

Analisi del PM10 in Italia

Daniele Cugnigni¹, Alessandro Fanesi², and Leonardo Genesin³

¹ Corso di laurea magistrale in Scienze Statistiche, matricola 2054519
daniele.cugnigni@studenti.unipd.it

² Corso di laurea magistrale in Scienze Statistiche, matricola 2057039
alessandro.fanesi@studenti.unipd.it

³ Corso di laurea magistrale in Scienze Statistiche, matricola 2056404
leonardo.genesin@studenti.unipd.it

1 Presentazione del problema

Il progetto consiste nell'analisi del PM10 in Italia. Il PM10 è una misura di particolato atmosferico, specificamente delle particelle sospese con un diametro inferiore a 10 micrometri. Queste particelle includono polvere, pollini, spore di muffe, particelle di carburante e altri materiali solidi o liquidi presenti nell'aria. Sono abbastanza piccole da poter essere inalate e possono penetrare nelle vie respiratorie superiori, raggiungendo anche i polmoni, pertanto possono essere un potenziale rischio per la salute umana. La presenza di PM10 nell'aria è spesso misurata in microgrammi per metro cubo ($\mu\text{g}/\text{m}^3$), in generale livelli più alti di PM10 sono associati a una maggiore preoccupazione per la salute collettiva.

L'analisi si sviluppa con l'obiettivo di rispondere a tre quesiti di interesse: la valutazione del livello di PM10 per area geografica; la valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti; la valutazione dell'effetto del lockdown sul livello di PM10.

2 Dati

I dati utilizzati per l'analisi sono stati reperiti da due Data Store, *Atmosphere Data Store*⁴ e *Climate Data Store*⁵, dell'organizzazione Copernicus, un ente finanziato dall'Unione Europea che si occupa di raccolta e analisi di dati di tipo ambientale. L'interesse si concentra su 30 città italiane: i capoluoghi di Regione (Ancona, Aosta, Bari, Bologna, Bolzano, Cagliari, Campobasso, Catanzaro, Firenze, Genova, L'Aquila, Milano, Napoli, Palermo, Perugia, Potenza, Roma, Torino, Trento, Trieste, Venezia) e altre nove città popolose (Alessandria, Brescia, Ferrara, Padova, Parma, Ravenna, Udine, Vicenza, Verona). In Figura 1 è riportata la disposizione geografica delle città considerate. Per ognuna di esse sono stati raccolti i dati relativi alle seguenti variabili:

- Livello di PM10 ($\mu\text{g}/\text{m}^3$)
- Velocità del vento a 10 metri di altezza (m/s)
- Temperatura a 2 metri di altezza (K)
- Precipitazioni totali (m per unità d'area)
- Livello di Co ($\mu\text{g}/\text{m}^3$)
- Livello di So₂ ($\mu\text{g}/\text{m}^3$)
- Livello di No₂ ($\mu\text{g}/\text{m}^3$)

⁴ <https://ads.atmosphere.copernicus.eu>

⁵ <https://cds.climate.copernicus.eu>

Le variabili relative ai fattori inquinanti sono state reperite dal dataset *CAMS European air quality reanalyses* del *Atmosphere Data Store*, mentre le variabili di natura metereologica sono contenute nel dataset *Climate and energy indicators for Europe from 1979 to present derived from reanalysis* del *Climate Data Store*. Il tempo di rilevazione va dal 1 Gennaio 2019 al 31 Dicembre 2020, con frequenza di rilevazione oraria.



Figura 1: Disposizione geografica delle città considerate

3 Analisi Preliminari

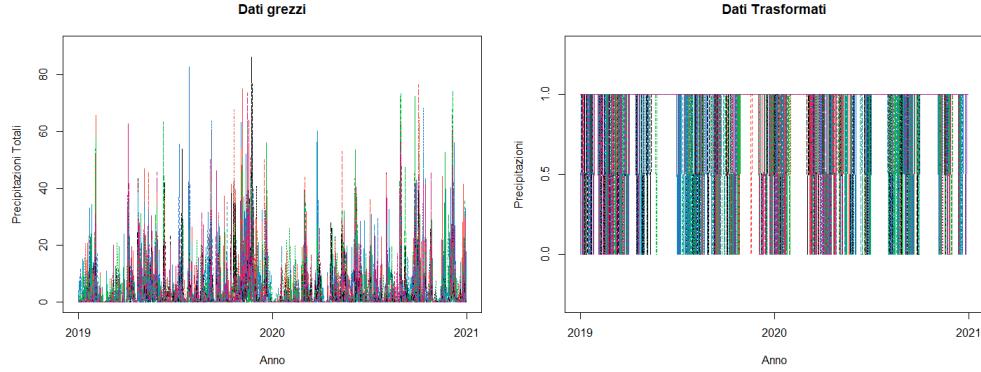
Prima di procedere con le analisi tramite modelli, sono state effettuate alcune operazioni di pulizia e lisciamento sui dati, oltre che alcune analisi esplorative.

3.1 Pulizia dei dati

Per ogni città le variabili rilevate con frequenza oraria sono state aggregate tramite la media giornaliera, in quanto è di interesse mantere l'informazione sull'intera giornata di rilevazione e non sulle singole ore. Poiché la variabile risposta *Livello di PM10* ha supporto positivo ne è stata considerata la trasformata logaritmica, in questo modo non è stato necessario imporre un vincolo di positività in fase di lisciamento ed in fase di modellazione. Per quanto riguarda le altre variabili considerate, queste sono state standardizzate (eccetto le *Precipitazioni totali*) in modo che l'ordine di grandezza delle singole non interferisse con le analisi effettuate con i modelli.

I dati relativi alla variabile *Precipitazioni totali* del dataset *CAMS European air quality* presentano 89 valori (orari) mancanti in corrispondenza del mese di Marzo 2020. Per questo

motivo, sono stati utilizzati per questo mese i dati sulle precipitazioni provenienti dal dataset *Agrometeorological indicators from 1979 to present derived from reanalysis* del *Climate Data Store* che fornisce i valori delle *Precipitazioni totali* con cadenza giornaliera. Questa variabile è stata in primo luogo dicotomizzata (1 se piove, 0 se non piove) poiché è di interesse mantenere solamente questo tipo di informazione piuttosto che quanti millimetri di pioggia sono stati registrati durante il giorno; in secondo luogo si è deciso di procedere con una regolarizzazione dell'andamento. Questa operazione è consistita nel considerare un giorno j -esimo di assenza di pioggia come un giorno di pioggia qualora il giorno $(j-1)$ -esimo ed il giorno $(j+1)$ -esimo fossero giorni di pioggia. L'idea alla base di questa operazione è legata all'ipotesi che se in un giorno non vengono rilevate piogge ma nel giorno precedente e successivo sì, evidentemente il periodo considerato è un periodo di maltempo in cui nel giorno in analisi non ha piovuto. In Figura 2 si mostrano i dati grezzi ed i dati trasformati.



