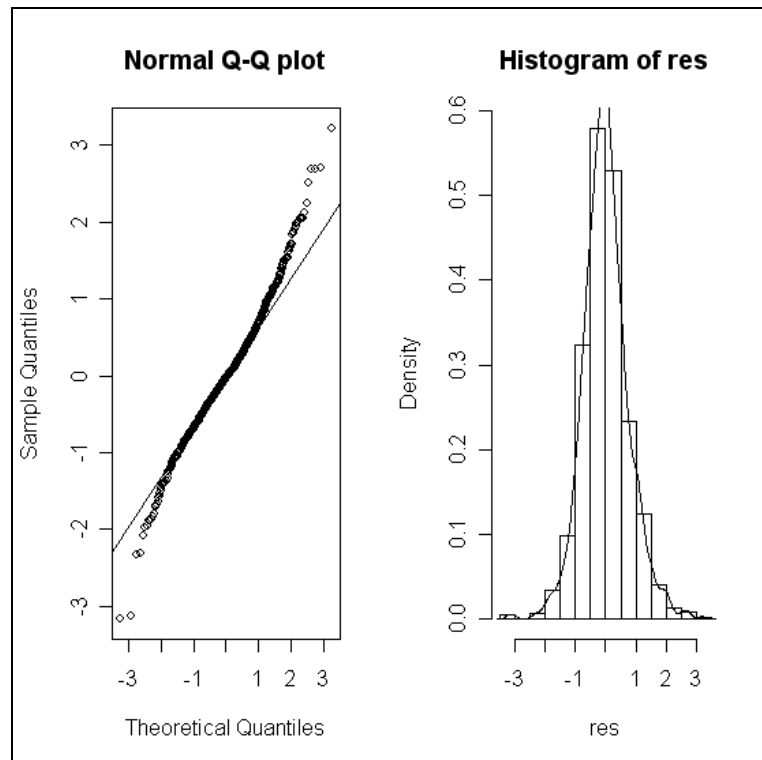


Figura 3.28 - Normal Q-Q plot e istogramma dei residui del modello Cd

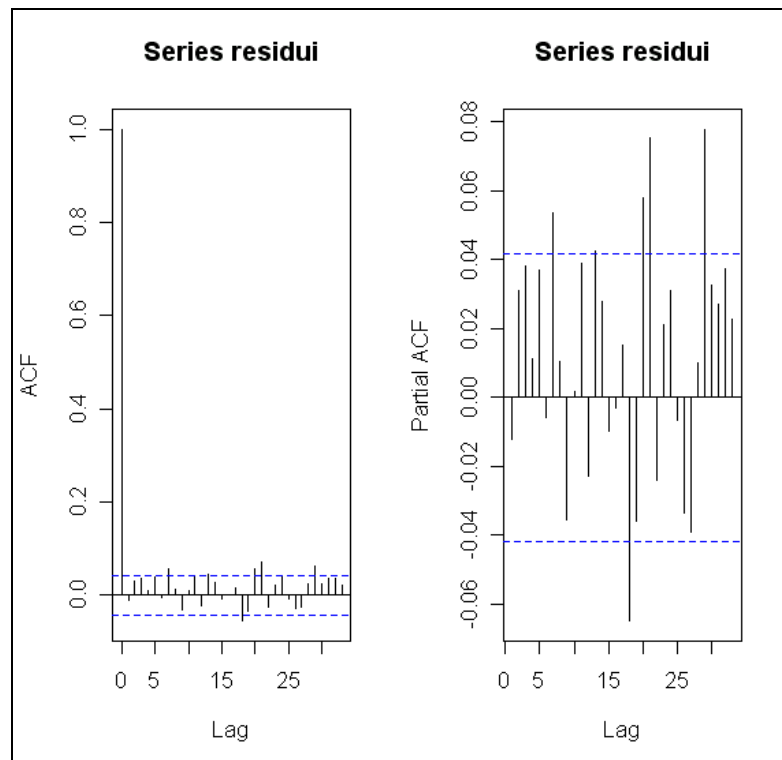


Figura 3.29 - Funzioni di autocorrelazione dei residui del modello Cd

Il modello del Piombo, invece, ha la seguente formula:

$$\log([Pb]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2 \cdot \text{spillato2}_t + \beta_3 \cdot \text{v1p.d}_t + \beta_4 \cdot \text{v1o.d}_t + \beta_5 \cdot \text{v1p.n}_t + \beta_6 \cdot \text{v1o.n}_t + \beta_7 \cdot \text{v1p.n}_{t-1} + \beta_8 \cdot \text{v1o.n}_{t-1} + \beta_9 \cdot \text{temp.n}_t + \beta_{10} \cdot \text{um.d}_{t-1} + \beta_{12}(\text{fz}_t) \cdot \text{v1p.d}_t + \varepsilon_t$$

Si sono ottenute le seguenti stime:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-3.341196	0.158755	-21.046	< 2e-16
spillato1	0.713379	0.142891	4.992	7.18e-07
spillato2	0.283339	0.124577	2.274	0.023181
v1p.d	-0.102248	0.060528	-1.689	0.091521
v1o.d	0.055701	0.032092	1.736	0.082975
v1p.n	0.067106	0.033759	1.988	0.047139
v1o.n	0.150085	0.047770	3.142	0.001735
v1p.n1	0.193404	0.034538	5.600	2.87e-08
v1o.n1	0.179029	0.048123	3.720	0.000212
temp.n	-0.011043	0.004215	-2.620	0.008946
um.d1	-0.005007	0.001711	-2.927	0.003511
fz1:v1p.d	0.279364	0.082816	3.373	0.000775
fz2:v1p.d	0.594369	0.119734	4.964	8.29e-07
fz3:v1p.d	0.449756	0.059241	7.592	8.02e-14

in questo caso il miglioramento dell'adattamento del modello ai dati è superiore rispetto al modello del Cadmio: in questo caso l'R<sup>2</sup> da 0.45 è passato al 0.4871.

Da questo punto si è cercato, poi, di modellare i residui. La loro forma finale è come quella scelta per il Cadmio: un SARIMA (1,0,0)x(1,0,1) con frequenza stagionale a 3 periodi. Le stime dei  $\beta_i$  nel modello finale risultano:

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.36093	0.03315	10.8891	0.0000
sar1	0.96270	0.03168	30.3899	0.0000
sma1	-0.93260	0.04213	-22.1366	0.0000
intercept	-3.27173	0.18576	-17.6123	0.0000
spillato1	0.56458	0.14378	3.9268	0.0001
spillato2	0.16133	0.13226	1.2197	0.2229
v1p.d	-0.08038	0.05620	-1.4301	0.1530

vlo.d	0.06608	0.02881	2.2936	0.0221
vlp.n	0.07709	0.03047	2.5302	0.0116
vlo.n	0.17596	0.04377	4.0202	0.0001
vlp.n1	0.19737	0.03122	6.3228	0.0000
vlo.n1	0.19227	0.04435	4.3356	0.0000
temp.n	-0.01889	0.00681	-2.7749	0.0056
um.d1	-0.00267	0.00191	-1.3995	0.1620
fz1_vlp.d	0.26283	0.07651	3.4353	0.0006
fz2_vlp.d	0.54248	0.10701	5.0695	0.0000
fz3_vlp.d	0.41077	0.05661	7.2556	0.0000

Anche qui i residui non hanno una distribuzione gaussiana a causa di code troppo poco pesanti (per cui l'indice di curtosi è pari a 1.9) ed è leggermente asimmetrica (indice di asimmetria pari a 0.16).

Anche in questo caso le autocorrelazioni sono molto basse: quelle significativamente diverse da zero non superano la soglia del 0.05 e per tanto le si può ignorare.

