

UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI SCIENZE STATISTICHE

CORSO DI LAUREA TRIENNALE IN
STATISTICA PER LE TECNOLOGIE E LE SCIENZE

**Studio dell'inquinamento atmosferico
nella Pianura Padana**

Relatrice:
PROF.SSA GIULIANA CORTESE

Laureando:
MARCO PICCOLO
1219147

Anno Accademico 2022/2023

Indice

Introduzione	1
1 L'inquinamento atmosferico	3
2 Raccolta e pulizia dei dati	5
2.1 Dati mancanti	7
2.2 Analisi esplorativa	9
3 I modelli e la loro applicazione	13
3.1 Applicazione del modello lineare normale	16
3.2 Applicazione del modello marginale	21
4 Analisi spaziale	25
5 Conclusioni	29

Introduzione

In questo studio viene analizzato l'inquinamento atmosferico presente nella Pianura Padana, concentrandosi sulle province delle cinque Regioni che la compongono, ovvero Piemonte, Lombardia, Veneto, Emilia Romagna e Friuli-Venezia Giulia. L'obiettivo è trovare un modello che spieghi il rapporto tra la qualità dell'aria nelle città e le emissioni provenienti dalle attività antropiche sparse nel territorio, tenendo in considerazione l'andamento negli anni di questi fenomeni e concentrandosi su tre sostanze inquinanti distinte: $PM_{2.5}$, PM_{10} e NO_2 .

Nel primo capitolo viene introdotto il fenomeno dell'inquinamento atmosferico nel Nord Italia, indicandone brevemente cause ed effetti, ma anche le misure legislative attuate per mitigare questo problema.

Nel secondo capitolo si spiega come sono stati raccolti e ripuliti i dati necessari per lo studio, utilizzandoli poi per una breve analisi esplorativa, necessaria per comprendere il fenomeno in maniera preliminare.

Nel terzo capitolo, invece, vengono approfonditi e applicati i modelli scelti per questa ricerca, ovvero il modello lineare normale e il modello marginale, spiegando le proprietà teoriche, interpretando i risultati e verificando la bontà di adattamento per ciascuno dei tre inquinanti scelti.

Nel quarto capitolo viene mostrata una sintetica analisi spaziale dei dati, considerata la natura dei dati atmosferici usati in questo studio.

Infine nel quinto capitolo vengono tratte le conclusioni su quanto trovato e i possibili approfondimenti futuri che si potrebbero compiere.

Capitolo 1

L'inquinamento atmosferico

Con l'inizio della rivoluzione industriale l'inquinamento atmosferico è diventato sempre più un problema, soprattutto per la salute delle persone, accelerando i cambiamenti climatici che viviamo anche oggi^[1]. Questo fenomeno è presente prevalentemente in zone altamente popolate ed industrializzate, proprio come la Pianura Padana. In questo territorio del Nord Italia infatti vivono circa 15 000 000 di italiani e grazie a fattori come la fertilità, la vicinanza geografica dei mercati esteri, la superficie pianeggiante, il conseguente agevole collegamento tra città e la costruzione di infrastrutture più facile rispetto al resto d'Italia, la Pianura Padana gode di una robusta economia diversificata, grazie anche alla buona distribuzione della popolazione in diversi centri urbani medio-grandi che costituiscono ottimi punti di riferimento produttivi, logistici e industriali^[2].

A questi fattori si aggiungono la scarsa ventilazione, dovuta alla presenza delle catene delle Alpi nella parte settentrionale e occidentale e degli Appennini in quella meridionale, e il trasporto di inquinanti atmosferici a lunga distanza verso l'Europa da altri paesi dell'emisfero settentrionale^[3], che hanno portato ad un drastico peggioramento della qualità dell'aria dagli anni 60 in poi, con un forte problema di smog in molte città, portando la Pianura Padana ad essere una delle regioni europee con la peggior situazione di inquinamento atmosferico^[4].

La gravità di questa situazione comporta diversi effetti:

- sull'ambiente, con la comparsa dell'effetto serra e delle piogge acide^[1];
- sulla salute, con l'aumento del rischio di patologie respiratorie e cardiovascolari e con la riduzione dell'aspettativa di vita^[5];
- sull'economia, costando 5 trilioni di dollari come risultato di una peggior qualità di vita e di una minor produzione a causa del maggior numero di

morti da malattie provocate dall'inquinamento^[6].

L'estensione dei danni del fenomeno dell'inquinamento atmosferico ha portato all'adozione di diverse misure legislative per controllarlo e ridurlo. Una delle prime direttive europee a riguardo è la Direttiva 96/62/CE del 1996, la quale mira a definire una strategia comune rivolta a stabilire standard di qualità dell'aria atti a prevenire o ridurre gli effetti nocivi degli inquinanti sull'ambiente. In Italia questa Direttiva è stata attuata con il Decreto Legislativo n. 351 del 1999, che definisce inoltre i ruoli delle Regioni e dello Stato: le Regioni si devono occupare dell'analisi della qualità dell'aria e dell'attuazione di piani sul territorio per prevenire o contenere situazioni oltre i valori limite degli inquinanti, lo Stato deve definire dei valori limite e dei valori obiettivo di qualità da raggiungere applicati a tutto il territorio. Fenomeni estremi, come quello dello smog, ha portato anche le amministrazioni provinciali e comunali a prendere provvedimenti sul tema dell'inquinamento atmosferico, istituendo zone a traffico limitato o creando grandi zone verdi nelle città.

L'implementazione di queste norme, insieme al miglioramento tecnologico delle sorgenti emissive, ha portato dall'inizio degli anni 2000 ad una graduale diminuzione della presenza di sostanze inquinanti nella zona della Pianura Padana^[7].

Capitolo 2

Raccolta e pulizia dei dati

Per l'obiettivo dello studio sono stati necessari dati storici sulla qualità dell'aria e sulle emissioni nelle province delle cinque regioni d'interesse, che inizialmente sono stati ricercati nei siti web delle rispettive Agenzie Regionali per la Protezione Ambientale (ARPA), ma non tutte rendono disponibili dati utili all'obiettivo dell'analisi poiché non presenti o non aggiornati. Tuttavia, dal 2011 è stata resa loro obbligatoria la condivisione dei dati all'Agenzia Europea sull'Ambiente (European Environment Agency) per monitorare la situazione atmosferica in tutta Europa^[8]. Nel sito web della EEA infatti è stato reperito il primo dataset, cioè l'Air Quality Health Risk Assessments, che raccoglie per ogni provincia degli Stati europei alcuni dati sugli inquinanti PM₁₀, PM_{2.5}, NO₂ e O₃ dal 2005 al 2020. In particolare sono raccolti: la concentrazione media annuale in $\mu\text{g}/\text{m}^3$, la concentrazione media annuale pesata per la popolazione della provincia in $\mu\text{g}/\text{m}^3$, le stime delle morti premature e degli anni di vita persi a causa dell'inquinamento, con il relativo intervallo di confidenza al 95%, calcolate con metodi definiti dall'OMS^{[9][10]}. La rilevazione degli inquinanti proviene da stazioni meteo sparse in tutto il territorio che, attraverso sensori che sfruttano proprietà fisiche e chimiche delle sostanze, permettono di avere ad ogni ora una misurazione affidabile della qualità dell'aria^[11].

Per le emissioni, inizialmente ci si è affidati all'Inventario Emissioni Aria (INEMAR), un database interregionale italiano che si occupa di stimare e salvare le emissioni in aria per ogni comune e per ogni attività umana dal 1999 in poi. Purtroppo però, per alcune regioni i dati non sono aggiornati e non è stato quindi possibile aggregarli coerentemente. In alternativa però, è stato reperito l'Inventario Nazionale delle Emissioni dell'Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), che raccoglie le emissioni di 42 inquinanti causate da 378

attività umane (selezionate secondo la nomenclatura SNAP97 della EEA^[12]) in ogni provincia italiana negli anni 1990, 1995, 2000, 2005, 2010, 2015 e 2019. La stima di queste emissioni avviene utilizzando il modello EMMA (EMission MAnager) dell'azienda Arianet, che combina la distribuzione spaziale delle attività umane e naturali e la speciazione chimica delle diverse sostanze inquinanti per dare un valore indicativo delle emissioni in un arco orario, giornaliero o annuale^[13]. Per il nostro scopo è stato deciso quindi di unire il dataset degli inquinanti della EEA e quello delle emissioni dell'ISPRA prendendo solo gli anni per cui i dati in entrambi i dataset sono disponibili, ovvero 2005, 2010, 2015 e 2019, e selezionando solamente le province che si trovano nella Pianura Padana completamente o che hanno la maggior parte degli abitanti nella Pianura. La figura 2.1 riporta la mappa della Pianura Padana con i capoluoghi di provincia selezionati.

