

Analisi del PM10 in Italia

Cugnigni Daniele, Fanesi Alessandro, Genesin Leonardo

Metodi statistici per dati funzionali
a.a. 2022-2023



Obiettivi dell'analisi

- 1 Valutazione del livello di PM10 per area geografica
- 2 Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti
- 3 Valutazione dell'effetto del lockdown sul livello di PM10



Dataset

- Fonti:
 - ▶ Atmosphere Data Store (Copernicus)¹
 - ▶ Climate Data Store (Copernicus)²
- Unità Statistiche: 30 città Italiane (20 capoluoghi di Regione e 10 città popolose).

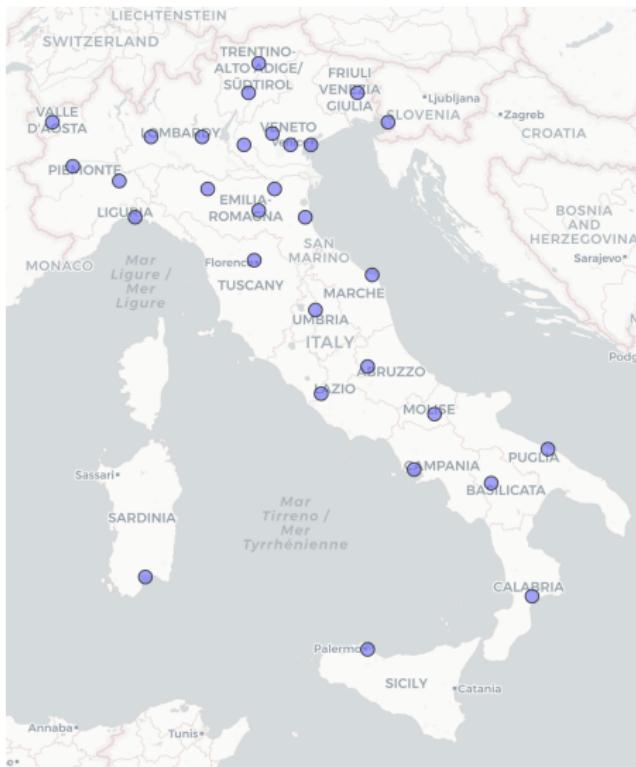
¹<https://ads.atmosphere.copernicus.eu>

²<https://cds.climate.copernicus.eu>



Dataset

Città rilevate



Dataset

- Sono state considerate le 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 ($\mu\text{g}/\text{m}^3$)*
- Tempo di rilevazione: dal 1 Gennaio 2019 al 31 Dicembre 2020 con frequenza oraria.



Preprocessing

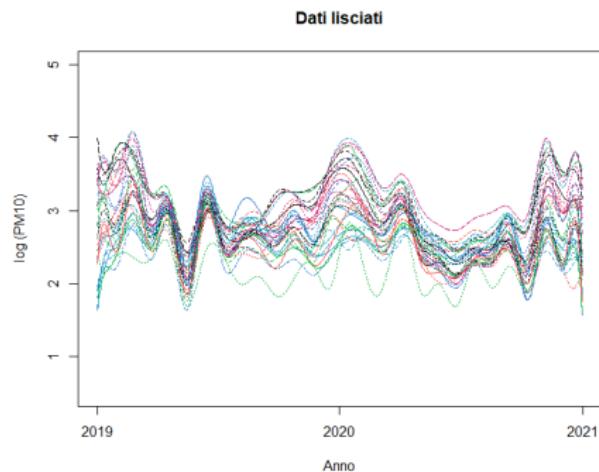
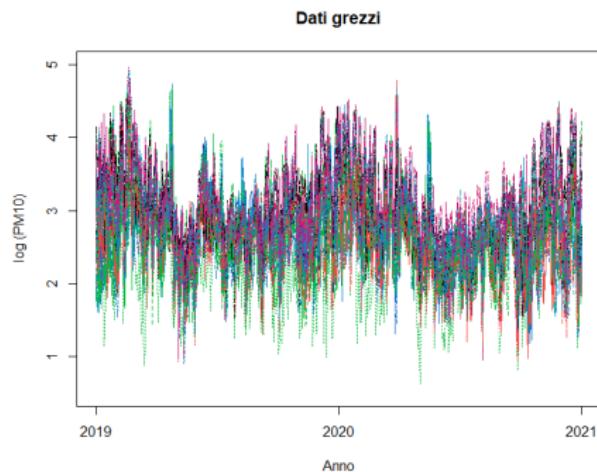
- Aggregazione dei dati orari in dati giornalieri
- Trasformazione logaritmica della variabile risposta (*Livello di PM10*)
- Standardizzazione delle variabili esplicative (*Precipitazioni totali esclusa*)
- Variabile *Precipitazioni totali*:
 - ▶ Dicotomizzazione (Piove = 1, Non piove = 0)
 - ▶ Regolarizzazione dell'andamento