3.2 Lisciamento

Avendo dei dati grezzi molto frastagliati, si è effettuato lisciamento delle variabili nel seguente modo:

- *Livello di PM10*: rappresentata con B-spline con $K = 30$ funzioni di base (Figura 3);
- *Velocità del vento*: rappresentata con B-spline con $K = 80$ funzioni di base (Figura 4);
- *Temperatura*: rappresentata con Basi di Fourier con $K = 25$ funzioni di base e periodo $T = 365$ (Figura 5);
- *Livello di Co*: rappresentata con B-spline con $K = 32$ funzioni di base (Figura 6);
- *Livello di So₂*: rappresentata con B-spline con $K = 40$ funzioni di base (Figura 7);
- *Livello di No₂*: rappresentata con B-spline con $K = 50$ funzioni di base (Figura 8);

3.3 Analisi esplorativa

Sono state condotte alcune brevi analisi esplorative sulla variabile risposta. In particolare in Figura 9 si riportano i grafici della media funzionale e dello standard error funzionale del *Livello*

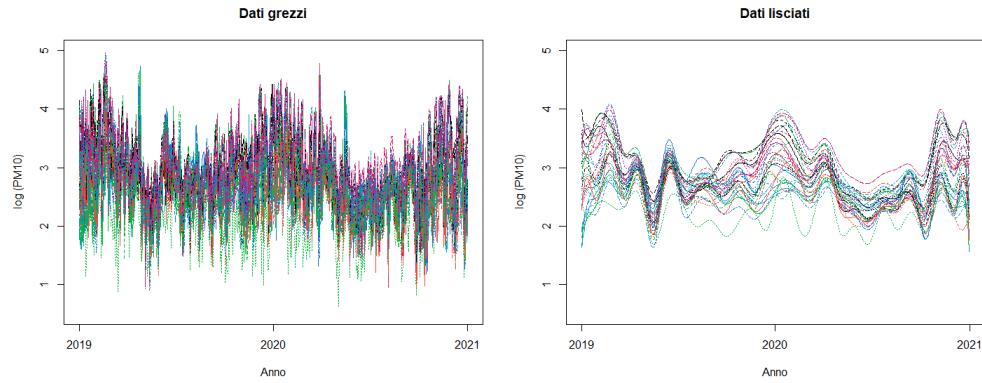


Figura 3: Variabile livello di PM10: dati grazzi (a sinistra) e dati lisciati (a destra).

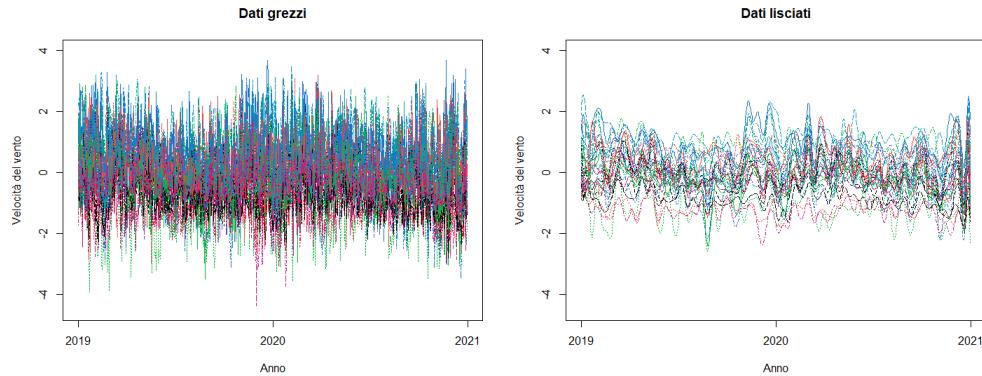


Figura 4: Variabile Velocità del vento: dati grazzi (a sinistra) e dati lisciati (a destra).

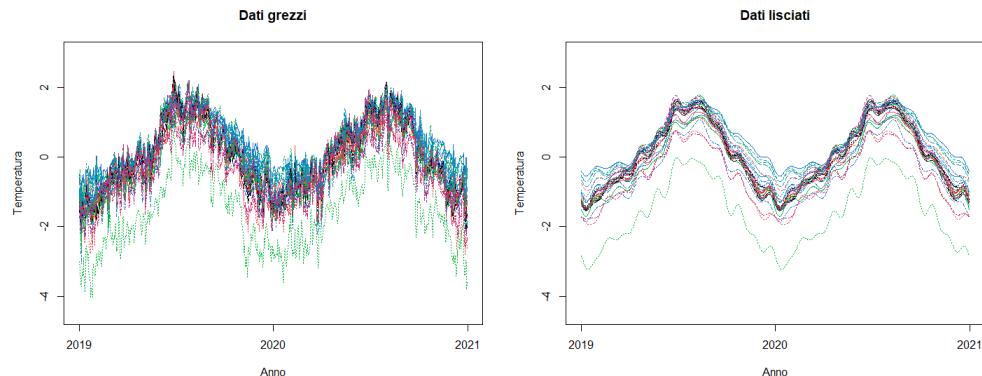


Figura 5: Variabile Temperatura: dati grazzi (a sinistra) e dati lisciati (a destra).

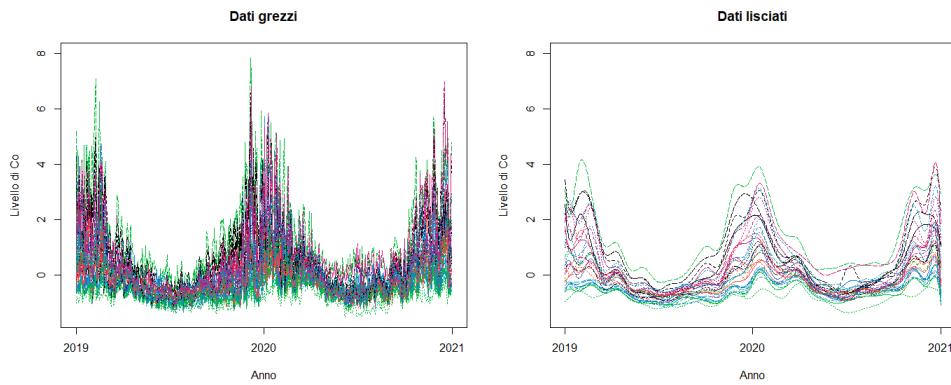


Figura 6: Variabile Livello di Co: dati grezzi (a sinistra) e dati lisciati (a destra).

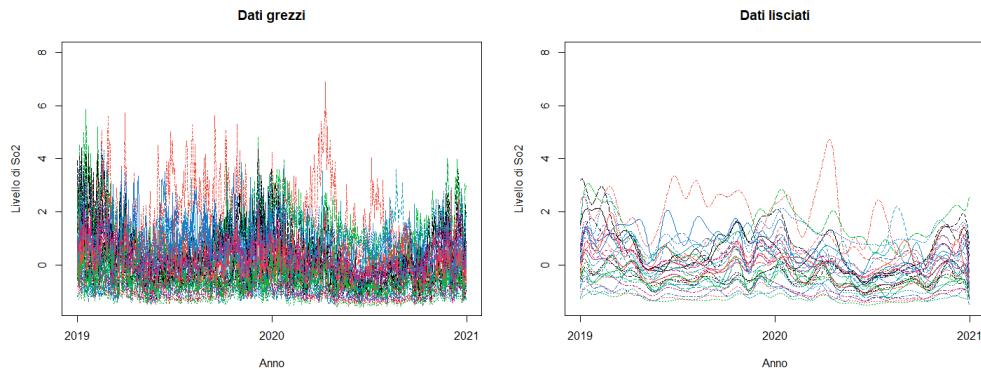


Figura 7: Variabile SO₂: dati grezzi (a sinistra) e dati lisciati (a destra).