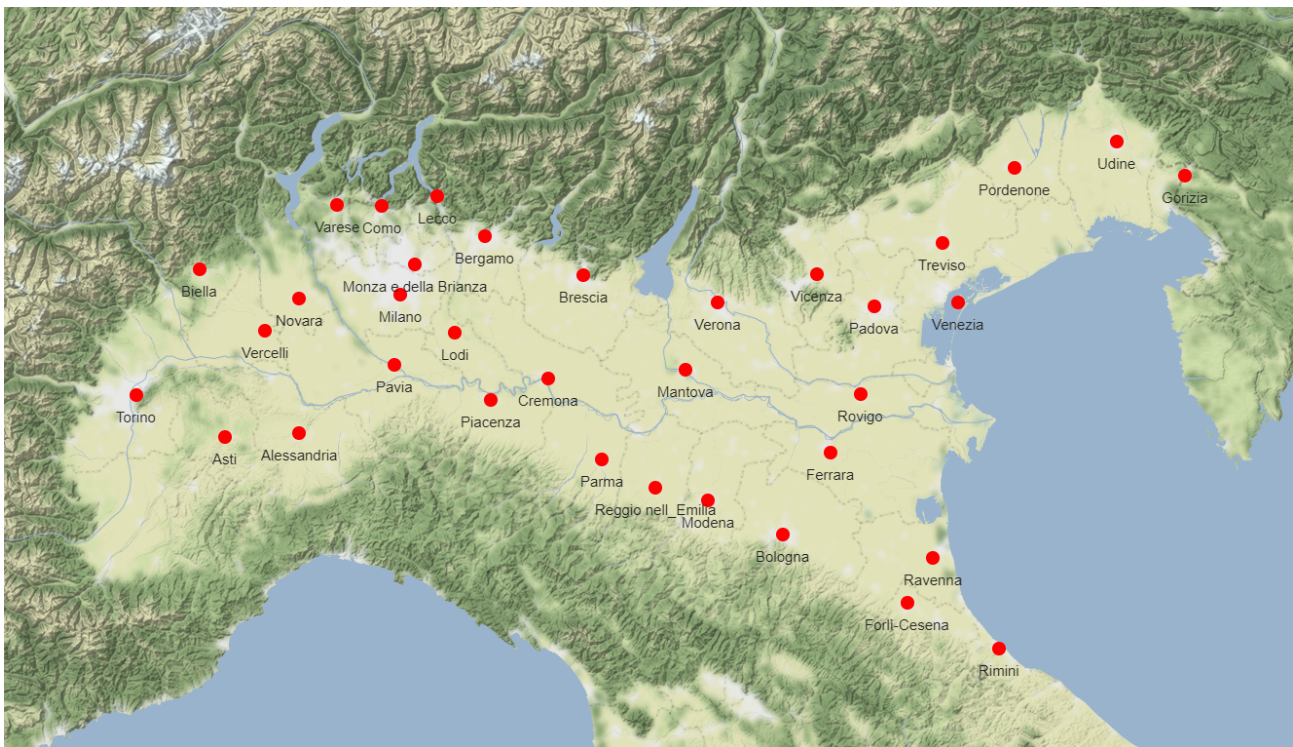


Figura 2.1: Mappa della Pianura Padana con i capoluoghi di provincia selezionati per lo studio.

Il dataset finale è formato da $n = 420$ osservazioni, equivalenti ai dati sui 3 inquinanti in 4 anni distinti per ciascuna delle 35 province, e $p = 21$ variabili, cioè:

<i>Nome variabile</i>	<i>Significato</i>
Prov	Provincia
Regione	Regione della provincia (Piemonte, Lombardia, Veneto, Emilia-Romagna, Friuli Venezia Giulia)
Anno	Anno di misurazione degli inquinanti e delle emissioni (2005, 2010, 2015, 2019)
Inquin	Tipo di inquinante (PM _{2.5} , PM ₁₀ , NO ₂)
Popol	Popolazione della provincia
Area	Superficie della provincia in km ²
Media_ann	Concentrazione media annuale di un certo inquinante misurata in $\mu\text{g}/\text{m}^3$
Media_ann_pes	Concentrazione media annuale di un certo inquinante pesata per la popolazione della provincia
Lat	Latitudine del capoluogo di provincia
Long	Longitudine del capoluogo di provincia
Agric	Emissioni in Mg dal settore dell'agricoltura
Solv	Emissioni in Mg dall'uso dei solventi
CombNonInd	Emissioni in Mg dalla combustione non industriale
CombInd	Emissioni in Mg dalla combustione industriale
ProcProd	Emissioni in Mg da processi produttivi senza combustione
ProdEn	Emissioni in Mg dalla combustione per la produzione e trasformazione di energia
Trasp	Emissioni in Mg dai trasporti su strada
AltriMacc	Emissioni in Mg da altri tipi di trasporti e macchinari
Rif	Emissioni in Mg dal trattamento e smaltimento dei rifiuti
EstrazFoss	Emissioni in Mg dall'estrazione e distribuzione dei combustibili fossili e dell'energia geotermica
AltreSorg	Emissioni in Mg da altre sorgenti

2.1 Dati mancanti

Per determinati macrosettori sono presenti dei dati mancanti in alcune province. Per evitare di perdere dati potenzialmente informativi per la stima dei modelli, sono stati stimati i valori utilizzando l'analisi spaziale. Viene assunto quindi un

modello spaziale gaussiano con variabile risposta l'emissione per il macrosettore d'interesse e come variabili esplicative le coordinate del capoluogo di provincia e la concentrazione media pesata annuale per inquinante. Per effettuare la previsione dell'emissione delle province con i dati mancanti si è sfruttato il Kriging, ovvero un metodo di previsione spaziale che, attraverso un sistema di pesi, dà più importanza alle osservazioni spaziali vicine al punto spaziale di previsione, con l'obiettivo di minimizzare l'errore quadratico medio. Dal variogramma empirico