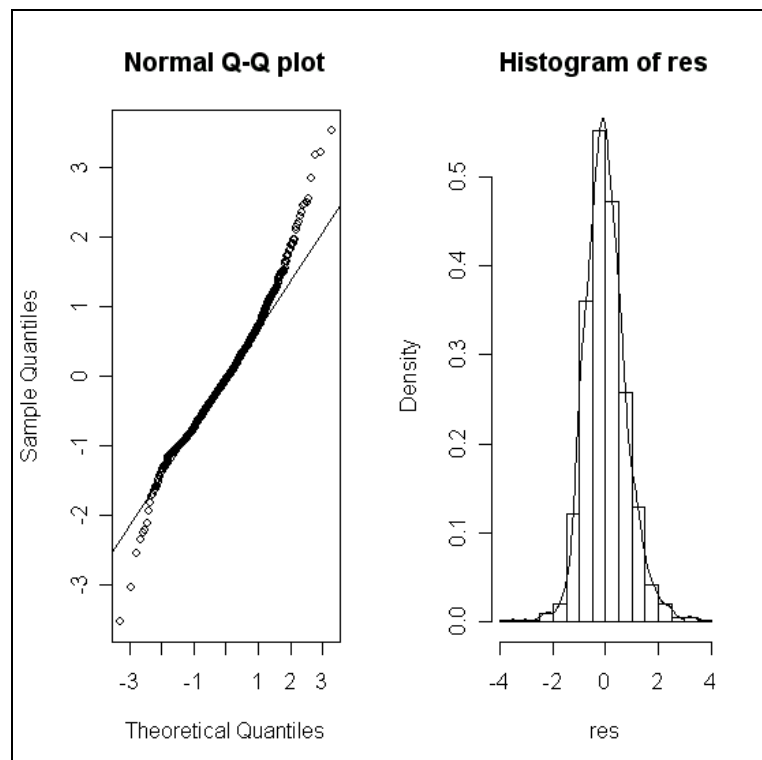


Figura 3.30 – Normal Q-Q plot e istogramma dei residui del modello Pb

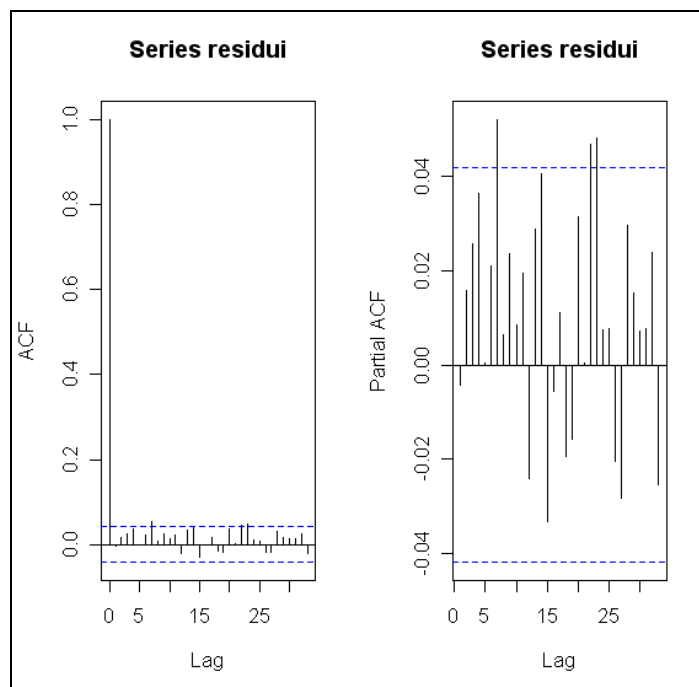


Figura 3.31 – Funzioni di autocorrelazione dei residui del modello Pb

Infine è stato stimato il modello dello Zinco:

$$\log([Zn]_t) = \beta_0 + \beta_1 \cdot \text{spillato1}_t + \beta_2 \cdot \text{spillato2}_t + \beta_3 \cdot \text{v1p.d}_t + \beta_4 \cdot \text{v1o.d}_t + \beta_5 \cdot \text{v1p.n}_t + \beta_6 \cdot \text{v1o.n}_t + \beta_7 \cdot \text{v1p.n}_{t-1} + \beta_8 \cdot \text{v1o.n}_{t-1} + \beta_9 \cdot \text{um.n}_t + \beta_{10}(\text{fz}_t) \cdot \text{v1p.d}_t + \varepsilon_t$$

cioè il modello stimato differisce dal modello spiegato nel paragrafo precedente 3.4.3. solo per le variabili del vento della notte precedente. Le stime sono le seguenti:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.895837	0.210518	-4.255	2.32e-05
spillato1	0.575556	0.136720	4.210	2.83e-05
spillato2	0.137123	0.118149	1.161	0.246136
v1p.d	-0.120516	0.058507	-2.060	0.039716
v1o.d	0.114950	0.030052	3.825	0.000140
v1p.n	0.020138	0.033569	0.600	0.548740
v1o.n	0.112370	0.044608	2.519	0.011950
v1p.n1	0.184657	0.031686	5.828	7.98e-09
v1o.n1	0.146184	0.045243	3.231	0.001281
temp.n	-0.019858	0.004071	-4.878	1.28e-06
um.n	-0.010308	0.002215	-4.655	3.77e-06

fz1:v1p.d	0.253469	0.077704	3.262	0.001151
fz2:v1p.d	0.622815	0.111596	5.581	3.22e-08
fz3:v1p.d	0.452714	0.056929	7.952	5.84e-15

Con l'inserimento di queste due ultime variabili, l'indice  $R^2$  arriva a 0.52. Da quel modello si è poi stimata la struttura dei residui che è un AR(1) non stagionale. I  $\beta_i$ -stimati nel modello complessivo sono di seguito riportati:

	Estimate	Std. Error	t value	Pr(> t )
ar1	0.34268	0.03437	9.9703	0.0000
(Intercept)	-1.29925	0.23766	-5.4669	0.0000
spillato1	0.52952	0.13723	3.8586	0.0001
spillato2	0.15690	0.12331	1.2724	0.2036
v1p.d	-0.09122	0.05636	-1.6186	0.1059
v1o.d	0.10397	0.02783	3.7358	0.0002
v1p.n	0.05537	0.03187	1.7372	0.0827
v1o.n	0.12724	0.04168	3.0529	0.0023
v1p.n1	0.18804	0.02970	6.3321	0.0000
v1o.n1	0.15127	0.04256	3.5544	0.0004
temp.n	-0.02431	0.00526	-4.6257	0.0000
um.n	-0.00496	0.00253	-1.9633	0.0499
fz1_v1p.d	0.22934	0.07415	3.0931	0.0020
fz2_v1p.d	0.52541	0.10339	5.0816	0.0000
fz3_v1p.d	0.41414	0.05627	7.3599	0.0000

L'indice di asimmetria del modello è pari a 0.13 mentre quello di curtosi è circa 1.82. In Figura 3.33 sono rappresentate le funzioni di autocorrelazione.

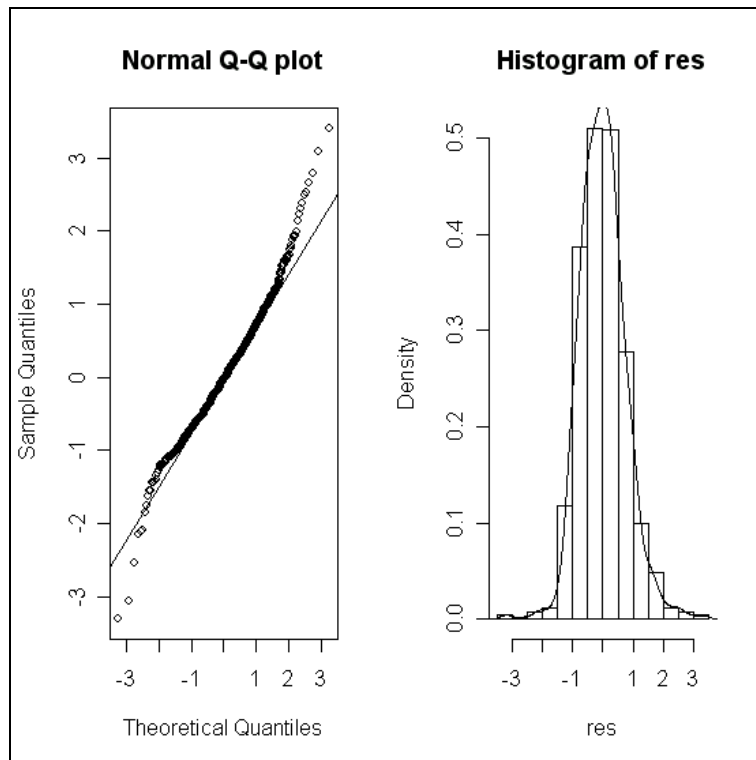


Figura 3.32 – Normal Q-Q plot e istogramma dei residui del modello Zn

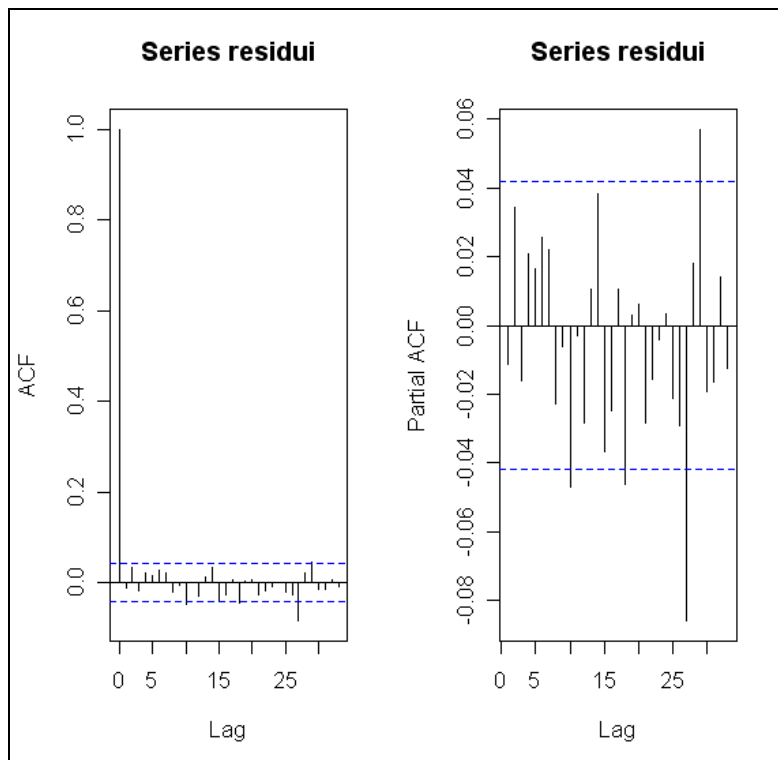


Figura 3.33 – Funzioni di autocorrelazione dei residui del modello Zn



Infine, si è stimato i modelli del Piombo e dello Zinco con il Cadmio come variabile esplicativa.