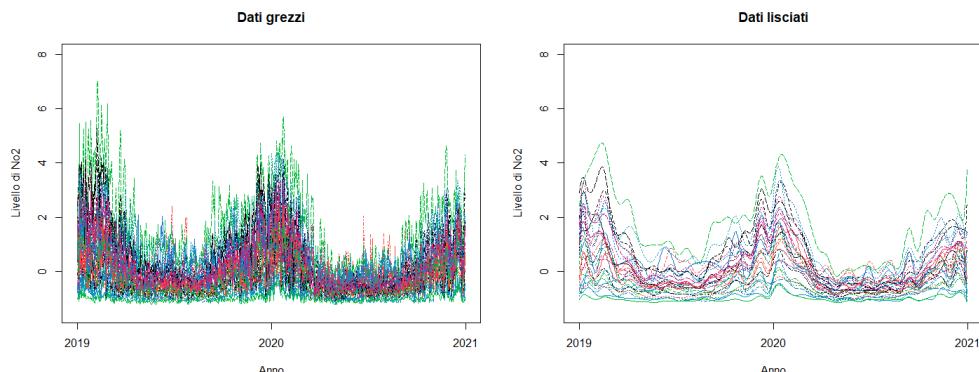


Figura 8: Variabile NO₂: dati grezzi (a sinistra) e dati lisciati (a destra).

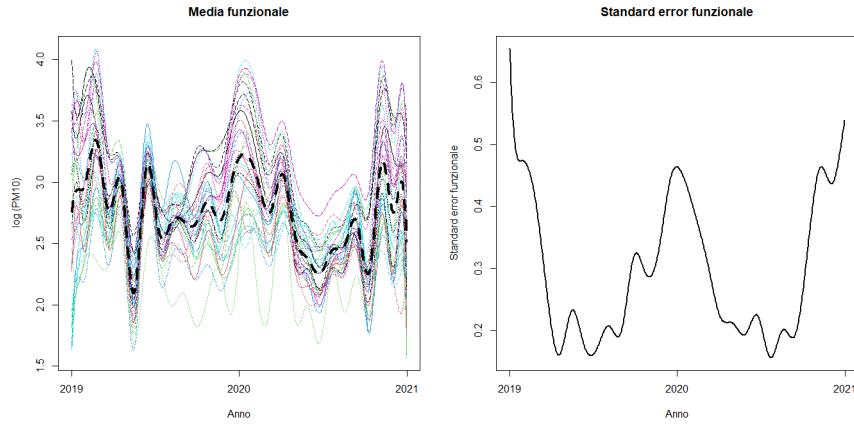


Figura 9: Media funzionale (a sinistra) e standard error funzionale (a destra) della variabile Livello di PM10.

di PM10. Per entrambe le quantità si nota un aumento nei mesi autunnali ed invernali mentre una riduzione nei mesi primaverili ed estivi. Inoltre è stato costruito un boxplot funzionale (Figura 10) basato sulla misura di profondità delle curve. Dal grafico è possibile notare come valgono analoghe considerazioni a quanto detto per lo standard error funzionale, per di più viene individuata una curva anomala di livello (e non di forma) che corrisponde alla città di Aosta.

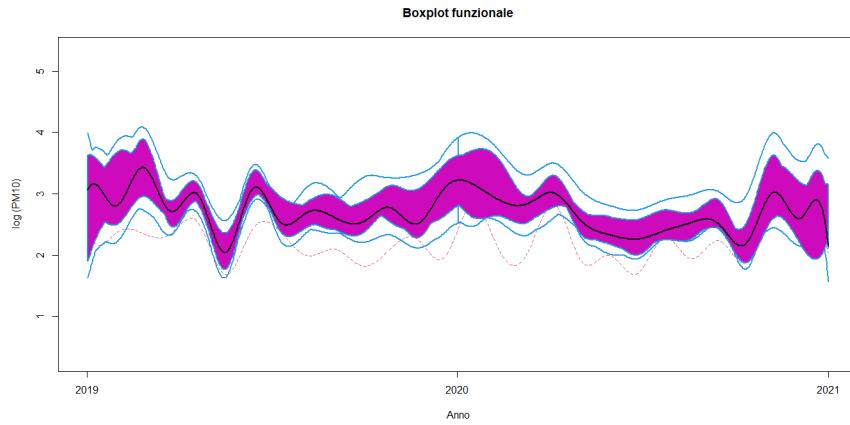


Figura 10: Boxplot funzionale del Livello di PM10; la linea rossa tratteggiata rappresenta la città di Aosta.

4 Valutazione del livello di PM10 per area geografica

Il primo obiettivo dell'analisi riguarda la valutazione del *Livello di PM10* a seconda dell'area geografica presa in considerazione. Prima di tutto si è creata la variabile categoriale *Area geografica* con 4 modalità (*Nord-Ovest*, *Nord-Est*, *Centro*, *Sud e Isole*) e ad ognuna delle 30 città si è assegnata la corrispondente modalità. La Figura 11 mostra che 6 città sono del Nord-Ovest, 5 città del Centro, 7 del Sud e 12 del Nord-Est.



Figura 11: Suddivisione delle città per area geografica

Dal punto di vista della modellazione statistica, si è preso in considerazione un modello di ANOVA funzionale

$$\log PM10_{ij}(t) = \mu(t) + area_j(t) + \varepsilon_{ij}(t), \quad i = 1, \dots, N_j, \quad j = 1, \dots, 4,$$

sotto il vincolo $\sum_{j=1}^4 area_j(t) = 0 \quad \forall t.$

Inizialmente si sono rappresentati i coefficienti funzionali tramite B-spline con $K = 15$ funzioni di base. Ottenendo da questa rappresentazione degli effetti stimati troppo frastagliati e quindi poco interpretabili, si sono penalizzati i coefficienti tramite un parametro di regolazione λ (comune per tutti) selezionato tramite convalida incrociata *leave-one-out* e pari a 173571, come mostrato in Figura 12. Ciò ha permesso di ottenere gli effetti stimati mostrati in Figura 13. Si nota come l'effetto del *Nord-Ovest* e del *Nord-Est* è praticamente identico: tende ad essere positivo nei mesi autunnali ed invernali e circa pari a 0 nei mesi primaverili ed estivi. Per quanto riguarda il *Sud e Isole* l'effetto stimato risulta della stessa intensità ma di segno opposto rispetto

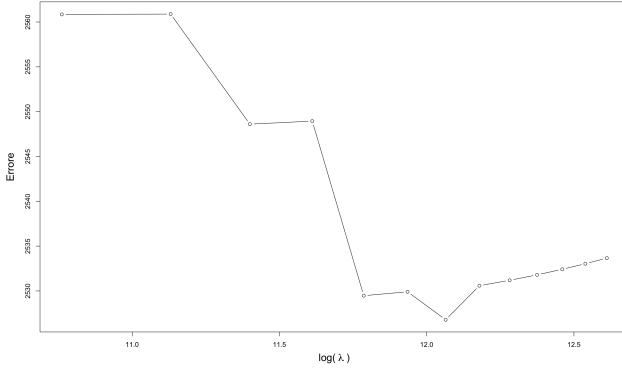


Figura 12: Errore in LOOCV per la scelta del valore del parametro di regolazione

all'effetto del *Nord-Ovest* e del *Nord-Est*, mentre l'effetto del *Centro* ha lo stesso segno ma con intensità minore rispetto all'effetto del *Sud* e *Isole* sul livello di PM10.