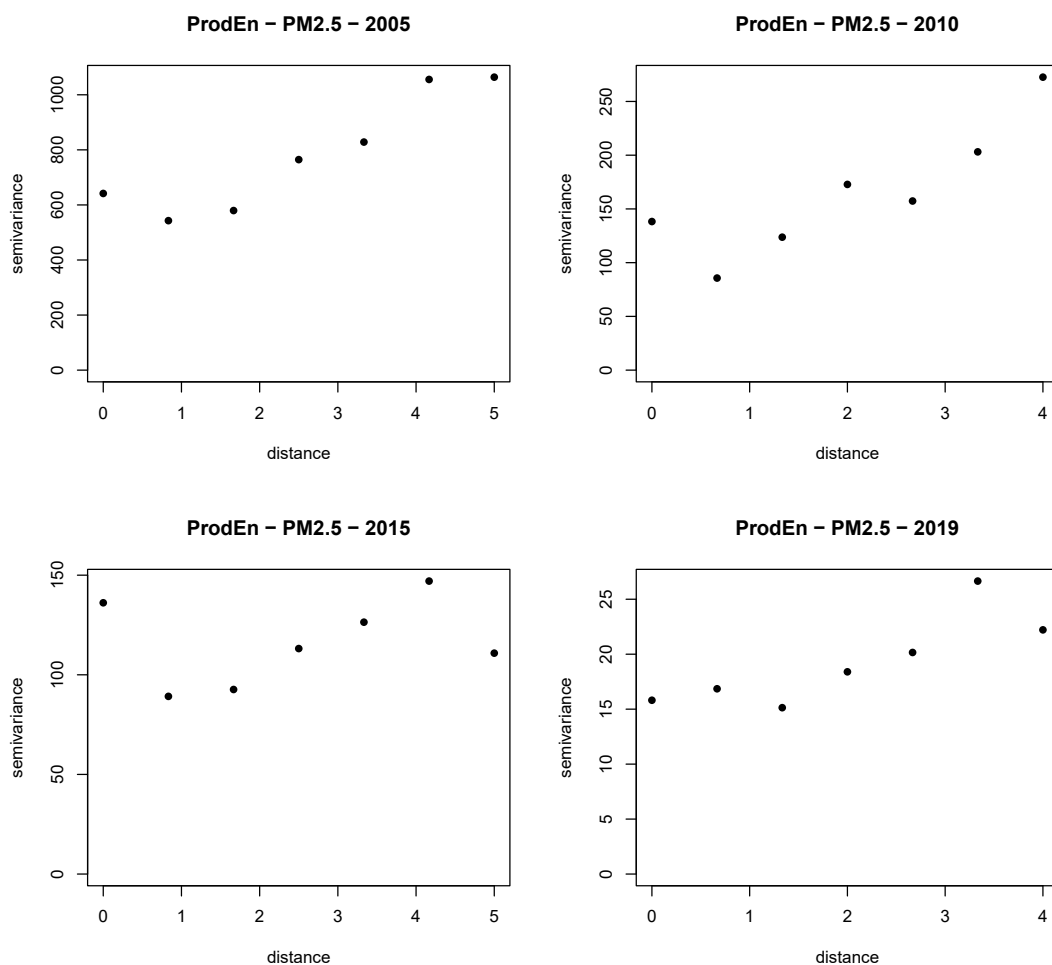


Figura 2.2: Variogrammi empirici per le emissioni di PM_{2.5} dal macrosettore *ProdEn* nei quattro anni 2005, 2010, 2015 e 2019

di questi dati, si vede però che i punti si distribuiscono in modo costante rispetto alla misura di distanza, piuttosto che avere un andamento crescente quando la distanza è piccola per poi stabilizzarsi all'aumentare della distanza. Questo suggerisce che non esiste una correlazione spaziale per questi punti o, in altre parole, che le emissioni di una certa provincia non sono influenzate dalle emissioni delle province vicine. Come esempio vengono riportati nella figura 2.2 i grafici dei va-

riogrammi empirici per le emissioni di $PM_{2.5}$ dal macrosettore della produzione di energia (*ProdEn*) nei quattro anni 2005, 2010, 2015 e 2019, dove si può notare come, per tutti i quattro grafici, la semivarianza resti circa costante al variare della distanza. Questo lo si nota per tutti i macrosettori e per ogni inquinante che presenta dati mancanti.

In questo caso il Kriging non risulta utile data l'assenza di correlazione spaziale tra punti vicini e, di conseguenza, i dati mancanti vengono sostituiti con la media delle rispettive emissioni di tutte le province di cui si hanno i dati. L'unico caso in cui non si è riusciti a sistemare i dati mancanti è l'emissione di NO_2 dal macrosettore dell'estrazione di combustibili fossili (*EstrazFoss*), in quanto non sono presenti per tutte le province in tutti gli anni.

2.2 Analisi esplorativa

Per un'analisi esplorativa più completa vengono considerati i dataset dell'EEA e dell'ISPRA, in modo tale da avere dati provenienti da più anni. Il grafico 2.3

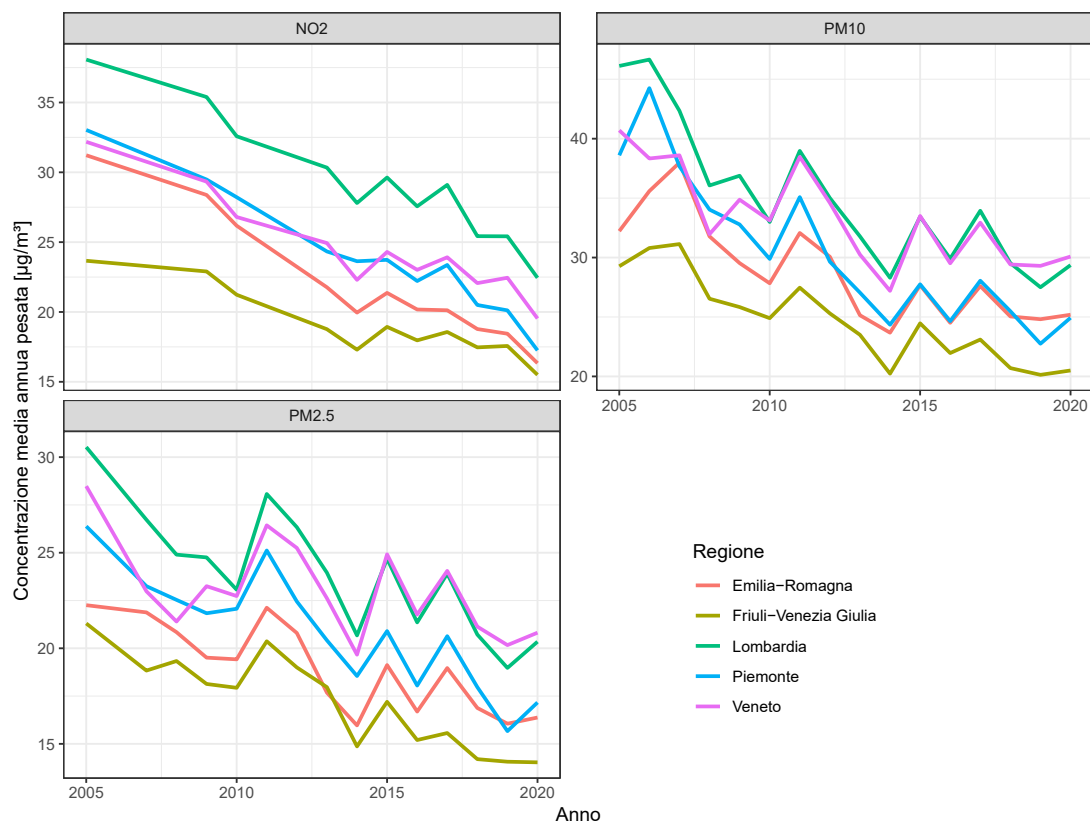


Figura 2.3: Andamento medio per regione della concentrazione media annua pesata dal 2005 al 2020

mostra come varia in media la concentrazione media annuale pesata dei tre in-

quinanti dal 2005 al 2020 in base alla Regione. In tutti i casi si riscontra un andamento decrescente, dovuto probabilmente all'introduzione di misure legislative mirate a contenere l'inquinamento. Nelle linee guida ONU del 2021 per la qualità dell'aria^[5], vengono poste delle soglie sulla concentrazione media annuale per ciascun inquinante con l'obiettivo di aiutare a comprendere e limitare il fenomeno dell'inquinamento atmosferico, nello specifico:

- $5 \mu\text{g}/\text{m}^3$ per $\text{PM}_{2.5}$
- $15 \mu\text{g}/\text{m}^3$ per PM_{10}
- $10 \mu\text{g}/\text{m}^3$ per NO_2

Guardando il grafico 2.3 vediamo come in media nelle diverse Regioni queste soglie siano ampiamente superate, con Lombardia e Veneto che presentano la situazione peggiore.

Nei grafici della figura 2.4, invece, sono rappresentati i totali delle emissioni per ogni inquinante nel 1990, 1995, 2000, 2005, 2010, 2015 e 2019 e differenziati per macrosettore. Anche in questo caso è presente un trend decrescente per le emissioni ed, inoltre, è possibile vedere come diversi macrosettori vadano ad impattare sulle emissioni. Tuttavia, si nota come la proporzione nel totale delle emissioni attribuita ad un certo macrosettore non resti sempre uguale, ma vari anch'essa, a causa probabilmente dell'entrata in vigore negli anni precedenti di normative ambientali mirate in quel macrosettore.



Figura 2.4: Totale delle emissioni di PM_{2.5}, PM₁₀ e NO₂ distinte per macrosettore negli anni 1990, 1995, 2000, 2005, 2010, 2015 e 2019.

Capitolo 3

I modelli e la loro applicazione

Nei modelli usati, come variabile risposta è stata scelta la concentrazione media annuale pesata per la popolazione della provincia (*Media_ann_pes*), una quantità che non è influenzata dalla numerosità di persone che abitano nella provincia, evitando quindi di inserire la variabile *Popol* come esplicativa. Il primo modello applicato ai dati è stato quello lineare normale, il quale impone la seguente relazione:

$$\mathbf{Y} = \mathbf{X}\underline{\beta} + \underline{\varepsilon}$$

o in forma matriciale:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & & \vdots \\ x_{n1} & \cdots & x_{np} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

con \mathbf{Y} vettore delle variabili risposta, \mathbf{X} matrice del modello, $\underline{\beta}$ vettore dei coefficienti delle variabili, $\underline{\varepsilon} \sim N(\mathbf{0}, \sigma^2 I_n)$ vettore degli errori casuali, $n = 420$ e p pari al numero di variabili selezionate. Le assunzioni di questo modello sono:

- *Linearità* nei coefficienti;
- *Omoschedasticità*, ovvero $\text{Var}(Y_i) = \sigma^2 \quad \forall i, i = 1, \dots, n$;
- *Normalità*, cioè $\mathbf{Y} \sim N_n(\mathbf{X}\underline{\beta}, \sigma^2 I_n)$;
- *Indipendenza* tra le osservazioni.