Per il primo sono stati scelti il vento diurno e l'umidità notturna (entrambe odierni), il vento notturno e la pressione diurna come variabili ritardate:

$$\log([\text{Pb}]_t) = \beta_0 + \beta_1 \cdot \log([\text{Cd}]_t) + \beta_2 \cdot v1p.d_t + \beta_3 \cdot v1o.d_t + \beta_4 \cdot v1p.n_{t-1} + \beta_5 \cdot v1o.n_{t-1} + \beta_{10} \cdot \text{press}.d_{t-1} + \beta_{11} \cdot \text{um}.d_t + \varepsilon_t$$

Le stime sono qui riportate:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	11.8095481	2.0704249	5.704	1.46e-08
log(Cd)	0.8185889	0.0180908	45.249	< 2e-16
v1p.d	0.1036862	0.0139652	7.425	2.09e-13
v1o.d	0.0145337	0.0161893	0.898	0.370
v1p.n1	0.0833622	0.0190541	4.375	1.32e-05
v1o.n1	-0.0119928	0.0266214	-0.450	0.652
press.d1	-0.0086317	0.0020293	-4.254	2.26e-05
um.d	-0.0037050	0.0009156	-4.046	5.52e-05

La differenza dell'indice  $R^2$  tra questo modello e quello precedente è quasi nulla. Da questo modello, i residui sono stati modellati come un ARMA(2,1). Il risultato è di seguito esposto.

	Estimate	Std. Error	t value	Pr(> t )
ar1	1.20333	0.08856	13.5878	0.0000
ar2	-0.23299	0.06403	-3.6389	0.0003
ma1	-0.88026	0.07679	-11.4639	0.0000
intercept	9.24726	2.65628	3.4813	0.0005
log(Cd)	0.81557	0.01796	45.4142	0.0000
v1p.d	0.09857	0.01238	7.9633	0.0000
v1o.d	0.01053	0.01452	0.7251	0.4686
v1p.n1	0.07938	0.01668	4.7588	0.0000
v1o.n1	0.01055	0.02372	0.4449	0.6565
um.d	-0.00219	0.00107	-2.0506	0.0406
press.d1	-0.00617	0.00262	-2.3513	0.0189

Con un'indice di asimmetria pari a 0.15 e di curtosi uguale a 2.88, la distribuzione dei residui del modello finale non è gaussiana (anche se gli indici si discostano di poco).

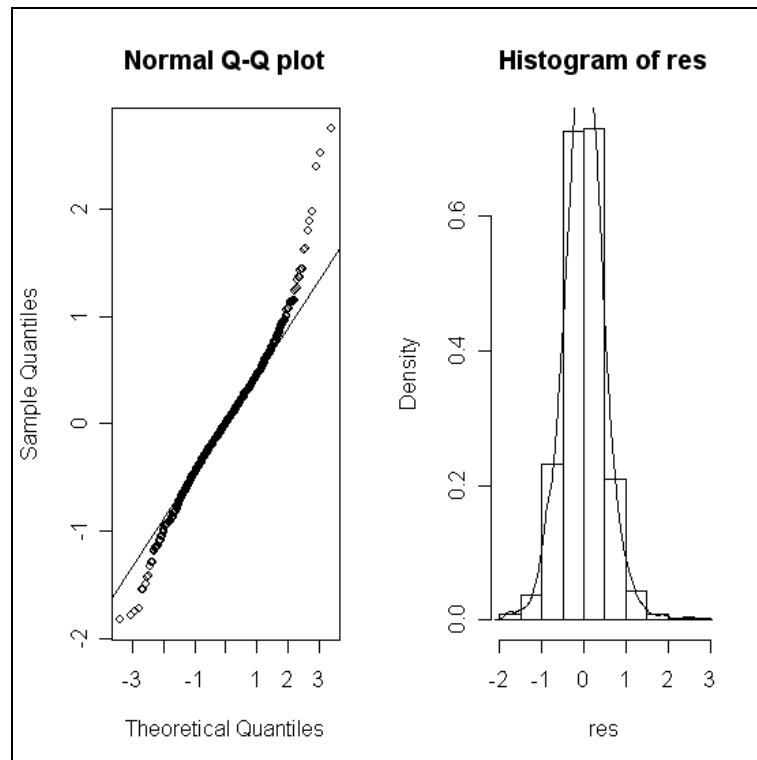


Figura 3.34 – Normal Q-Q plot e istogramma dei residui del secondo modello Pb

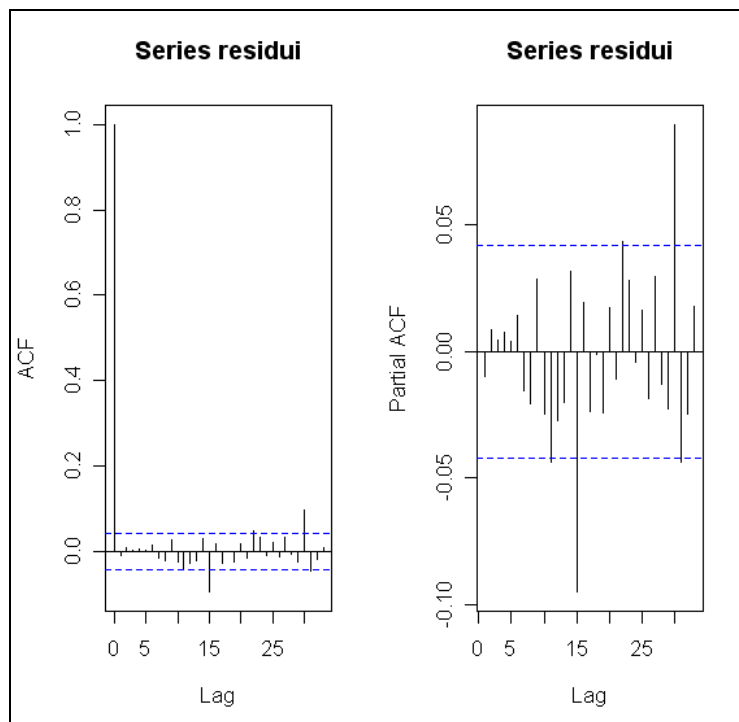


Figura 3.35 – Funzioni di autocorrelazione dei residui del secondo modello Pb

Per lo Zinco, invece, il numero di variabili è molto minore: si è inserito solo il vento diurno di oggi e l'umidità notturna del giorno precedente:

$$\log([Zn]_t) = \beta_0 + \beta_1 \cdot \log([Cd]_t) + \beta_2 \cdot v1p.d_t + \beta_3 \cdot v1o.d_t + \beta_4 \cdot um.n_{t-1} + \varepsilon_t$$

Si sono ottenuti i seguenti risultati di stima:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	4.389957	0.174994	25.086	< 2e-16
log(Cd)	0.701413	0.019971	35.122	< 2e-16
v1p.d	0.174450	0.013928	12.525	< 2e-16
v1o.d	0.050924	0.018072	2.818	0.00491
um.n1	-0.006522	0.001318	-4.948	8.52e-07

Anche in questo caso il modello non ha registrato dei miglioramenti, a livello di  $R^2$ .

I residui sono modellati, poi, come un ARMA(2,2) per lo Zinco.