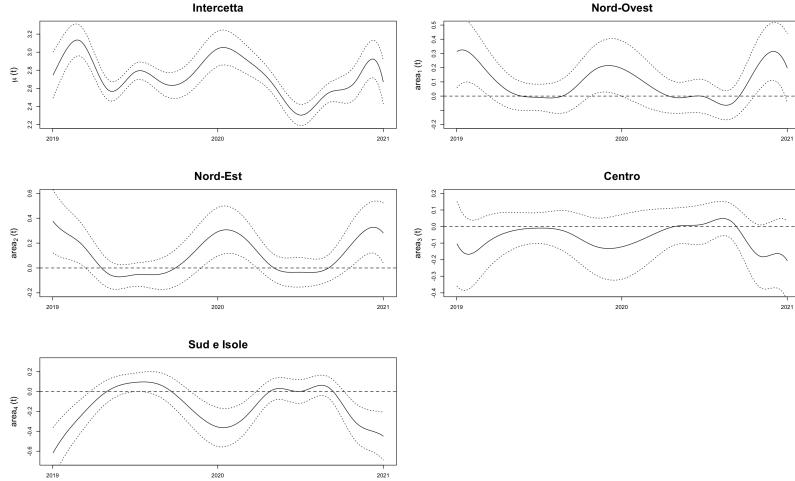


Figura 13: Effetti stimati e intervalli di confidenza punto a punto per area geografica

Dal punto di vista degli intervalli di confidenza punto a punto, l'effetto del *Centro* sembra non essere significativo sul livello di PM10, mentre l'effetto di *Nord-Ovest*, *Nord-Est* e *Sud e Isole* sembrerebbe essere significativo nei mesi freddi e non significativo nei mesi caldi.

Successivamente si è passati alla diagnostica del modello. In termini di bontà di adattamento ai dati, il grafico di sinistra della Figura 14 mostra l' R^2 funzionale, mettendo in luce come il modello riesca a spiegare sempre, ad esclusione degli ultimi giorni del 2020, almeno il 70% della variabilità totale. Dal punto di vista della significatività dell'effetto della variabile *Area*

geografica, si è effettuato un test F di permutazione per valutare il sistema di ipotesi

$$\begin{cases} H_0 : area_1(t) = area_2(t) = area_3(t) = area_4(t) = 0 \\ H_1 : \overline{H_0} \end{cases}$$

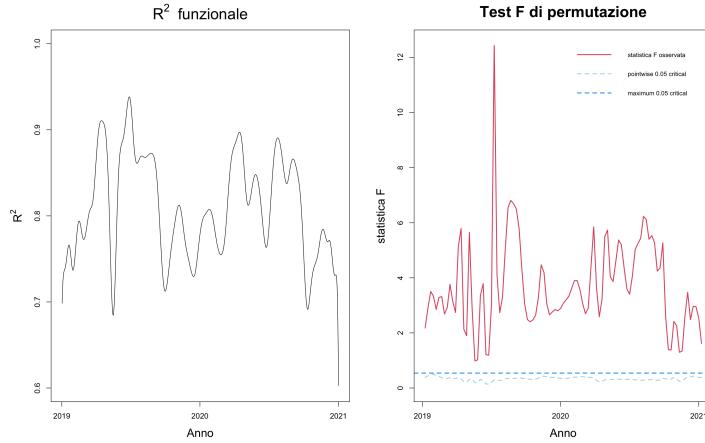


Figura 14: Diagnostica del modello ANOVA funzionale

Il grafico di destra della Figura 14 evidenzia che, ad un livello di significatività del 5%, si rifiuta l'ipotesi nulla sia puntualmente (ovvero ad ogni t) sia complessivamente (ovvero per ogni t), pertanto l'inclusione dell' *Area geografica* è utile per spiegare il comportamento del *Livello di PM10*.

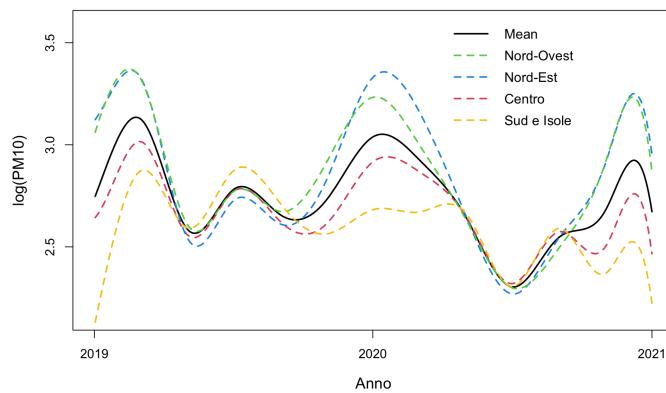


Figura 15: Effetti stimati nel modello ANOVA funzionale

In conclusione, è riportato in Figura 15 un grafico del *Livello di PM10* stimato per ogni area geografica, dove si conferma quanto detto in precedenza: nei mesi caldi il *Livello di PM10* è circa lo stesso in tutta Italia, mentre nei mesi freddi è più elevato rispetto alla media generale nel *Nord-Ovest* e nel *Nord-Est*, è inferiore rispetto alla media generale nel *Sud e Isole* ed è circa pari alla media generale al *Centro*.

5 Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti

Il secondo obiettivo dell'analisi è valutare se ci siano o meno delle relazioni tra il *Livello di PM10* e le altre variabili che sono state rilevate, ossia i tre fattori metereologici e gli altri tre fattori inquinanti.