L'ultima assunzione però si va a scontrare con i risultati ottenuti nell'analisi esplorativa, dove si trova sia una certa correlazione temporale sia spaziale. I dati in analisi infatti sono dati longitudinali, ovvero raccolti nelle stesse province in anni

diversi, quindi è normale aspettarsi correlazione tra gli anni. Inoltre trattandosi di dati riguardanti l'atmosfera, sarebbe scorretto pensare che le sostanze emesse in una certa provincia non si spostino in quelle limitrofe, che invece possono avere un peggioramento della qualità dell'aria. Nonostante ciò, questo modello è stato applicato nella fase iniziale delle analisi, per la sua semplicità di interpretazione dei risultati e facilità di adattamento ai dati.

Per questi motivi il secondo modello utilizzato è stato il modello marginale per risposte normali, il quale fa parte dei modelli lineari generalizzati, in particolare dei modelli di quasi-verosimiglianza. Questo modello è utile proprio per dati longitudinali^[14]. Infatti si possono considerare le $m = 35$ province come unità statistiche tra di loro indipendenti e su cui vengono effettuate le misurazioni nei quattro anni presi in esame, quindi la variabile risposta per ogni provincia $\mathbf{Y}_k, k = 1, \dots, m$, sarà multivariata con $s = 4$ componenti dipendenti. Con questo modello, quindi, la correlazione spaziale non viene considerata preferendo di conseguenza quella temporale.

Questo modello viene definito marginale perché quello che si vuole descrivere è l'effetto delle esplicative sul valore atteso marginale della nostra variabile risposta, cioè:

$$E(\mathbf{Y}_k) = \underline{\mu}_k, \quad k = 1, \dots, m$$

con $\underline{\mu}_k$ vettore con componenti $\mu_{kj} = \mathbf{x}_{kj}\underline{\beta}$, $j = 1, \dots, s$, con \mathbf{x}_{kj} vettore delle variabili esplicative della provincia k al tempo j . Sotto questo modello inoltre la matrice di varianza e covarianza di ogni provincia è pari a:

$$\text{Var}(\mathbf{Y}_k) = \mathbf{V}_k = \sigma^2 R(\underline{\alpha}), \quad k = 1, \dots, m$$

dove $\sigma^2 = \text{Var}(Y_{kj}), j = 1, \dots, s$, che è quindi costante per ogni tempo misurazione, e $R(\underline{\alpha})$ è la matrice di correlazione $s \times s$ di \mathbf{Y}_k con parametri di correlazione $\underline{\alpha}$ che può essere definita a priori. Per stimare $\underline{\alpha}$ e σ^2 si può usare il metodo dei momenti, ottenendo $\hat{\underline{\alpha}}$ e $\hat{\sigma}^2$, invece per stimare $\underline{\beta}$, si utilizzano i minimi quadrati generalizzati, ovvero:

$$\hat{\underline{\beta}} = \left(\sum_{k=1}^m \mathbf{X}_k^\top \mathbf{V}_k \mathbf{X}_k \right)^{-1} \left(\sum_{k=1}^m \mathbf{X}_k^\top \mathbf{V}_k \mathbf{Y}_k \right)$$

con \mathbf{X}_k matrice delle variabili esplicative per la provincia k .

Nel 1986 Liang e Zeger^{[15][16]} mostrarono come, qualunque sia la struttura di correlazione di $R(\underline{\alpha})$, gli stimatori $\hat{\underline{\beta}}$ siano consistenti per $\underline{\beta}$ e che, sotto condizioni

di regolarità, per $m \rightarrow +\infty$ si ha che:

$$\sqrt{m} \left(\underline{\beta} - \hat{\underline{\beta}} \right) \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

Inoltre Σ/m è asintoticamente equivalente alla matrice di covarianza robusta o sandwich, definita come:

$$\left(\sum_{k=1}^m \mathbf{X}_k^\top \mathbf{V}_k^{-1} \mathbf{X}_k \right)^{-1} \left(\sum_{k=1}^m \mathbf{X}_k^\top \mathbf{V}_k^{-1} \text{Var}(\mathbf{Y}_k) \mathbf{V}_k^{-1} \mathbf{X}_k \right) \left(\sum_{k=1}^m \mathbf{X}_k^\top \mathbf{V}_k^{-1} \mathbf{X}_k \right)^{-1} \quad (3.1)$$

che viene stimata sostituendo le stime $\hat{\underline{\alpha}}$, $\hat{\sigma}^2$ ai rispettivi parametri $\underline{\alpha}$, σ^2 e stimando $\text{Var}(\mathbf{Y}_k)$ con $(\mathbf{y}_k - \hat{\underline{\mu}}_k)(\mathbf{y}_k - \hat{\underline{\mu}}_k)^\top$, con $\hat{\underline{\mu}}_k$ vettore della media campionaria della provincia k . Specificare la corretta struttura di correlazione, quella più vicina alla realtà, migliora tuttavia l'efficienza degli stimatori $\hat{\underline{\beta}}$, o in altre parole diminuisce il loro standard error. Le possibili strutture di correlazione prese in esame sono:

- *Indipendenza:*

$$R(\underline{\alpha}) = I_s$$

- *Equicorrelazione:*

$$R(\underline{\alpha}) = \begin{pmatrix} 1 & \alpha & \cdots & \alpha \\ \alpha & 1 & \cdots & \alpha \\ \vdots & \vdots & \ddots & \vdots \\ \alpha & \alpha & \cdots & 1 \end{pmatrix}$$

- *Autoregressione di ordine 1, o AR(1):*

$$R(\underline{\alpha}) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{s-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{s-2} \\ \alpha^2 & \alpha & 1 & \cdots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{s-1} & \alpha^{s-2} & \cdots & \alpha & 1 \end{pmatrix}$$

I parametri α in questo caso sono chiamati coefficienti di correlazione intra-classe. La libreria R utilizzata per applicare il modello marginale si chiama **gee** e per stimare i coefficienti β utilizza le equazioni di stima generalizzate (Generalized Estimating Equation, GEE), un metodo sviluppato sempre da Liang e Zeger nel

1986^{[15][16]} che può essere applicato a qualsiasi tipo di variabile risposta, non solo normale come nel caso di questo studio, e sono definite come:

$$\sum_{k=1}^m \mathbf{D}_k^\top \mathbf{V}_k^{-1} (\mathbf{y}_k - \underline{\mu}_k) = 0$$

dove \mathbf{D}_k è la matrice con elementi $\frac{\partial \mu_{kj}}{\partial \beta_r}$, $j = 1, \dots, s$, $r = 1, \dots, p$. La funzione `gee()` restituisce, oltre alle stime dei coefficienti $\underline{\beta}$, anche due tipi di stime per i rispettivi standard error naive e robusti. Le stime naive vengono calcolate ipotizzando che la struttura di correlazione specificata sia quella vera, usando la formula:

$$se(\hat{\beta}) = \sqrt{(\mathbf{X}^\top \mathbf{V}^{-1} \mathbf{X})^{-1}} = \sqrt{\left(\sum_{k=1}^m \mathbf{X}_k^\top \mathbf{V}_k^{-1} \mathbf{X}_k \right)^{-1}} \quad (3.2)$$

mentre le stime robuste vengono calcolate a partire dalla matrice di varianza e covarianza robusta (3.1). Più le stime naive e robuste sono simili più la struttura di correlazione assunta è quella corretta, poichè porta all'efficienza migliore per gli stimatori dei β .

3.1 Applicazione del modello lineare normale

Per applicare il modello lineare normale sono stati scelti per ciascun inquinante tre configurazioni con variabili diverse, per poi confrontarle e sceglierne la migliore. Qui di seguito vengono riportati i risultati dei vari modelli.