	Estimate	Std. Error	t value	Pr(> t )
ar1	1.52667	0.09411	16.2221	0.0000
ar2	-0.52876	0.09328	-5.6683	0.0000
ma1	-1.19526	0.10601	-11.2749	0.0000
ma2	0.22266	0.09800	2.2719	0.0233
intercept	4.48603	0.23905	18.7661	0.0000
log(Cd)	0.71114	0.01922	36.9992	0.0000
v1p.d	0.14528	0.01245	11.6739	0.0000
v1o.d	0.03695	0.01550	2.3838	0.0173
um.n1	-0.00598	0.00144	-4.1704	0.0000

L'indice di asimmetria, in questo caso, è prossimo allo zero (unico caso in cui l'indice sia negativo) e la curtosi è calcolata a 2.29. Anche qui le funzioni di autocorrelazione, benchè ci sia qualche valore significativamente diverso da zero, in valore assoluto non superano la soglia del 0.1, per cui si può ipotizzarle non influenti nel modello.

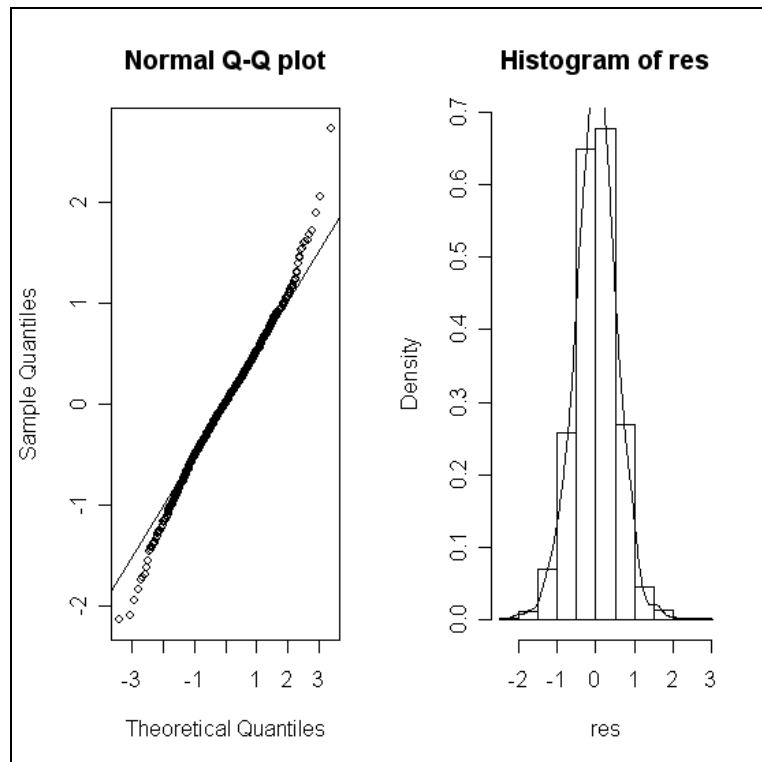


Figura 3.36 – Normal Q-Q plot e istogramma dei residui del secondo modello Zn

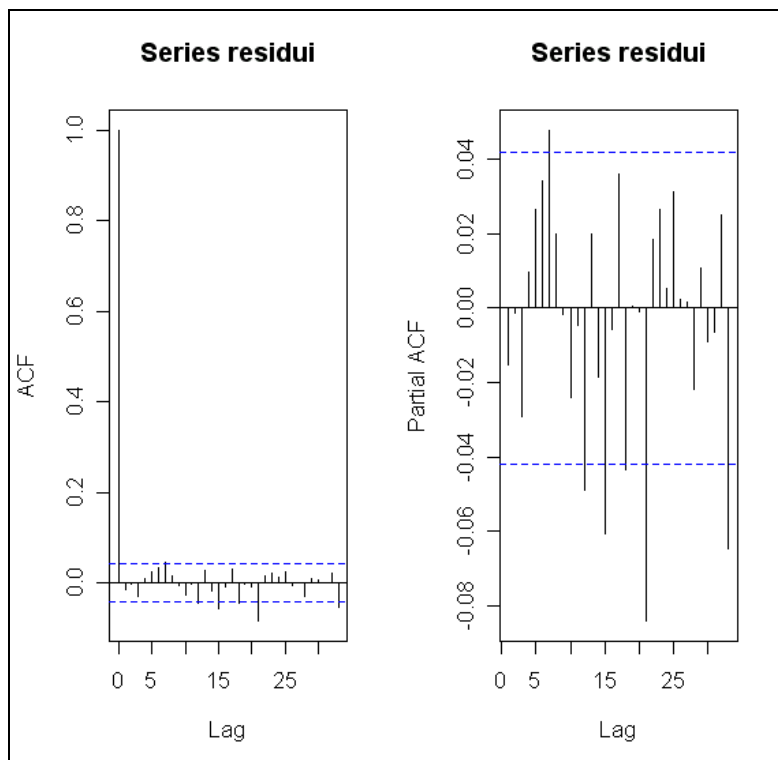


Figura 3.37 – Funzioni di autocorrelazione dei residui del secondo modello Zn

### 3.6 - La validazione incrociata

Per valutare la bontà dei modelli ottenuti, si è pensato di effettuare un controllo attraverso una validazione incrociata.

La validazione incrociata è stata effettuata su un gruppo di dati in cui tutti i modelli potessero venir stimati: dal 1 luglio 2003 fino al 16 aprile 2006, che corrisponde alla lunghezza del dataset dei dati di produzione.

Vengono inoltre confrontati i modelli finali (basati sulle variabili meteorologiche ed in cui viene modellato anche l'errore) e la camminata casuale (*random walk*). La formula di quest'ultima è:

$$Y_t = Y_{t-1} + \varepsilon_t$$

Nel *random walk*, se le osservazioni arrivano fino al tempo  $t$ , la miglior previsione al tempo  $t+1$  è l'osservazione fatta al tempo  $t$ .

Per decidere quale indice utilizzare al fine di valutare gli scarti tra le previsioni e le osservazioni (errori di previsione) ci si è basati sulle considerazioni seguenti.

Ci si è posti il seguente problema: all'interno del dataset ci sono delle osservazioni anomale? La risposta, purtroppo, è affermativa.

Dagli esperti dell'ARPA FVG che seguono il caso dell'acciaieria ci è stato riferito che nel periodo (all'incirca) tra il settembre 2004 fino a marzo 2005 si sono osservate delle concentrazioni di metalli molto più alte del normale. L'azienda in questo caso, avvertita del fenomeno, non è intervenuta effettuando dei cambiamenti di tipo strutturale all'impiantistica, bensì ha effettuato un maggiore e più stretto controllo sulle emissioni e sulle strutture; dall'aprile 2005 i valori delle emissioni sono ritornate sui valori più bassi e maggiormente conformi. Quanto è stato appena detto lo si può osservare nella Figura 4.1, in cui è rappresentato il Manganese, che è uno dei migliori traccianti di questa attività industriale.

La lunghezza temporale di queste fuoriuscite è decisamente significativa, se si pensa che il dataset completo ha una lunghezza totale pari a quasi 34 mesi.

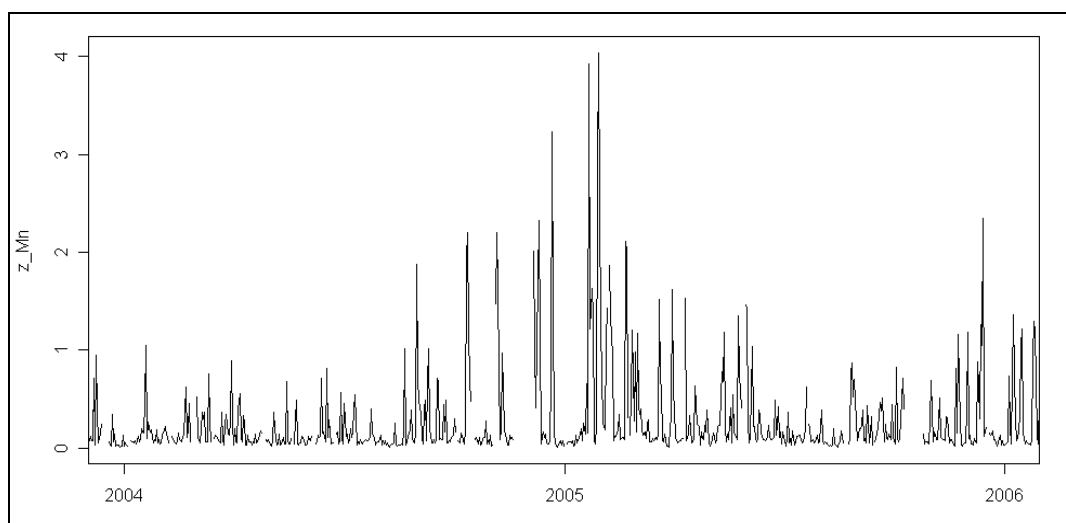


Figura 4.1 – Andamento nel Manganese negli anni 2004 e 2005

Per questo motivo, l'utilizzo di una media degli errori di previsione sarebbe distorta da tali anomalie. Serve un indice più robusto per risolvere tali problematiche.

Un indice che è meno influenzato, e quindi più robusto, è la mediana: in questo caso si è utilizzata la mediana del valore assoluto degli errori.