5.1 Modello a effetti simultanei

Per rispondere al quesito d'interesse è stato innanzitutto formulato un modello funzione contro funzione a effetti simultanei. La specificazione del modello nella sua forma completa, ossia con tutte le possibili variabili esplicative incluse, è

$$\begin{aligned} \log PM10_i(t) = & \mu(t) + temp_i(t)\beta_{temp}(t) + wind_i(t)\beta_{wind}(t) + \\ & + prec_i(t)\beta_{prec}(t) + no2_i(t)\beta_{no2}(t) + \\ & + so2_i(t)\beta_{so2}(t) + co_i(t)\beta_{co}(t) + \varepsilon_i(t), \quad i = 1, \dots, 30. \end{aligned}$$

Per selezionare quali tra le variabili a disposizione includere nel modello finale è stata implementata una procedura passo-passo sia in avanti sia all'indietro, valutando ad ogni passo rispettivamente l'aggiunta o l'eliminazione di una variabile alla volta tramite il criterio d'informazione automatica AIC. In Figura 16 sono riportati i valori di AIC ottenuti da ogni modello

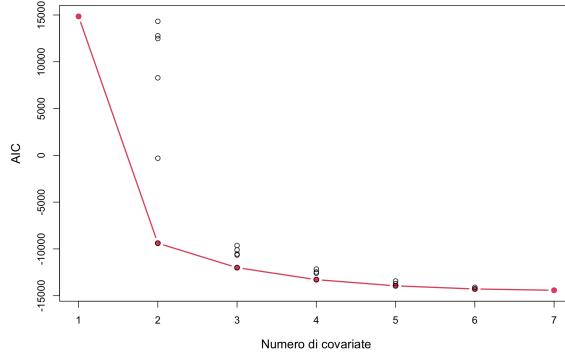


Figura 16: Valori di AIC ottenuti dai modelli a effetti simultanei adattati con la procedura passo-passo in avanti. In rosso sono indicati i valori di AIC corrispondenti al modello migliore di ogni passo.

stimato nella procedura passo-passo in avanti a partire dal modello nullo fino al modello completo. Il modello che minimizza l'AIC è il modello completo ($AIC=-14316$), che viene quindi selezionato come il modello migliore.

Una conclusione analoga si ottiene con la procedura all'indietro che, partendo dal modello completo, si conclude al primo passo. In Figura 17 sono riportate le stime dei coefficienti funzionali associati ad ognuna delle variabili incluse nel modello. Per quanto riguarda i fattori metereologici, si osserva che la temperatura ha un effetto positivo sul *Livello di PM10*, più marcato nei periodi freddi, la velocità del vento ha un effetto tendenzialmente negativo e più accentuato anch'esso nei periodi freddi, mentre le precipitazioni hanno un effetto tendenzialmente nullo, leggermente negativo in alcuni intervalli di tempo. I tre inquinanti hanno un'associazione positiva con il *Livello di PM10*, in particolare il monossido di carbonio, *Co*, il cui effetto è particolarmente accentuato nei periodi invernali, mentre l'effetto degli altri due inquinanti, *No2* e *So2*, è di un'entità minore e si vede soprattutto nei periodi estivi.

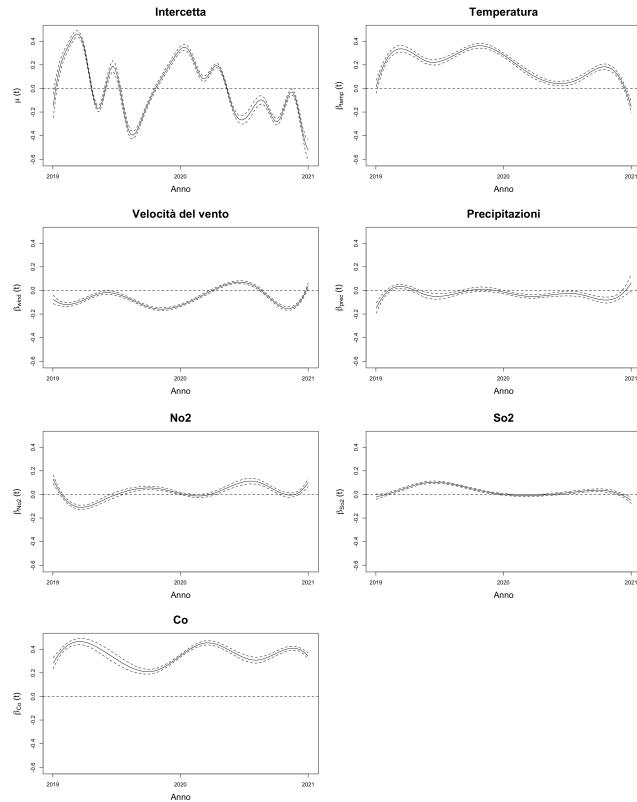


Figura 17: Effetti stimati nel modello simultaneo.

5.2 Modello a effetti bivariati

Il modello a effetti simultanei ha permesso di evidenziare importanti relazioni tra la variabile d'interesse e le concomitanti e, per sua costruzione, permette una facile interpretazione dei risultati che si ottengono. Tuttavia è limitato a valutare la relazione tra variabili misurate solamente nel medesimo istante e spesso può invece essere d'interesse indagare se ci siano degli effetti di alcune variabili sulla risposta in un istante di tempo differente. Per questo motivo è stato adattato ai dati in analisi un secondo modello, a effetti bivariati. In particolare, si è

considerato un modello in cui le variabili esplicative hanno un effetto sul *Livello di PM10* sia per le loro misurazioni in tempi s precedenti al tempo t considerato per la variabile risposta, sia per misurazioni in tempi s successivi a t . Questa scelta è giustificata due ragioni: in fase esplorativa si è osservato che i dati rilevati seguono una certa stagionalità nel tempo e questo *pattern* può rendere interessante l'informazione che viene data dalle variabili esplicative in periodi anche successivi a quello considerato per la variabile risposta (ad esempio, ci si può aspettare che, se la temperatura ha un effetto ritardato sui livelli di PM10, si possano utilizzare i dati sulla temperatura di Dicembre 2019 per dire qualcosa sui livelli di PM10 di Gennaio 2019); la seconda ragione riguarda il fatto che l'obiettivo d'interesse è puramente interpretativo e non previsivo, per questo motivo non è strettamente necessario il vincolo di utilizzare ai fini delle analisi esclusivamente le informazioni che si hanno fino all'istante temporale considerato.

Si specifica quindi un modello funzione contro funzione a effetti bivariati che nella sua forma di modello completo è del tipo

$$\begin{aligned} \log PM10_i(t) = & \mu(t) + \int temp_i(s)\beta_{temp}(s,t)ds + \\ & + \int wind_i(s)\beta_{wind}(s,t)ds + \int no2_i(s)\beta_{no2}(s,t)ds + \\ & + \int so2_i(s)\beta_{so2}(s,t)ds + \int co_i(s)\beta_{co}(s,t)ds + \\ & + \int prec_i(s)\beta_{prec}(s,t)ds + \varepsilon_i(t), \quad i = 1, \dots, 30. \end{aligned}$$

Anche in questo caso per selezionare il modello finale è stato implementata una procedura di selezione delle variabili passo-passo sia in avanti sia all'indietro basata sull'AIC. In Figura 18 sono riportati i valori di AIC ottenuti da ogni modello stimato nella procedura passo-passo in avanti e si nota che, come nel caso a effetti simultanei, il miglior modello risulta essere quello completo ($AIC=-13417$). Una conclusione analoga si ottiene con la procedura passo-passo all'indietro.