Prima configurazione

Nella prima configurazione si sono considerate come variabili le province e le diverse emissioni per macrosettore. Per adattare i modelli è stata utilizzata una selezione *backward* basata sul criterio AIC, ovvero inserendo inizialmente nel modello tutte le variabili d'interesse, per poi rimuovere iterativamente quelle che non risultano significative. Di seguito sono riportati i risultati dei modelli, da cui vengono omesse le stime per le singole province per una migliore leggibilità, trattandosi di 35 fattori.

Tabella 3.1: PM_{2.5} - prima configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	15.4709	2.4330	6.36	0.0000
Prov
AltreSorg	0.0065	0.0032	2.00	0.0483
Solv	-0.3857	0.1220	-3.16	0.0021
CombNonInd	0.0025	0.0016	1.55	0.1237
Trasp	0.0247	0.0066	3.75	0.0003
AltriMacc	0.0238	0.0073	3.27	0.0015
ProdEn	0.0369	0.0270	1.37	0.1753

Tabella 3.2: PM₁₀ - prima configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	24.7375	2.0682	11.96	0.0000
Prov
AltreSorg	0.0094	0.0037	2.52	0.0133
Solv	-0.3593	0.1154	-3.11	0.0024
Trasp	0.0297	0.0077	3.87	0.0002
AltriMacc	0.0340	0.0092	3.71	0.0003
ProdEn	0.0306	0.0225	1.36	0.1755

Tabella 3.3: NO₂ - prima configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	23.2689	2.0215	11.51	0.0000
Prov
AltreSorg	0.0058	0.0034	1.70	0.0931
Solv	-7.4534	1.8303	-4.07	0.0001
CombInd	0.0020	0.0006	3.26	0.0015
Trasp	0.0018	0.0003	6.27	0.0000

Seconda configurazione

Alla luce dei risultati dell'analisi esplorativa, in particolare dai grafici in figura 2.3, nella seconda configurazione è stata aggiunta la variabile continua *Anno* insieme all'interazione tra *Anno* e le diverse province, applicando sempre l'approccio *backward*. I modelli trovati sono i seguenti:

Tabella 3.4: PM_{2.5} - seconda configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1279.1633	108.0653	11.84	0.0000
Prov
Anno	-0.6240	0.0534	-11.69	0.0000
Solv	-0.1749	0.0828	-2.11	0.0372
CombInd	-0.0040	0.0027	-1.47	0.1437
CombNonInd	-0.0016	0.0011	-1.43	0.1562
Trasp	0.0137	0.0045	3.06	0.0028
AltriMacc	-0.0211	0.0063	-3.35	0.0012
EstrazFoss	0.6234	0.3466	1.80	0.0752

Tabella 3.5: PM₁₀ - seconda configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1490.6813	441.5428	3.38	0.0013
Prov
Anno	-0.7218	0.2194	-3.29	0.0017
Agric	0.0023	0.0019	1.16	0.2512
AltreSorg	0.0041	0.0022	1.88	0.0644
Solv	-0.5517	0.0876	-6.30	0.0000
CombInd	-0.0044	0.0035	-1.28	0.2038
ProcProd	-0.0107	0.0056	-1.93	0.0584
Trasp	0.0413	0.0073	5.65	0.0000
AltriMacc	-0.0167	0.0138	-1.21	0.2306
Rif	-0.1331	0.0629	-2.12	0.0383
Anno:Prov

Tabella 3.6: NO₂ - seconda configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1498.8593	270.8915	5.53	0.0000
Prov
Anno	-0.7328	0.1345	-5.45	0.0000
AltreSorg	0.0021	0.0016	1.31	0.1950
Solv	-2.0497	0.9014	-2.27	0.0262
Trasp	0.0004	0.0002	2.34	0.0223
AltriMacc	0.0004	0.0003	1.30	0.1994
Anno:Prov

Terza configurazione

Nella terza configurazione per completezza sono state aggiunte le interazioni tra *Anno* e i diversi macrosettori. I risultati dei modelli con queste variabili sono:

Tabella 3.7: PM_{2.5} - terza configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3715.6646	2271.7636	1.64	0.1079
Anno	-1.8366	1.1308	-1.62	0.1103
Prov
Agric	0.0189	0.0103	1.83	0.0732
AltreSorg	-0.8958	0.8728	-1.03	0.3094
Solv	-287.3240	88.9951	-3.23	0.0021
CombNonInd	-2.8496	0.8484	-3.36	0.0015
ProcProd	-0.0135	0.0074	-1.81	0.0761
Trasp	12.4236	4.2589	2.92	0.0052
AltriMacc	-11.6465	4.7812	-2.44	0.0182
Rif	33.5712	34.7342	0.97	0.3382
ProdEn	22.7158	11.4496	1.98	0.0524
EstrazFoss	0.9833	0.5480	1.79	0.0785
Anno:Prov
Anno:AltreSorg	0.0004	0.0004	1.03	0.3084
Anno:Solv	0.1430	0.0443	3.23	0.0022
Anno:CombNonInd	0.0014	0.0004	3.35	0.0015
Anno:Trasp	-0.0062	0.0021	-2.90	0.0053
Anno:AltriMacc	0.0058	0.0024	2.43	0.0183
Anno:Rif	-0.0168	0.0173	-0.97	0.3359
Anno:ProdEn	-0.0113	0.0057	-1.98	0.0527

Tabella 3.8: PM₁₀ - terza configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	649.9428	2606.1741	0.25	0.8041
Anno	-0.3048	1.2972	-0.23	0.8152
Prov
Agric	0.0040	0.0020	1.96	0.0557
AltreSorg	0.7081	0.7891	0.90	0.3738
Solv	-171.0062	84.1272	-2.03	0.0474
CombInd	2.6523	1.7799	1.49	0.1425
CombNonInd	-0.9133	1.0306	-0.89	0.3798
ProcProd	-0.0132	0.0060	-2.21	0.0314
Trasp	8.3870	4.5856	1.83	0.0734
AltriMacc	-17.0749	6.0378	-2.83	0.0067
Rif	37.2853	34.8266	1.07	0.2895
ProdEn	24.8240	11.2592	2.20	0.0321
EstrazFoss	-30.0875	25.9523	-1.16	0.2518
Anno:Prov
Anno:AltreSorg	-0.0004	0.0004	-0.89	0.3762
Anno:Solv	0.0849	0.0419	2.03	0.0481
Anno:CombInd	-0.0013	0.0009	-1.49	0.1424
Anno:CombNonInd	0.0005	0.0005	0.88	0.3824
Anno:Trasp	-0.0041	0.0023	-1.82	0.0752
Anno:AltriMacc	0.0085	0.0030	2.83	0.0068
Anno:Rif	-0.0186	0.0173	-1.07	0.2881
Anno:ProdEn	-0.0124	0.0056	-2.20	0.0322
Anno:EstrazFoss	0.0150	0.0129	1.16	0.2512

Tabella 3.9: NO₂ - terza configurazione

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	3532.5125	862.8976	4.09	0.0001
Anno	-1.7504	0.4310	-4.06	0.0001
Prov
Agric	-0.0010	0.0011	-0.93	0.3579
AltreSorg	0.0025	0.0017	1.48	0.1448
Solv	-2973.7047	1406.0678	-2.11	0.0387
CombNonInd	-1.8974	0.8459	-2.24	0.0287
Trasp	0.2710	0.1629	1.66	0.1015
AltriMacc	0.0006	0.0003	1.82	0.0732
Rif	0.0136	0.0148	0.92	0.3634
ProdEn	0.1309	0.1027	1.27	0.2075
Anno:Prov
Anno:Solv	1.4800	0.7004	2.11	0.0389
Anno:CombNonInd	0.0009	0.0004	2.25	0.0286
Anno:Trasp	-0.0001	0.0001	-1.65	0.1034
Anno:ProdEn	-0.0001	0.0001	-1.28	0.2069

Scelta della configurazione ed interpretazione

Per scegliere la configurazione migliore sono state prese in considerazione tre misure:

- l'*Akaike Information Criterion* (AIC):

$$AIC = -2 \ln(L) + 2p$$

- il *Bayesian Information Criterion* (BIC):

$$BIC = -2 \ln(L) + p \ln(n)$$

- l' R^2 corretto:

$$R_{adj}^2 = 1 - \left(\frac{n-1}{n-p-1} (1 - R^2) \right)$$

Tutte queste misure permettono di confrontare i modelli delle diverse configurazioni per capire quale si adatti meglio ai dati. I risultati sono i seguenti:

	PM _{2.5}			PM ₁₀			NO ₂		
	Mod. 1	Mod. 2	Mod. 3	Mod. 1	Mod. 2	Mod. 3	Mod. 1	Mod. 2	Mod. 3
AIC	718.1	604.95	588.6	816.05	631.84	613.1	781.72	512.46	508.05
BIC	841.65	731.43	847.47	936.65	864.23	880.7	899.39	733.09	752.2
R_{adj}^2	0.632	0.837	0.859	0.655	0.913	0.921	0.748	0.965	0.966

Da questi valori si può notare come, secondo l'AIC e l' R^2 corretto, la configurazione migliore per tutti e tre gli inquinanti sia la terza, invece per il BIC è la seconda. Tuttavia, prendendo in considerazione l'AIC e l' R^2 corretto, si può evincere come tra i modelli della seconda e della terza configurazione ci sia un miglioramento, ma meno importante rispetto a quello tra i modelli della prima e della seconda configurazione, questo per ciascun inquinante. Quindi, sfruttando il principio di parsimonia, la configurazione di variabili scelta per le tre sostanze inquinanti è la seconda.