Per effettuare la validazione incrociata, il set dei dati è stato suddiviso in 10 gruppi di uguale numerosità (102 dati per gruppo): su nove di questi viene stimato il modello e nel decimo rimanente si effettua la previsione. Con una rotazione di tutti i gruppi si è potuta esaminare la capacità previsiva in tutto l'arco temporale.

In Tabella 4.1 riportati i valori degli indici risultanti della validazione incrociatae quelli della camminata casuale.

Nel modello finale, la previsione del modello sono calcolate come

$$\hat{Y} = \exp(\beta_0 + \beta_1 \cdot x_1 + \dots + \beta_p \cdot x_p)$$

e sono espresse in  $\mu\text{g}/\text{m}^3$ .

Sia per il random walk che per i modelli finali, si è stimata la mediana dei valori assoluti degli errori (espressi in  $\mu\text{g}/\text{m}^3$ ):

$$\text{mediana errore assoluto} = \text{median}(|Y_t - \hat{Y}_t|)$$

In Tabella 4.2 sono espresse le mediane degli errori percentuale, calcolate come segue:

$$\text{mediana errore percentuale} = \text{median} \left( \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \right)$$

Come si può vedere dalle tabelle, l'errore dei modelli finali è sempre minore di quello della passeggiata casuale.

L'unico metallo che ha errori di previsione quasi simili al random walk è il Nichel, il quale è l'unico a rigettare le variabili inerenti alla produzione.

Tabella 4.1 - Tabella mediane degli errori di previsione (valori assoluti) calcolati per la validazione incrociata

<b>Metallo da spiegare</b>	<b>Modello finale (in <math>\mu\text{g}/\text{m}^3</math>)</b>	<b>Random Walk (in <math>\mu\text{g}/\text{m}^3</math>)</b>
<b>Manganese</b>	0.04698	0.06672
<b>Ferro</b>	0.34630	0.41647
<b>Nichel</b>	0.00289	0.00428
<b>Cadmio</b>	0.00032	0.00039
<b>Piombo</b>	0.02557	0.03274
<b>Zinco</b>	0.12457	0.18685

Tabella 4.2 – Tabella mediane delle percentuali d'errore di previsione calcolati per la validazione incrociata

<b>Metallo da spiegare</b>	<b>Modello finale</b>	<b>Random Walk</b>
<b>Manganese</b>	44.978%	57.899%
<b>Ferro</b>	34.045%	41.853%
<b>Nichel</b>	38.169%	39.917%
<b>Cadmio</b>	49.389%	58.509%
<b>Piombo</b>	53.252%	56.805%
<b>Zinco</b>	50.223%	56.477%





## **CAPITOLO 4**

### **Considerazioni conclusive**

Dall'analisi descrittiva dei dati sono emerse numerose informazioni utili.

In primo luogo è importante la trasformazione logaritmica che è stata applicata alle variabili risposta dei vari modelli.

Le variabili trasformate si sono rilevate simmetriche e con una varianza stabile (si veda il grafico di dispersione di Figura 4.1: nella scala originale, al crescere dei valori misurati aumenta la loro dispersione).

Analizzando le matrici di correlazione dei campionatori A e C, poi, sono emersi vari punti.

1. In primo luogo si sono distinti dei gruppi tra i vari metalli: il primo composto da Ferro e Manganese, il secondo che include Cadmio, Zinco e Piombo e il terzo dove si possono includere i rimanenti elementi.

Questa distinzione è importante, in quanto concorda con le conoscenze di cui si disponeva a priori, relativamente alle associazioni fra i vari metalli nelle emissioni dovute alle diverse fasi di produzione.

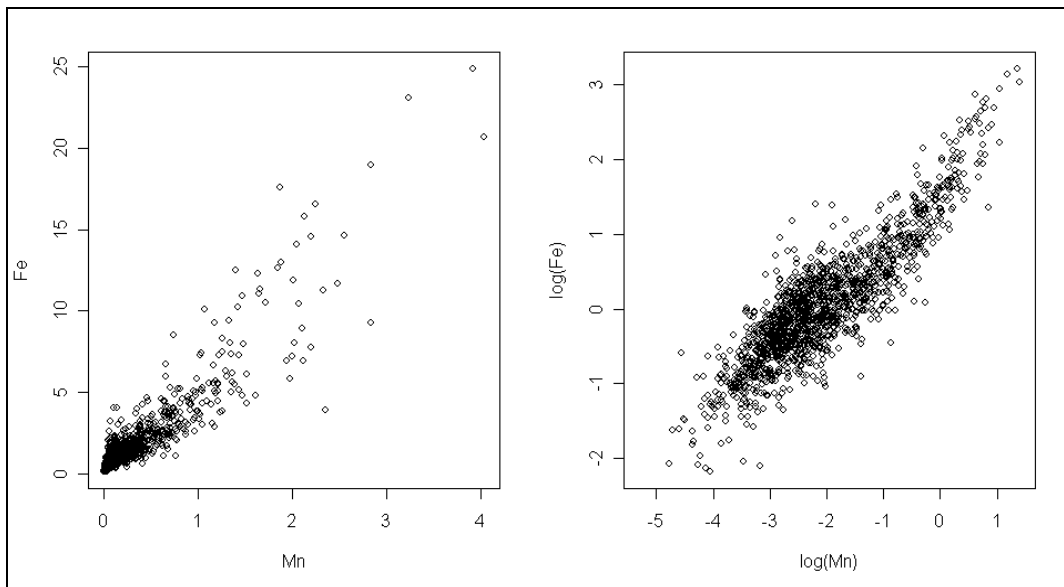


Figura 4.1 – Grafico tra Manganese e Ferro, senza e con la trasformazione logaritmica

Le informazioni a disposizione riguardano:

- Ferro e Manganese, tipici della produzione siderurgica; il primo è presente soprattutto nelle emissioni primarie (fase fusoria), il secondo è riscontrato prevalentemente nelle emissioni secondarie;
  - Nichel, Cromo e Vanadio sono metalli utilizzati per la produzione di acciai speciali (fase di affinazione);
  - Cadmio, Piombo e Zinco: il minerale di Cadmio è associato a quello di Zinco e il Piombo è un'impurezza del minerale di Zinco; inoltre lo Zinco e il Piombo sono basso fondenti e la loro presenza è principalmente nelle emissioni secondarie (fase di carica dei forni).
2. Si è osservato che le matrici di correlazione dei dati di concentrazione, condizionati alla tipologia di funzionamento dell'impianto, cambiano. Nel caso in cui sia attivo il solo Forno1, il Forno2 o entrambi simultaneamente, queste matrici sembrano avere una struttura simile. Invece nel caso in cui l'azienda abbia entrambe i forni spenti si può evidenziare che solamente la correlazione tra Ferro e Manganese rimane significativa.

E' interessante osservare che quest'ultima matrice corrisponde alla matrice di correlazione (generale) del campionatore B, il quale è situato in posizione ortogonale rispetto alla direzione dei venti dominanti.

Da questo risultato è emersa l'importanza di includere nei modelli le variabili che descrivono il vento, prevedendo anche termini di interazione con gli indicatori di attività dello stabilimento.

Date queste osservazioni, ci si aspettava che l'analisi delle componenti principali desse dei risultati analoghi, essendo basata, a sua volta, sull'utilizzo della matrice di correlazione.

La PCA non ha dato i risultati sperati, che si sono invece ottenuti nelle successive modellazioni.

A causa della censura effettuata sui dati, il modello generale è stato scomposto in due. Il primo è un modello binomiale, con il quale si intende prevedere il superamento o meno della soglia di rilevabilità; il secondo un modello lineare per i dati il cui valore supera la soglia.

A causa di un forte sbilanciamento delle variabili binomiali verso il valore 1, la stima effettuata tramite modelli binomiali è risultata difficoltosa. L'unico a non dare alcun problema è risultato il Nichel: si è evidenziato che le variabili esplicative scelte corrispondono sostanzialmente alle stesse scelte nel modello lineare, sviluppato per i dati sopra soglia. Si può ipotizzare che, coerentemente, le variabili che spiegano l'andamento delle concentrazioni di un determinato metallo siano le stesse che concorrono a determinare il superamento o meno della soglia di rilevabilità.

Per la stima dei modelli si è supposta una distribuzione gaussiana dell'errore. In questi modelli l'andamento della variabile risposta (la concentrazione del metallo in esame) è spiegato dalle variabili di produzione e dalle variabili meteorologiche. A seconda delle variabili esplicative scelte si possono distinguere gli stessi gruppi evidenziati dalle matrici di correlazione. In Tabella 4.1 sono rappresentate tutte le variabili scelte nei vari modelli (con  $fz*v1p.d$  e  $fz*v1o.d$  si intende l'interazione tra le due variabili).