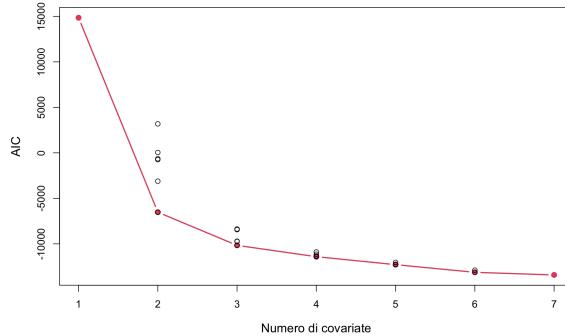


Figura 18: Valori di AIC ottenuti dai modelli a effetti bivariati adattati con la procedura passo-passo in avanti. In rosso sono indicati i valori di AIC corrispondenti al modello migliore di ogni passo.

In Figura 19 sono riportate le stime dei coefficienti funzionali associati ad ognuna delle variabili incluse nel modello. In questo caso, essendo i coefficienti funzioni di due variabili (a eccezione

dell'intercetta funzionale), le stime vengono riportate sottoforma di *heatmap* che presentano un colore chiaro (giallo) in corrispondenza di valori alti delle stime e un colore scuro (rosso) in corrispondenza di valori bassi. Sono inoltre sovraimposte delle curve di livello riportanti i valori stimati. L'interpretazione dei coefficienti è un po' più complicata rispetto al modello con effetti simultanei e come prima osservazione si nota che le stime sono generalmente piccole in modulo.

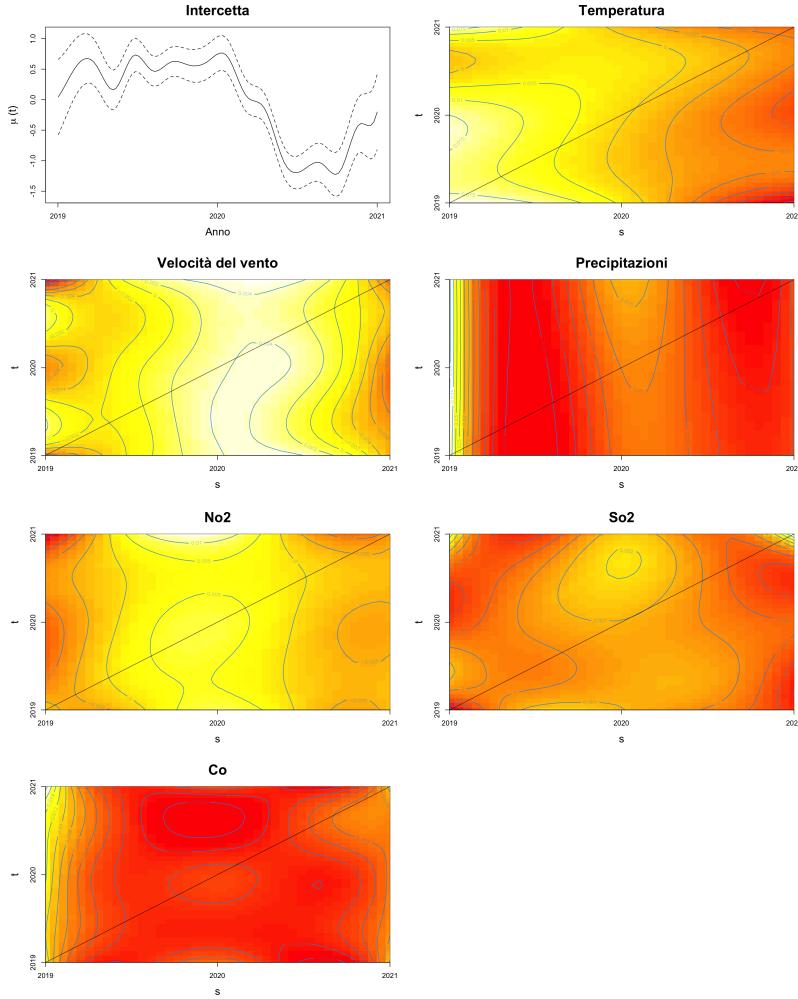


Figura 19: Effetti stimati nel modello a effetti bivariati. Le heatmap presentano un colore chiaro (giallo) in corrispondenza di valori alti delle stime e un colore scuro (rosso) in corrispondenza di valori bassi.

Per la temperatura si può dire sommariamente che se rilevata ad un tempo s precedente all'istante t considerato si ha un'associazione positiva con il *Livello di PM10* mentre ad un tempo successivo a quello considerato questa associazione è perlopiù negativa. Per la velocità del vento invece si nota principalmente che quella presente nel periodo invernale del 2020 ha un'associazione positiva sul *Livello di PM10* in tutti e due gli anni considerati. Le precipitazioni hanno

un effetto quasi sempre negativo sul *Livello di PM10* misurati in tutto il biennio considerato, a eccezione di quelle rilevate nei periodi invernali 2019-2020 che hanno al contrario un effetto debolmente positivo. Per quanto riguarda le stime dei coefficienti relative ai fattori inquinanti, sembra che l'effetto sia totalmente catturato dalla variabile *No2*, di cui si nota in particolare l'associazione positiva che hanno i suoi valori dell'inverno 2020 sul *Livello di PM10* di tutto l'arco temporale considerato; gli altri due inquinanti hanno invece un effetto tendenzialmente nullo o poco interpretabile.

Dati i risultati difficilmente interpretabili e un valore dell'AIC più alto del miglior modello a effetti bivariati, ci si sente di considerare più affidabili le conclusioni che sono state fatte con il modello ad effetti simultanei.

6 Valutazione dell'effetto del lockdown sul livello di PM10

Il terzo e ultimo obiettivo dell'analisi riguarda la valutazione di un eventuale effetto del lockdown sul *Livello di PM10*. Per rispondere al quesito, il primo passo effettuato è stato dividere, per ognuna delle 30 città, la funzione lisciata di ciascuna variabile in due funzioni: una per il 2019 e una per il 2020. In questo modo si è passati da 30 a 60 unità statistiche. Inoltre è stata creata la variabile esplicativa scalare *Anno*, che assume valore 0 nel 2019 e 1 nel 2020. In Figura 20 sono riportati il *Livello di PM10* nelle 30 città suddiviso per anno.

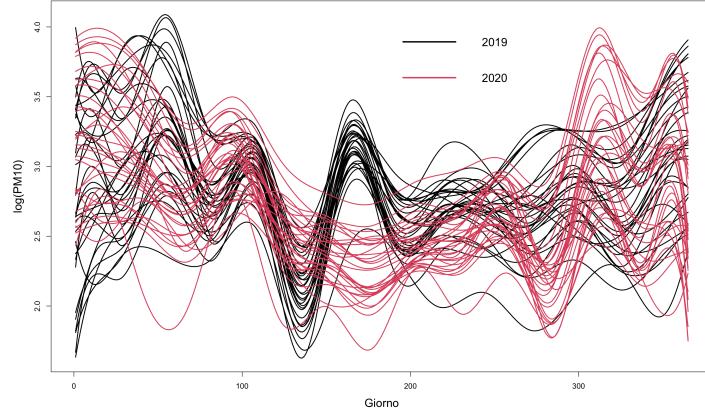


Figura 20: *Livello di PM10* nelle 30 città: in nero nel 2019 e in rosso nel 2020

6.1 Modello a effetti simultanei

Il primo modello preso in considerazione è stato il modello simultaneo con l'inclusione delle tre variabili metereologiche (*Temperatura*, *Velocità del vento*, *Precipitazioni*) e della variabile *Anno*

$$\begin{aligned} \log PM10_i(t) = & \mu(t) + anno_i \beta_{anno}(t) + temp_i(t) \beta_{temp}(t) + \\ & + wind_i(t) \beta_{wind}(t) + prec_i(t) \beta_{prec}(t) + \varepsilon_i(t), \quad i = 1, \dots, 60. \end{aligned}$$

Nel modello sono state incluse le variabili metereologiche come variabili di controllo, mentre non sono state prese in considerazione le variabili relative al *Livello di NO₂*, *Livello di SO₂*, *Livello di Co* in quanto è ragionevole ipotizzare che anche su queste tre variabili ci sia stato un effetto del lockdown, pertanto una loro inclusione nel modello avrebbe rischiato di mascherare un eventuale effetto del lockdown sul *Livello di PM10*.