È possibile osservare come le variabili *Anno* e *Prov* risultino fortemente significative per tutti e tre i modelli di questa configurazione, a prova del fatto che la concentrazione media annuale varia molto da anno ad anno e da una provincia ad un'altra. Per i PM₁₀ e l'NO₂ inoltre è significativa anche l'interazione tra le due variabili, probabilmente a causa di diverse misure messe in atto in certi anni da alcune province rispetto ad altre per contrastare queste sostanze. Per tutti e

tre i modelli il coefficiente della variabile *Anno* è negativo, a conferma del trend decrescente della concentrazione degli inquinanti trovato nell'analisi esplorativa preliminare. Per le emissioni dei diversi macrosettori, invece, risultano molte stime negative per i coefficienti, suggerendo che all'aumentare delle emissioni la concentrazione degli inquinanti dovrebbe diminuire, l'opposto di quello che ci si aspetterebbe. Questa discordanza, però, può essere causata dall'andamento non sempre lineare sia delle emissioni totali delle diverse province sia della concentrazione media annuale degli inquinanti, come visto anche nell'analisi esplorativa, che quindi va ad influenzare le stime dei coefficienti.

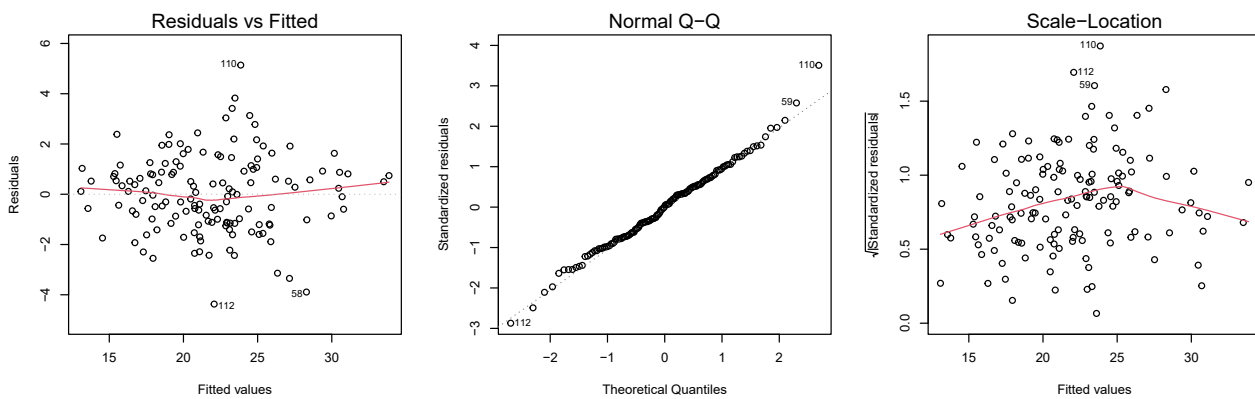


Figura 3.1: $PM_{2,5}$ - bontà di adattamento

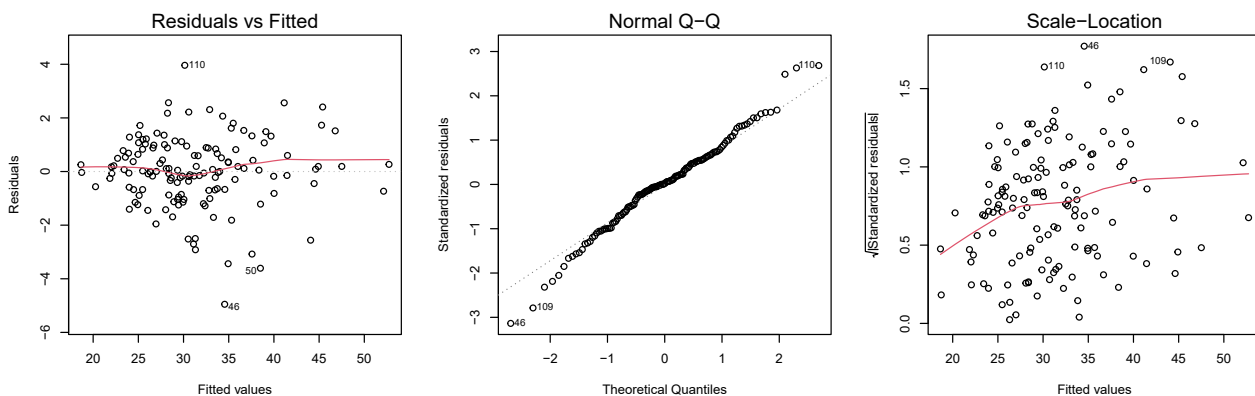


Figura 3.2: PM_{10} - bontà di adattamento

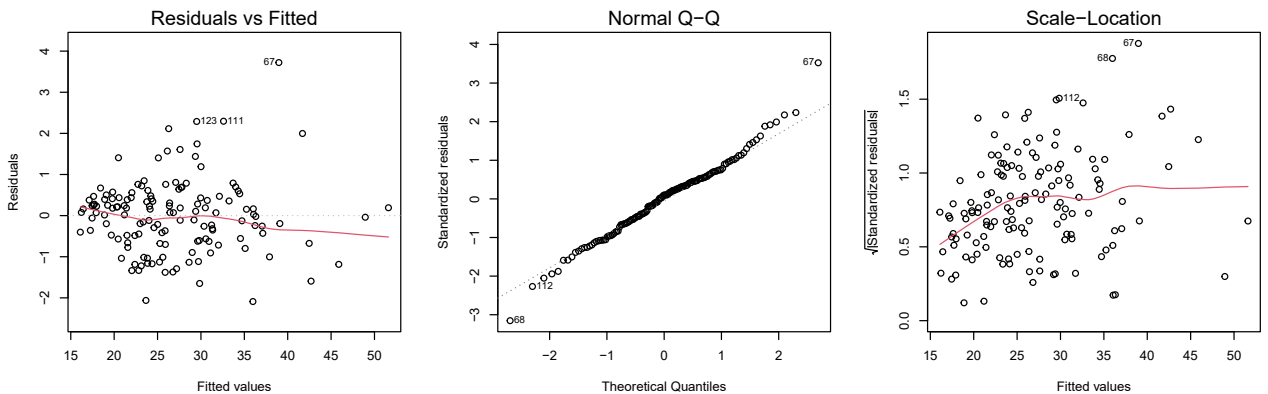


Figura 3.3: NO₂ - bontà di adattamento

Dai grafici 3.1, 3.2 e 3.3 si può trarre che i tre modelli hanno un buon adattamento ai dati, anche se si nota una leggera eteroschedasticità per il modello dei PM₁₀ e dell'NO₂.