Tabella 4.1 - Variabili scelte dai modelli finali

	log(Mn)	log(Fe)	log(Ni)	log(Cd)	log(Pb)	log(Zn)
<b>spillato1</b>	X	X		X	X	X
<b>spillato2</b>	X	X		X	X	X
<b>wd</b>	X	X				
<b>v1p.d</b>	X	X	X	X	X	X
<b>v1o.d</b>	X	X	X	X	X	X
<b>v1p.n</b>	X	X		X	X	X
<b>v1o.n</b>	X	X		X	X	X
<b>v1p.n1</b>	X	X		X	X	X
<b>v1o.n1</b>	X	X		X	X	X
<b>temp.n</b>				X	X	X
<b>um.n</b>						X
<b>um.d1</b>					X	
<b>um.n1</b>	X	X				
<b>press.d1</b>				X		
<b>press.n1</b>		X	X			
<b>fz*v1p.d</b>	X	X		X	X	X
<b>fz*v1p.n</b>	X					

Va notato che solamente il Nichel rigetta le variabili inerenti alla produzione.

Il gruppo del Manganese-Ferro, invece, è l'unico a richiedere la variabile *giorno della settimana* tra le esplicative del modello. Questi due metalli, infatti, registrano mediamente una concentrazione minore durante il sabato e la domenica, comportamento del tutto contrastante con il fatto che l'acciaiera in tali giorni lavora normalmente a pieno regime.

Il motivo di tale andamento sarà uno degli obiettivi delle analisi che verranno effettuate in futuro.

Tra le variabili di produzione, si vede che il coefficiente dello *spillato1* ha sempre un valore di stima maggiore di quello dello *spillato2*. Una spiegazione si può cercare nel fatto che il Forno1 è più grande e quindi la sua produzione giornaliera, a pieno regime, è nettamente superiore a quella del Forno2: cosa che non traspare dalle variabili normalizzate.

Tabella 4.2 - Variabili scelte dai modelli all'interno dei gruppi

	<b>log(Fe)</b>	<b>log(Pb)</b>	<b>log(Zn)</b>
<b>log(Mn)</b>	<b>X</b>		
<b>log(Cd)</b>		<b>X</b>	<b>X</b>
<b>v1p.d</b>		<b>X</b>	<b>X</b>
<b>v1o.d</b>		<b>X</b>	<b>X</b>
<b>v1p.n1</b>	<b>X</b>	<b>X</b>	
<b>v1o.n1</b>	<b>X</b>	<b>X</b>	
<b>um.n</b>		<b>X</b>	
<b>um.n1</b>			<b>X</b>
<b>press.d1</b>		<b>X</b>	
<b>press.n1</b>	<b>X</b>		

Tuttavia, in contrasto con quanto detto, questo forno dispone di una segregazione più completa (*elephant house*) rispetto all'altro e ciò induce a pensare che produca una quantità minore di emissioni diffuse rispetto al Forno2, segregato con una *dog-house*.

Per ciascun gruppo di metalli, fra quelli evidenziatisi, sono stati creati dei modelli in cui la concentrazione di uno dei metalli è stata inserita come variabile esplicativa. La creazione di tali modelli può essere utile per valutare l'opportunità di una riduzione del numero di elementi chimici da analizzare. In Tabella 4.2 sono evidenziate le variabili scelte.

Va sottolineato poi che per lo Zinco e per il Piombo, pur essendo all'interno dello stesso gruppo, in questi modelli vengono inserite un numero e tipi di variabili esplicative diverse tra loro.

Nei modelli finali non è mai stata inserita la variabile *pioggia*. Questo è un fatto molto strano, in quanto la pioggia gioca un ruolo importante nella rimozione particolato atmosferico. Il motivo è da ricercare nella relazione che intercorre tra questa variabile e le variabili *umidità* e *pressione*, le quali sono, invece, inserite come variabili esplicative in tali modelli.

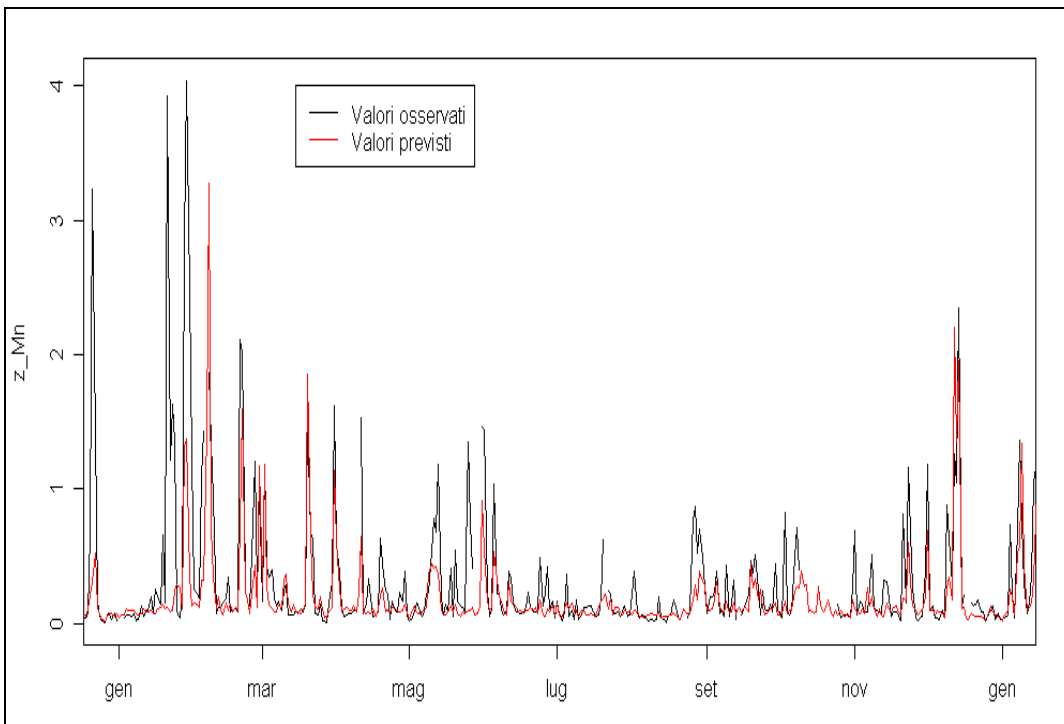


Figura 4.2– Andamento del Manganese nel 2005 raffrontato con il suo modello finale

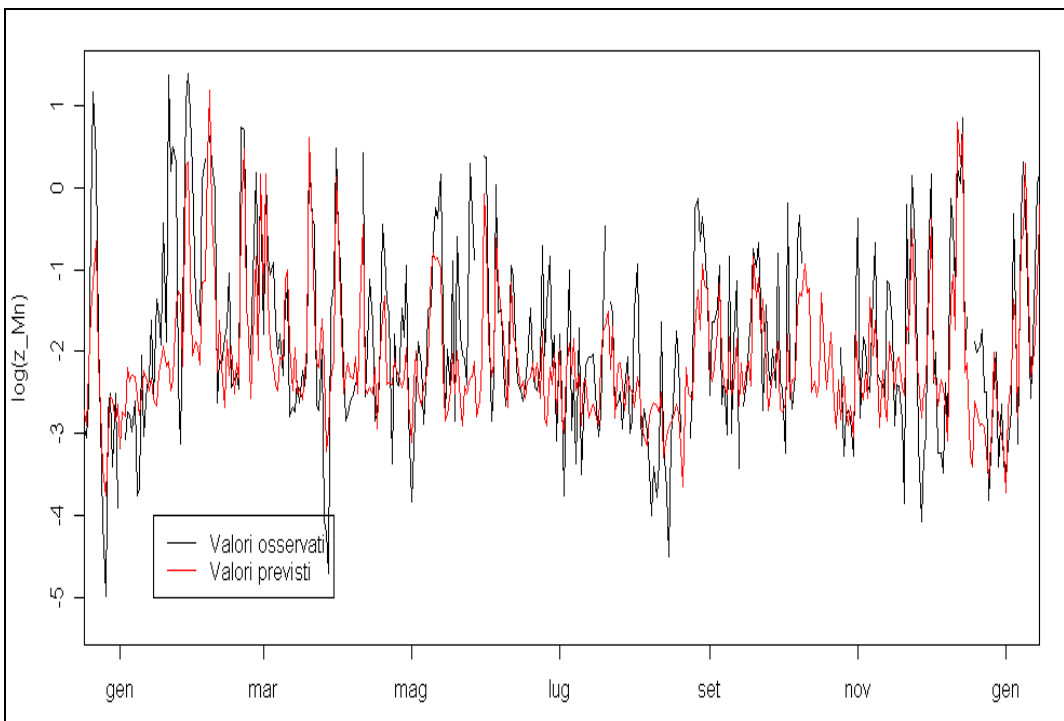


Figura 4.3 - Andamento del log(Manganese) nel 2005 raffrontato con il modello finale

Infine è importante notare dai grafici di Figura 4.2 e 4.3 (relativi alle concentrazioni di Manganese) come questi modelli abbiano delle difficoltà a prevedere i valori molto alti, anche se riescono, comunque, a seguire l'andamento della serie. Questo suggerisce l'opportunità di orientarsi verso modelli capaci di cogliere le caratteristiche di non linearità del problema; un altro possibile approccio potrà essere quello dell'analisi dei valori estremi. Ulteriori miglioramenti delle capacità predittive dei modelli sono attesi dall'estensione, mirata, dell'insieme delle variabili utilizzabili come predittori:

- con l'implementazione, presso l'ARPA FVG, di algoritmi di calcolo capaci di stimare le grandezze che descrivono la stabilità atmosferica nello Strato Limite Planetario (altezza di rimescolamento, flusso di calore sensibile dal suolo, lunghezza di Monin Obukhov, classe di stabilità di Pasquill-Gifford);
- con l'implementazione dell'inventario delle emissioni in atmosfera nella provincia di Udine e la conseguente capacità di stimare gli apporti inquinanti dovuti ad altre sorgenti (riscaldamento, traffico, impianti industriali, agricoltura, ecc.).