In Figura 21 sono riportati i coefficienti funzionali stimati dal modello: si nota che le bande di confidenza risultano molto strette per tutti gli effetti. Da un punto di vista interpretativo, focalizzando l'attenzione sulla variabile di interesse *Anno*, è possibile constatare che l'effetto risulta essere monotono decrescente per tutta la durata del lockdown, con un effetto positivo nella prima metà e negativo nella seconda metà. Sembra quindi esserci un effetto a livello di intensità per tutta la durata del lockdown, mentre una diminuzione del livello di PM10 a partire dalla seconda metà del lockdown.

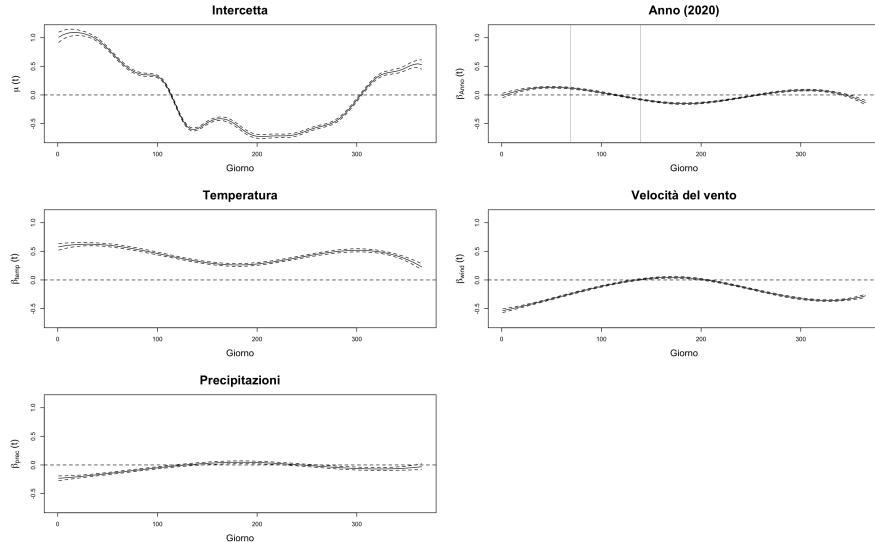


Figura 21: Effetti stimati dal modello simultaneo

Queste considerazioni riguardano gli effetti e le bande di confidenza stimate punto a punto, tuttavia per poter dire che tali considerazioni risultano valide bisognerebbe valutare la significatività della variabile *Anno* complessivamente. A tal fine, si è effettuato un t-test di permutazione basato sulla procedura di Freedman-Lane⁶ per saggiare l'ipotesi

$$\begin{cases} H_0 : \beta_{anno}(t) = 0 & \forall t \\ H_1 : \overline{H_0} \end{cases}$$

Il grafico in Figura 22 mostra come, ad un livello di significatività del 5%, si rifiuta l'ipotesi nulla, pertanto l'effetto della variabile *Anno* risulta significativamente diverso da 0 e le valutazioni fatte possono essere ritenute valide.

⁶ Freedman D., Lane D. (1983), *A nonstochastic interpretation of reported significance levels*.

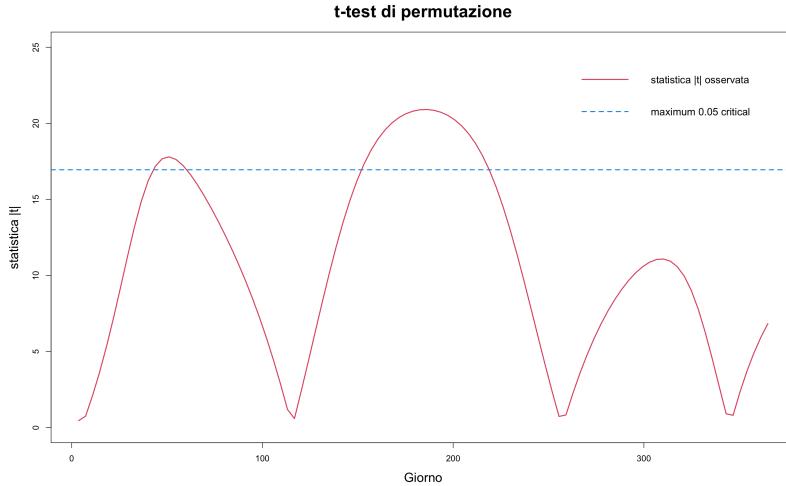


Figura 22: *t-test di permutazione basato sulla procedura di Freedman-Lane*

6.2 Modello simultaneo con intercetta casuale

Il modello utilizzato precedentemente ha permesso di verificare la significatività di una differenza del *Livello di PM10* tra il 2019 e il 2020. Tuttavia questo modello non tiene conto della dipendenza che vi è tra le coppie di rilevazioni effettuate per la stessa città nei due anni (2019 e 2020); per questo motivo si è deciso di considerare un modello con intercetta casuale funzionale specifica per ciascuna città.

Di seguito si riporta la specificazione del modello simultaneo con intercetta casuale funzionale

$$\begin{aligned} \log PM10_i(t) = & \mu(t) + anno_i \beta_{anno}(t) + città_j[i](t) + temp_i(t) \beta_{temp}(t) + \\ & + wind_i(t) \beta_{wind}(t) + prec_i(t) \beta_{prec}(t) + \varepsilon_i(t), \quad i = 1, \dots, 60, \quad j = 1, \dots, 30, \end{aligned}$$

con generica intercetta casuale rappresentata tramite espansione in funzioni di base del tipo $città(t) = \mathbf{b}'\Phi(t)$, dove per il vettore di coefficienti \mathbf{b} si assume una distribuzione a priori $\mathbf{b} \sim N(0, (\lambda R)^{-})$ con $R = \int [L\Phi(t)]' [L\Phi(t)] dt$ matrice di penalizzazione.

In Figura 23 vengono riportati i coefficienti stimati dal modello ed è possibile osservare come questi non differiscano sostanzialmente nei livelli e nell'andamento dagli effetti stimati con il modello senza intercetta casuale, eccetto che per la variabile *Temperatura*. La differenza importante da notare è legata alle bande di confidenza calcolate, poiché in questo caso risultano più ampie.