3.2 Applicazione del modello marginale

Per applicare il modello marginale per risposte normali è necessario scegliere quale struttura di correlazione usare tra: indipendenza, equicorrelazione e AR(1). Per farlo si è preso, per ogni inquinante e per ogni struttura di correlazione, il modello con la variabile *Anno*, le variabili sui macrosettori e l'interazione tra *Anno* e quest'ultime, e si sono confrontate le stime degli standard error di ciascun coefficiente β calcolate con la matrice di varianza robusta 3.1 con quelle naive calcolate con la formula 3.2, in quanto sono misure di efficienza degli stimatori $\hat{\beta}$. Per fare questo si è usato il rapporto

$$\frac{SE_{robust}}{SE_{naive}} \quad (3.3)$$

per tutti i coefficienti, calcolandone infine la media complessiva del modello. Sapendo che più gli standard error robusti e naive sono simili più la struttura di correlazione assunta è quella corretta, poichè porta all'efficienza migliore, allora il modello con il rapporto (3.3) medio più vicino ad 1 sarà quello da scegliere. È importante notare come l'assunzione della struttura d'indipendenza ci porterebbe ad avere un modello lineare normale, quindi l'inclusione dell'indipendenza tra le possibili strutture di correlazione è utile anche per capire se effettivamente il modello marginale possa portare un miglioramento d'efficienza rispetto al modello

lineare normale applicato nella sezione precedente. I rapporti (3.3) medi dei vari modelli sono i seguenti:

Tabella 3.10: Rapporti (3.3) medi

	PM _{2.5}	PM ₁₀	NO ₂
Indipendenza	0.745	0.738	0.641
Equicorrelazione	0.707	0.804	0.671
AR(1)	0.724	0.764	0.7

Da questi risultati è interessante notare come la miglior struttura di correlazione non sia uguale per tutti gli inquinanti. In particolare:

- Per i PM_{2.5} la struttura migliore risulta essere quella d'indipendenza;
- Per i PM₁₀ la struttura migliore è l'equicorrelazione;
- Per l'NO₂ la struttura migliore è l'AR(1).

Bisogna sottolineare, però, come per ogni inquinante la differenza tra i tre rapporti medi degli standard error non sia così grande, ma si tratti di valori molto vicini tra di loro, da cui quindi non emerge una struttura migliore definitiva.

Eseguendo una selezione *backward* per i modelli con la struttura scelta per ogni inquinante, si ottengono i seguenti modelli:

Tabella 3.11: PM_{2.5} - modello marginale con struttura di correlazione d'indipendenza

	Estimate	Naive S.E.	Robust S.E.	Wald (robust)	Pr(> W)
(Intercept)	1225.25	152.37	136.89	80.11	0.00
Anno	-0.60	0.08	0.07	78.06	0.00
Agric	0.02	0.00	0.00	32.25	0.00
Solv	-8.05	11.20	7.90	1.04	0.31
Trasp	2.78	1.28	0.94	8.72	0.00
AltriMacc	-2.07	1.23	0.66	9.91	0.00
Rif	-0.02	0.01	0.00	24.56	0.00
EstrazFoss	0.89	0.28	0.34	6.84	0.01
Anno:Solv	0.00	0.01	0.00	1.09	0.30
Anno:Trasp	-0.00	0.00	0.00	8.76	0.00
Anno:AltriMacc	0.00	0.00	0.00	9.88	0.00

Tabella 3.12: PM₁₀ - modello marginale con struttura di correlazione di equicorrelazione

	Estimate	Naive S.E.	Robust S.E.	Wald (robust)	Pr(> W)
(Intercept)	1.6e+03	1.6e+02	2e+02	71	0
Anno	-0.79	0.078	0.097	69	1.1e-16
Solv	-0.11	0.065	0.053	4.4	0.036
CombNonInd	-0.0023	0.001	0.00087	7.1	0.0076
ProcProd	0.54	0.36	0.21	7.3	0.0069
Trasp	0.015	0.0047	0.0039	15	0.00013
AltriMacc	-1.5	1.4	0.64	5.7	0.017
EstrazFoss	0.091	0.046	0.03	9.7	0.0018
Anno:ProcProd	-0.00027	0.00018	0.0001	7.3	0.007
Anno:AltriMacc	0.00074	0.00067	0.00032	5.7	0.017
α	0.702	se = 0.042		279.37	0

Tabella 3.13: NO₂ - modello marginale con struttura di correlazione AR(1)

	Estimate	Naive S.E.	Robust S.E.	Wald (robust)	Pr(> W)
(Intercept)	1.7e+03	1.3e+02	1.1e+02	2.7e+02	0
Anno	-0.81	0.062	0.053	2.6e+02	0
AltreSorg	0.0023	0.0013	0.0007	13	0.00039
Solv	-2.5e+02	83	65	11	0.0008
Trasp	0.00069	0.00021	0.00018	16	7.8e-05
Rif	1.2	1.1	0.42	10	0.0015
ProdEn	-0.00033	0.0002	0.0001	10	0.0015
Anno:Solv	0.12	0.041	0.032	11	0.00082
Anno:Rif	-0.00061	0.00057	0.00021	10	0.0015
α	0.912	se = 0.029		31.45	2.05e-08

Anche in questo caso la variabile *Anno* risulta significativa e con coefficiente negativo, come nel modello lineare normale, dimostrando la progressiva diminuzione della concentrazione media annua degli inquinanti che sta avvenendo negli ultimi anni^[7]. Anche per i modelli marginali, inoltre, sono presenti coefficienti delle emissioni con segno negativo, ma la motivazione di questi risultati controintuitivi si può ritrovare nel trend storico irregolare delle emissioni e della concentrazione degli inquinanti. Ad eccezione del modello per i PM_{2.5}, la scelta di strutture di correlazione diverse dall'indipendenza fa comprendere che il modello marginale adottato sembra spiegare meglio il fenomeno dello studio per i PM₁₀ e per l'NO₂ rispetto al modello lineare normale, avendo inoltre per entrambi una stima della correlazione intra-classe α molto alta.

Capitolo 4

Analisi spaziale

In questo capitolo si è eseguita una breve analisi con lo scopo di verificare la presenza della correlazione spaziale tra le province.

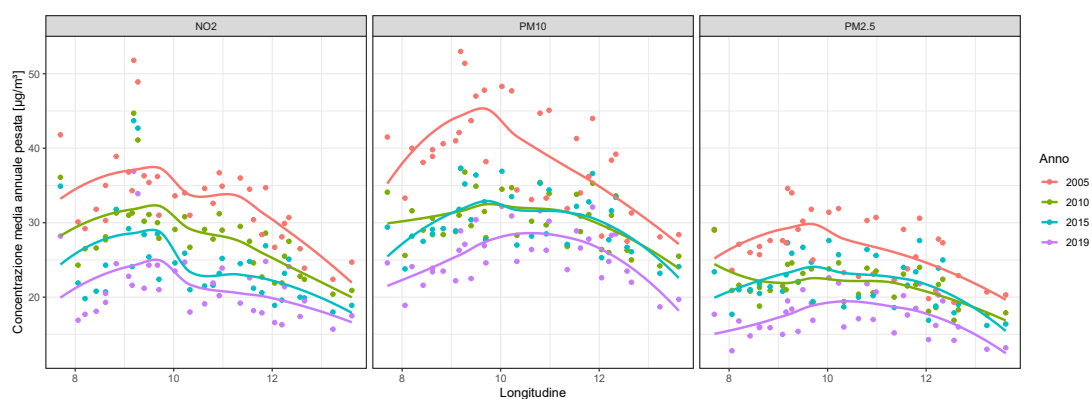


Figura 4.1: Concentrazione media annuale pesata dei tre inquinanti negli anni 2005, 2010, 2015 e 2019, al variare della longitudine dei capoluoghi di provincia.

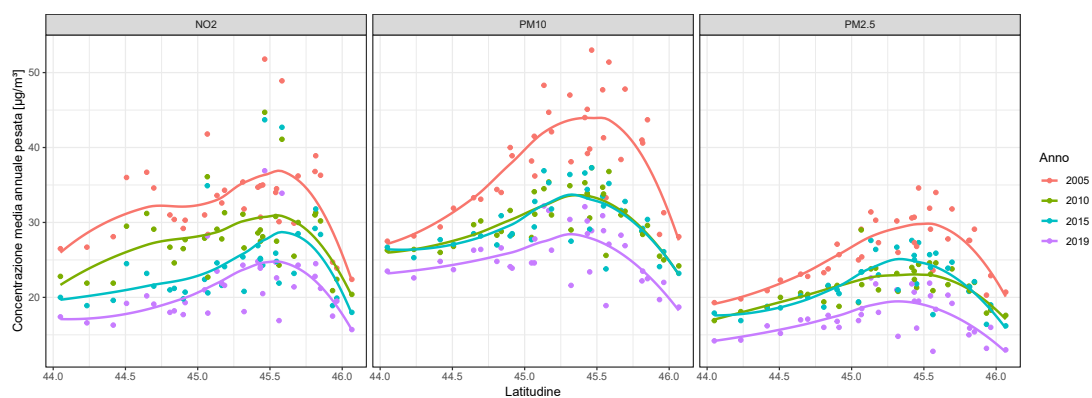


Figura 4.2: Concentrazione media annuale pesata dei tre inquinanti negli anni 2005, 2010, 2015 e 2019, al variare della latitudine dei capoluoghi di provincia.