Importanti benefici, in termini di conoscenza del problema ed interpretazione dei dati, potranno avvenire:

- dall'implementazione di modelli deterministici di dispersione degli inquinanti, capaci di fornire simulazioni attendibili del fenomeno di interesse;
- dall'analisi dei dati di una campagna di misura svolta durante l'intero 2006, in un sito urbano, in cui sono stati rilevati gli stessi parametri acquisiti nel sito industriale di cui si tratta.





# APPENDICE

## A.1 - L'implementazione di funzioni in R

In questo paragrafo viene esposta la funzione che è stata utilizzata per estrapolare i dati dal dataset del NCDC.

Di seguito è riportato il modo in cui si presentano i dati di un lancio:

```

                                ---COLUMN NUMBER---
      1           2           3           4           5           6           7
LINTYP
                                header lines
254      HOUR      DAY      MONTH      YEAR      (blank)      (blank)
  1      WBAN#      WMO#      LAT D      LON D      ELEV      RTIME
  2      HYDRO      MXWD      TROPL      LINES      TINDEX      SOURCE
  3      (blank)      STAID      (blank)      (blank)      SONDE      WSUNITS

                                data lines
  9      PRESSURE      HEIGHT      TEMP      DEWPT      WIND DIR      WIND SPD
  4
  5
  6
  7
  8
```

dove (sono riportate dalla leggenda solo le informazioni importanti ai fini dell'estrazione dei dati):

- LINTYP: è il tipo di identificazione di riga;
- 254 = indica un nuovo sondaggio nel file di output;
- HOUR: ora di lancio, riportata in UTC;

- DAY: giorno di lancio;
- MONTH: mese di lancio;
- YEAR: anno di lancio;
- WSUNITS: unità di misura della velocità del vento (in questo caso era stato selezionato "ms" che sta per decimetri di metri al secondo);
- PRESSURE: in millibars;
- HEIGHT: altezza, in metri, in cui si è raggiunta la pressione;
- TEMP: temperatura, in decine di gradi Celsius;
- DEWPT: temperatura del punto di rugiada;
- WIND DIR: direzione del vento, in gradi;
- WIND SPD: velocità del vento (espresso a seconda dell'unità di misura scelta);

Quindi per ogni lancio ci sono innumerevoli informazioni e i dati di ogni lancio è posto uno sotto l'altro in modo continuo. Il problema è l'impossibilità di utilizzare i dati posti in tale modo.

Da qui la necessità di implementare una funzione in R. Questa funzione è creata in modo che dato il file originale si estrae i dati di una pressione assegnata.

Qui di seguito è riportato il codice di tale funzione (le righe precedute dal simbolo # rappresentano i commenti):

```
radiosondaggio<-function(file,pressione)
{
  #metto il sistema in inglese
  Sys.setlocale("LC_TIME", "C")
  #numero delle colonne
  rows = nrow(file)
  #vedo in che righe del file ci sono indicazioni sulle date
  begin<-(1:rows)[file$V1==254]
  n<-length(begin)
  #estraggo la prima data
  inizio<- as.Date(paste(file[begin[1],3], file[begin[1],4],
    file[begin[1],5]), "%d %b %Y")
}
```

```

#estraggo l'ultima data
fine<- as.Date(paste(file[begin[n],3], file[begin[n],4],
                    file[begin[n],5]), "%d %b %Y")
#creo una sequenza senza ripetizioni di date
giorni<-as.Date(seq(inizio,fine,by="days"))
# calcolo il numero di giorni (senza ripetizioni)
m<-length(giorni)
#creo il dataframe minimo (21 colonne: una per le date e poi per ogni
#lancio giornaliero (al max 4) estraggo 5 variabili: Height, Temp,
#Dewpt, Wind.dir e Wind.spd
out <- as.data.frame(matrix(NA, m, 21))
#variabile che indica la riga
line <- 0
for (i in 1:rows)
{
  #se nella colonna 1 del file originale i valori sono 1,2 e 3, questi
  #non vanno bene
  if (file[i, 1] > 3)
  {
    #vedo la data
    if (file[i, 1] == 254)
    {
      #prendo la data
      data <- as.character(paste(file[i,3],file[i,4],file[i,5]))
      #prendo l'ora
      ora <- as.numeric(as.character(file[i,2]))
      #a seconda dell'ora assegno un valore a offset
      #la riga di codice sotto equivale a questi 4 if considerando
      #4 intervalli di tempo:
      #ora=5<-(3,8) ; ora=11<-(9,14) ; ora=17<-(15,20) ; ora=23<-(21,2)
      #if (ora == 5) {offset <- 0}
      #if (ora == 11) {offset <- 1}
      #if (ora == 17) {offset <- 2}
      #if (ora == 23) {offset <- 3}
      offset <- as.numeric((abs(ora-11)<3) + 2*(abs(ora-17)<3) +
                          3*(abs(ora-23)<3) + 3*(ora<3))
      #nel caso in cui la prima osservazione ha ora=(0,1,2) allora
      #devo cambiargli la data (corrisponde al lancio delle 23)
      if (line == 0 && ora<3) {data<-as.character(as.Date(data, "%d %b
                          %Y") - 1)}
      #se è la prima data aggiungo 1 a line e scrivo la data
      if ( line == 0 )
      {

```

```

    line <- line + 1
    out[line, 1] <- as.character(data)
  }
#se la data è diversa da quella già scritta aggiungo 1 a line e
#scrivo la data
if ((data != out[line, 1]) & (ora>=3) )
{
  #differenza di giorni
  b<-as.Date(out[line,1],"%d %b %Y")
  a<-as.numeric(as.Date(data,"%d %b %Y") - b)
  #vedo se ci sono giorni mancanti nel mezzo,e se si li aggiungo
  if (a>1)
  {
    for (i in 1:(a-1) )
    {
      line<-line + 1
      out[line, 1] <- (b + i)
    }
  }
  #dopo aver aggiunto i giorni mancanti metto l'ultima data
  #osservata
  line <- line + 1
  out[line, 1] <- as.character(data)
}

}

#se nella prima colonna il valore non è 254, allora ho il dato a
#seconda della pressione
else
{
  #a seconda della pressione scrivo il dato nel dataframe
  if (file[i,2] == pressione)
  {
    out[line, 1 + (offset * 5) + 1] <-
      as.numeric(as.character(file[i,3]))
    out[line, 1 + (offset * 5) + 2] <-
      as.numeric(as.character(file[i,4]))
    out[line, 1 + (offset * 5) + 3] <-
      as.numeric(as.character(file[i,5]))
    out[line, 1 + (offset * 5) + 4] <-
      as.numeric(as.character(file[i,6]))
  }
}

```

```

        out[line, 1 + (offset * 5) + 5] <-
            as.numeric(as.character(file[i,7]))
    }
}
}
}
#trasformo la data in tipo Date
#metto direttamente il vettore giorni!!!
out[,1]<-giorni
#assegno i nomi alle colonne
colnames(out)<-c("Date", "Height5", "Temp5", "Dewpt5", "WindDir5",
    "WindSpd5", "Height11", "Temp11", "Dewpt11", "WindDir11",
    "WindSpd11", "Height17", "Temp17", "Dewpt17", "WindDir17",
    "WindSpd517", "Height23", "Temp23", "Dewpt23", "WindDir23",
    "WindSpd23")
#cambio la dicitura dei valori mancanti
out[out[,]==99999]<-NA
return (out)

```

Alla fine viene ritornata la matrice out, una matrice che contiene i valori di tutte le variabili dato un valore di pressione pre-indicata.