In Figura 24 si riportano le previsioni delle intercette casuali funzionali e i residui del modello. Si osserva che vi è eterogeneità tra le diverse città, infatti le previsioni delle intercette casuali presentano ampia variabilità confrontandola con quella dei residui. Questa eterogeneità nel modello simultaneo con soli effetti fissi non veniva catturata. Inoltre guardando solo il grafico dei residui si nota come, particolarmente nei mesi primaverili, vi sia una forte correlazione, questo indica che le covariate utilizzate nel modello non permettono di spiegare del tutto il *Livello di PM10* per questi mesi.

Andrebbe individuata una procedura per valutare la ragionevolezza dell'inserimento di un'intercetta casuale: in termini di variabilità spiegata si nota che questa aumenta considerevolmente nella maggior parte dell'arco temporale considerato (Figura 25), tuttavia sarebbe necessario un

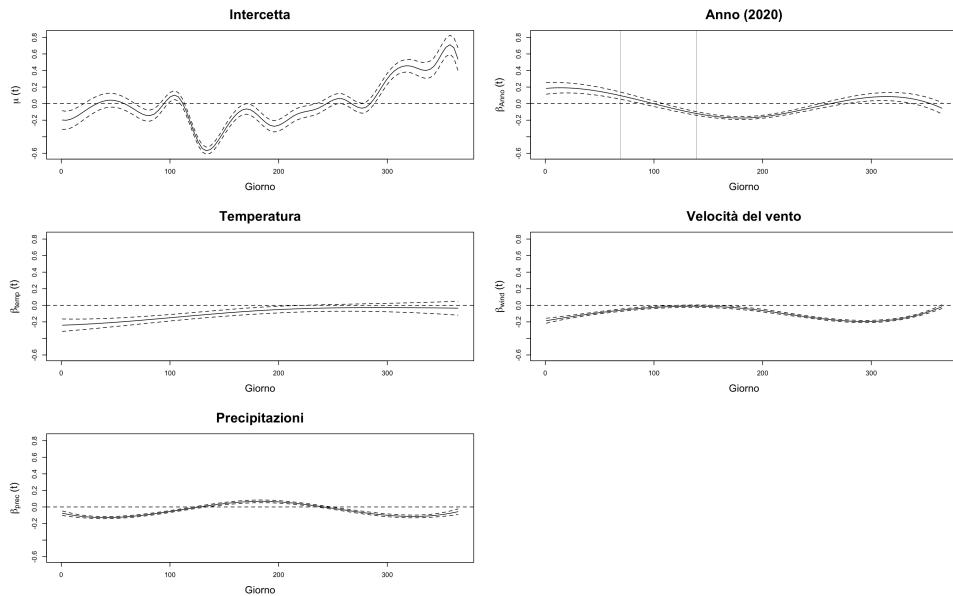


Figura 23: Effetti stimati dal modello simultaneo con intercette casuali.

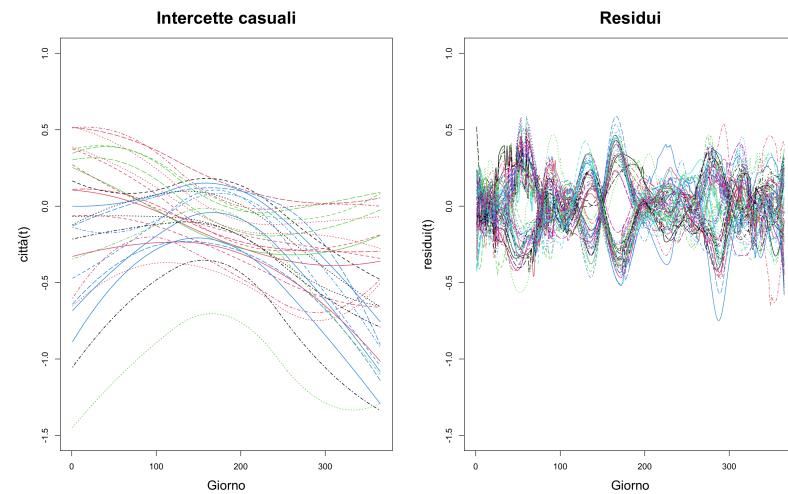


Figura 24: Previsioni delle intercette casuali (a sinistra) e residui del modello simultaneo con intercetta casuale (a destra).

metodo che consideri anche il costo dell'inserimento del termine casuale nel modello. Anche per quanto riguarda la variabile *Anno*, sarebbe opportuno utilizzare un test adeguato per valutarne la significatività in un modello a effetti misti.

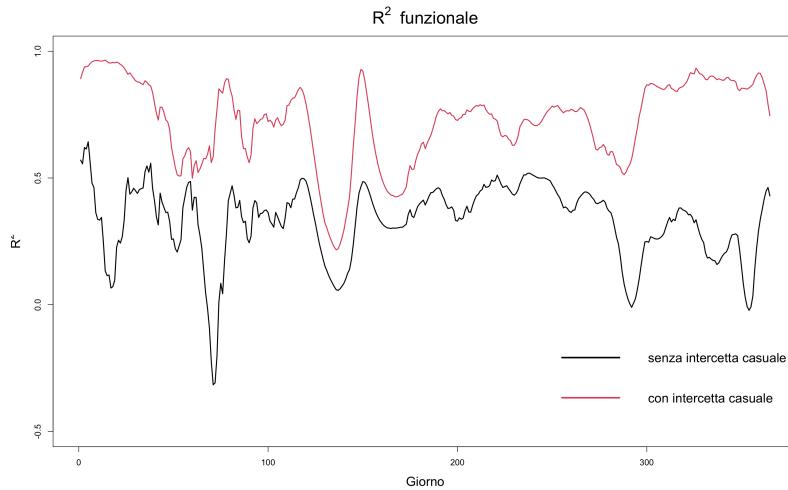


Figura 25: R^2 funzionale per il modello simultaneo con intercetta casuale (in rosso) e senza intercetta casuale (in nero).

7 Conclusioni

In questo lavoro sono stati adattati diversi modelli che hanno permesso di dire in conclusione che esistono differenze nel *Livello di PM10* tra aree geografiche d'Italia, che ci sono delle relazioni tra *Livello di PM10* e alcuni fattori metereologici e inquinanti e infine che sembra esserci una differenza nel *Livello di PM10* tra il 2020 e il 2019.

Alcuni possibili miglioramenti possono essere i seguenti: l'analisi relativa alle differenze tra aree geografiche in Italia potrebbe essere integrata alle due analisi che sono state svolte in seguito; il modo in cui è stata trattata la variabile *Precipitazioni totali* costituisce una delle varie possibilità, la questione meriterebbe un approfondimento ulteriore per individuarne una eventualmente migliore; l'inserimento degli effetti nei modelli simultanei potrebbe essere ritardato per descrivere meglio le relazioni tra esplicative e risposta; nel modello a effetti bivariati si potrebbero considerare gli effetti delle esplicative solo in tempi precedenti a quello considerato per la variabile risposta ($s \leq t$); si potrebbe affinare l'analisi relativa all'effetto del lockdown sul *Livello di PM10*, anche alla luce delle considerazioni fatte alla Sezione 6.2.