Nei grafici 4.1 e 4.2 sono rappresentate le concentrazioni medie annue pesate dei tre inquinanti negli anni 2005, 2010, 2015 e 2019 al variare della longitudine e latitudine dei capoluoghi delle province della Pianura Padana, a cui è stato applicato anche un lisciamento per far risaltare meglio il trend. Si può osservare come per entrambe le coordinate sia presente un andamento spaziale parabolico per la concentrazione, ovvero c'è una più alta presenza di inquinanti nella parte centrale della Pianura rispetto alle estremità, dove troviamo le catene delle Alpi e degli Appennini e il Mar Adriatico.

Una quantità utile per verificare la presenza di correlazione spaziale è l'indice I di Moran, che si calcola con la formula:

$$I = \frac{m}{W} \frac{\sum_{i=1}^m \sum_{j=1}^m w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^m (x_i - \bar{x})^2}$$

dove, in questo caso, m è il numero di province, x_i è la concentrazione media annuale pesata della provincia i , \bar{x} è la media campionaria della concentrazione, w_{ij} sono dei pesi che sono stati posti pari all'inverso della distanza tra le province i e j e W è la somma di tutti i pesi. L'indice I varia da -1 a 1 e un valore pari a 0 indica che non è presente correlazione spaziale. Sotto l'ipotesi di non correlazione, il valore atteso dell'indice I è:

$$E(I) = \frac{-1}{m-1}$$

mentre la varianza è:

$$\text{Var}(I) = \frac{mS_4 - S_3S_5}{(m-1)(m-2)(m-3)(\sum_i \sum_j w_{ij})^2}$$

dove

$$\begin{aligned} S_1 &= \frac{1}{2} \sum_i \sum_j (w_{ij} + w_{ji})^2 & S_2 &= \sum_i \left(\sum_j w_{ij} + \sum_j w_{ji} \right)^2 \\ S_3 &= \frac{m^{-1} \sum_i (x_i - \bar{x})^4}{(m^{-1} \sum_i (x_i - \bar{x})^2)^2} & S_4 &= (m^2 - 3m + 3)S_1 - mS_2 + 3W^2 \\ S_5 &= S_1 - 2mS_1 + 6W^2 \end{aligned}$$

Assumendo la normalità dell'indice I , si possono usare queste due quantità, quindi, per condurre un test Z per verificare che I sia significativamente diverso da 0.

Calcolando questo indice per ogni inquinante e per ogni anno si ottengono i

	PM _{2.5}	PM ₁₀	NO ₂
2005	0.198	0.24	0.215
2010	0.085	0.093	0.235
2015	0.098	0.103	0.186
2019	0.1	0.119	0.146

Tabella 4.1: Indice I di Moran per le province calcolato sulla concentrazione media annuale pesata di ogni inquinante nel 2005, 2010, 2015 e 2019

valori nella tabella 4.1, i quali sono significativamente maggiori di 0, per cui si può concludere che è presente una correlazione spaziale positiva tra le diverse province, ma che negli anni si è ridotta.

Capitolo 5

Conclusioni

L'inquinamento atmosferico è un fenomeno molto complesso, soprattutto in aree popolate come la Pianura Padana, ed è, quindi, difficile concentrare i tanti fattori che lo causano in un unico risultato. Nonostante ciò, l'implementazione di modelli che spieghino correttamente la relazione tra le diverse unità statistiche e tra le variabili, come nel caso del modello marginale, si può rivelare molto utile a comprendere meglio questo problema, per poi poter prevederlo e mitigarlo. Ci si può, quindi, ritenere soddisfatti dei risultati trovati, poichè dimostrano che l'idea di includere la correlazione temporale nel modello marginale è la giusta intuizione per spiegare meglio questo fenomeno.

Tuttavia, la presenza di risultati discordanti dalle aspettative, come le stime negative dei coefficienti delle emissioni o l'uso della struttura di indipendenza per il modello marginale dei $PM_{2.5}$, fanno capire come sia necessario approfondire meglio questi modelli. È importante sottolineare, però, come l'utilizzo di dati aggregati per provincia e per quattro anni distinti non sia stato completamente sufficiente per lo scopo di questo studio, per cui sarebbero stati preferibili dati riguardanti le singole stazioni meteo e con misurazioni mensili in anni contigui, ma che per problemi esterni non è stato possibile reperire.

L'aggiunta della correlazione spaziale nel modello e l'inserimento di variabili che tengano conto sia delle misure legislative messe in atto contro l'inquinamento sia dei fattori meteorologici, insieme a dati migliori e all'uso di modelli più complicati, sono delle possibili idee su come continuare questo tema e che porteranno sicuramente ad una miglior comprensione dell'inquinamento atmosferico.

Bibliografia

- [1] Ioannis Manisalidis et al. «Environmental and Health Impacts of Air Pollution: A Review». In: *Frontiers in Public Health* 8 (2020).
- [2] Piero Dagradi. *Pianura Padana*. 1994. URL: https://www.treccani.it/enciclopedia/pianura-padana_%28Enciclopedia-Italiana%29/.
- [3] European Environment Agency. *Inquinamento atmosferico*. 2020. URL: <https://www.eea.europa.eu/it/themes/air/intro>.
- [4] The World Air Quality Project. *Air Pollution in Europe: Real-time Air Quality Index Visual Map*. 2023. URL: <https://aqicn.org/map/europe/>.
- [5] World Health Organization. *7 million premature deaths annually linked to air pollution*. 2014. URL: <https://www.who.int/news/item/25-03-2014-7-million-premature-deaths-annually-linked-to-air-pollution>.
- [6] Institute for Health Metrics World Bank e Evaluation at University of Washington. *The Cost of Air Pollution - Strengthening the Economic Case for Action*. 2016.
- [7] A. Bigi e G. Ghermandi. «Long-term trend and variability of atmospheric PM₁₀ concentration in the Po Valley». In: *Atmospheric Chemistry and Physics* 14.10 (2014), pp. 4895–4907.
- [8] Commissione Europea, cur. *2011/850/UE - disposizioni di attuazione delle direttive 2004/107/CE e 2008/50/CE del Parlamento europeo e del Consiglio per quanto riguarda lo scambio reciproco e la comunicazione di informazioni sulla qualità dell'aria ambiente*. 2011.
- [9] *Health risks of air pollution in Europe – HRAPIE project. Recommendations for concentration–response functions for cost–benefit analysis of particulate matter, ozone and nitrogen dioxide*. WHO Regional Office for Europe, 2013.

- [10] World Health Organization. *WHO global air quality guidelines: particulate matter (PM_{2.5} and PM₁₀), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide*. World Health Organization, 2021, xxi, 273 p.
- [11] European Environment Agency. *How is air pollution measured?* 2023. URL: <https://eeadmz1-cws-wp-air02.azurewebsites.net/index.php/testing-about-aq-ereporting/>.
- [12] Agenzia Regionale per la Protezione Ambientale Veneto. *Metodologia EMEP-EEA*. 2022. URL: <https://www.arpa.veneto.it/++api++/temi-ambientali/aria/emissioni-di-inquinanti/metodologia-emep-corinair>.
- [13] Agenzia Regionale per la Protezione Ambientale Piemonte. *La modellizzazione delle emissioni*. 2019. URL: <https://www.arpa.piemonte.it/approfondimenti/temi-ambientali/aria/pressioni-emissive/cartella-pressioni-emissive-documentazione-e-dati-ambientali/la-modellizzazione-delle-emissioni>.
- [14] Alessandra Salvan, Nicola Sartori e Luigi Pace. *Modelli Lineari Generalizzati*. 2020.
- [15] Scott L. Zeger e Kung-Yee Liang. «Longitudinal data analysis using generalized linear models». In: *Biometrika* 73.1 (1986), pp. 13–22.
- [16] Scott L. Zeger e Kung-Yee Liang. «Longitudinal Data Analysis for Discrete and Continuous Outcomes». In: *Biometrics* 42.1 (1986), pp. 121–130.