## ***A.2 - Elenco delle variabili***

In questo paragrafo sono indicate tutte le variabili che sono state utilizzate ai fini della modellazione.

<i>spillato1</i>	quantità acciaio spillato (cioè fuso e colato) dal forno 1
<i>spillato2</i>	quantità acciaio spillato dal forno 2
<i>fz</i>	variabile fattore che indica il modo in cui opera l'azienda: assume valore 0 (entrambe i forni spenti), 1 (lavora solo il forno 1), 2 (lavora solo il forno 2) e 3 (lavorano entrambe i forni)
<i>v1p.d</i>	vento medio diurno parallelo (con fuoco il forno1) al tempo t
<i>v1o.d</i>	vento medio diurno ortogonale al tempo t
<i>v1p.n</i>	vento medio notturno parallelo al tempo t

<i>v1o.n</i>	vento medio notturno ortogonale al tempo t
<i>v1p.d1</i>	vento medio diurno parallelo al tempo t-1
<i>v1o.d1</i>	vento medio diurno ortogonale al tempo t-1
<i>v1p.n1</i>	vento medio notturno parallelo al tempo t-1
<i>v1o.n1</i>	vento medio notturno ortogonale al tempo t-1
<i>temp.d</i>	temperatura media diurna al tempo t
<i>temp.n</i>	temperatura media notturna al tempo t
<i>temp.d1</i>	temperatura media diurna al tempo t-1
<i>temp.n1</i>	temperatura media notturna al tempo t-1
<i>um.d</i>	umidità media diurna al tempo t
<i>um.n</i>	umidità media notturna al tempo t
<i>um.d1</i>	umidità media diurna al tempo t-1
<i>um.n1</i>	umidità media notturna al tempo t-1
<i>press.d</i>	pressione media diurna al tempo t
<i>press.n</i>	pressione media notturna al tempo t
<i>press.d1</i>	pressione media diurna al tempo t-1
<i>press.n1</i>	pressione media notturna al tempo t-1
<i>pioggia</i>	pioggia caduta al tempo t
<i>pioggia1</i>	pioggia caduta al tempo t-1
<i>radiazioni</i>	irraggiamento solare al tempo t
<i>radiazioni1</i>	irraggiamento solare al tempo t-1
<i>radio.height11</i>	altezza per una pressione di 8500 mmbar del radiosondaggio delle 11
<i>radio.temp11</i>	temperatura del radiosondaggio delle 11
<i>radio.dewpt11</i>	temperatura del punto di rugiada del radiosondaggio delle 11
<i>radio.height23</i>	altezza del radiosondaggio delle 23
<i>radio.temp23</i>	temperatura del radiosondaggio delle 23
<i>radio.dewpt23</i>	temperatura del punto di rugiada del radiosondaggio delle 23

---

<i>Mn</i>	Manganese della centralina A
<i>Fe</i>	Ferro
<i>Ni</i>	Nichel
<i>Cd</i>	Cadmio
<i>Pb</i>	Piombo
<i>Zn</i>	Zinco
<i>Cr</i>	Cromo
<i>V</i>	Vanadio
<i>As</i>	Arsenico





## BIBLIOGRAFIA

- [1] Borroni (2005), *Profilo di rischio e soluzioni: acciaieria elettrica*, Dipartimento di Chimica, Materiali e Ingegneria Chimica, Politecnico di Milano, Milano (Italia), cap. 2; pagg. 34-45;
- [2] Borroni (2005), *Profilo di rischio e soluzioni: acciaieria elettrica*, Dipartimento di Chimica, Materiali e Ingegneria Chimica, Politecnico di Milano, Milano (Italia), cap. 4; pagg. 190-220;
- [3] APAT (2003), *Il ciclo industriale dell'acciaio del forno elettrico in Italia*, rapporto 38/2003, Roma, cap. 2, pagg. 75-123;
- [4] ARPA FVG (2005), Dipartimento provinciale di Udine, Servizio Tematico Analitico, nota Prot. n. 4987/05 dd 12 maggio 2005;
- [5] A. Borroni, B. Mazza, G. Nano (1989), *Controllo delle emissioni eariformi da forno elettrico ad arco: un obiettivo per l'ambiente di lavoro e per l'ambiente esterno*, Dipartimento Chimica Fisica Applicata, Politecnico di Milano, Milano;
- [6] A. Stefana, S. Tenini (2003), *Considerazioni sulle emissioni originate dal ciclo produttivo dell'acciaio attraverso l'utilizzo di impianti dotati di forno elettrico ad arco*, Brescia Ricerche, n. 42, pagg.38-45;

- [7] C. Grassotti, S. Tenini, ARPA Brescia (2005), *Confronto sulle emissioni originate dal ciclo produttivo dell'acciaio mediante forno tradizionale eaf e forno a caricamento continuo tipo "Conteel"*, Brescia Ricerche, n. 52, pagg. 21-24;
- [8] A. Borroni, G. Nano, R. Rota (2003), *Criticità delle emissioni diffuse con forno elettrico per la produzione di acciaio*, Giornale degli Igienisti Industriali, vol. 28, n.4, pagg. 203-214;
- [9] P. Prati, A. Zucchiatti, F. Lucarelli, P. A. Mandò (2000), *Source apportionment near a steel plant in Genoa (Italy) by continuous aerosol sampling and PIXE analysis*, Atmospheric Environment, n.34, pagg. 3149-3157;
- [10] F. Valerio, A. Stella, E. Daminelli (2004), *Identificazione delle fonti di emissioni di Idrocarburi Policiclici Aromatici (IPA) e benzene: l'esperienza di Genova-Cornigliano*, IST – Istituto Nazionale per la Ricerca sul Cancro, Conferenza su "La prevenzione primaria dei tumori di origine industriale ed ambientale in una società moderna", 8 novembre 2004;
- [11] Ente Nazionale Italiano di Unificazione – UNI (2005), *Norma Europea UNI EN 14902 – Qualità dell'aria ambiente: metodo normalizzato per la misurazione di Pb, Cd, As e Ni nella frazione PM10 del particolato in sospensione*, Milano;
- [12] R Development Core Team - R Foundation for Statistical Computing (2006) *A Language and Environment for Statistical Computing*, <http://www.R-project.org>, Vienna (Austria).

## RINGRAZIAMENTI

Dopo aver trascorso altri due anni, di studio e divertimento mi ritrovo a dire un grazie grande come una casa a tantissime persone. Per questo se mi dimentico di qualcuno (e spero proprio di no) chiedo venia!!!!

Per cominciare devo tutto questo a mamma e papà: mi domando ancora come siano riusciti a sopportarmi prima di un esame...ok, adesso capisco perché a settembre mentre preparavo "statistica cp" se ne sono scappati in vacanza!!!

Grazie a Clara, il mio grillo parlante: se non ci fossi ti dovrei inventare!

Irene, sempre splendida e di cui le invidio l'allegria e spensieratezza (ma prima o poi imparerò a esserlo).

Anna, la mia Chipsyna, un pezzettino del mio cuore resta sempre prenotato per te.

Maida...si sa che gli incontri del sabato mattina in biblioteca per "studiare" mi mancano tanto tanto tanto!

Lele, prima o poi mi spiegherai come fai a fare 4 lavori contemporaneamente.

Grazie a Enrico per la concessione a termine del suo portatile.

I nonni, Lole, Gigio, Romina, Alex e gli zii Annibale e Ovidio a cui voglio tanto bene.

Un grazie a Davide, Andrea, Giorgia (mi mancano le nostre chiacchierate pre-notturne), Federica e Flavia (due ragazze splendide), Elena,

## RINGRAZIAMENTI

---

Antonella, Silvia e Claudia e tutti i miei compagni della Banda Tita Michelas.

Ringrazio il prof. S. Bordignon, sempre disponibile e il dott. F. Pauli che mi ha seguito con molta pazienza.

Francesco, un tutor veramente speciale e con cui ho discusso fino all'ultimo (in realtà mi son divertita un casino a creare sti confronti in cui ci si scannava a vicenda e alla fine non si arrivava a niente!!!)

Il dott. R. Villalta, Responsabile del Settore Tutela Qualità dell'Aria dell'ARPA FVG, che ha voluto questo lavoro (lo ammetto, a me incute sempre un po' di sana "paura da capo"!!!).

Stefania, dolce e simpatica al punto giusto (con un'aggiunta di peperoncino).

Tom, i suoi discorsi a doppio senso e inutili già mi mancano.

Luca, che rimmarrà sempre il dolce Luchino!

Il dott. E. Baiutti, è stata una piacevole scoperta ritrovarti all'ARPA e ti ringrazio ancora per i lavori e le informazioni che mi hai dato.

Il dott. F. Moimas, il dott. E. Salvagni per i dati, la documentazione e le profique discussioni.