Lisciamento - Variabile risposta

Livello di PM10

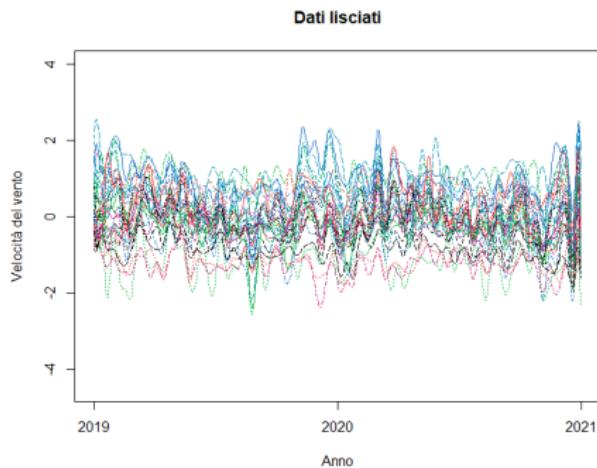
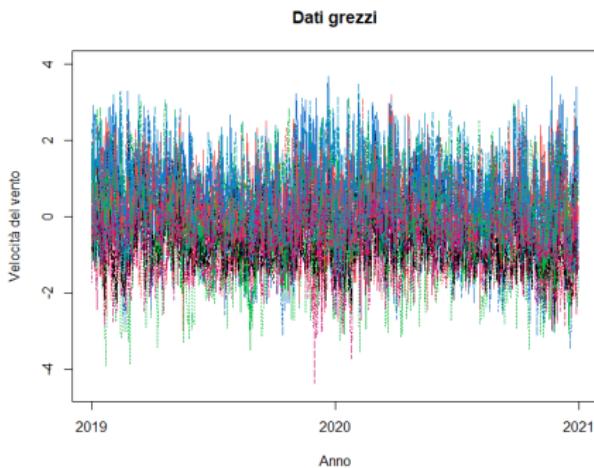
- Rappresentazione tramite *B-spline* con $K = 30$ funzioni di base



Lisciamento - Variabili metereologiche

Velocità del vento

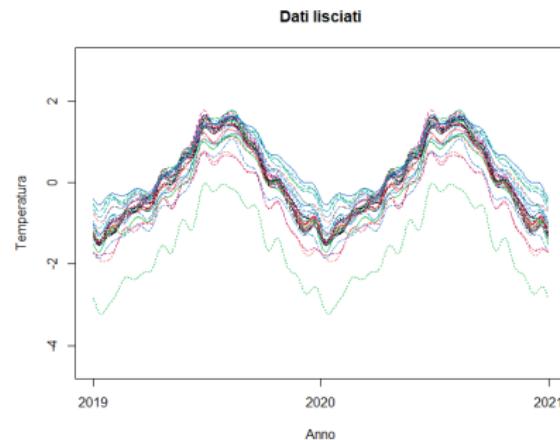
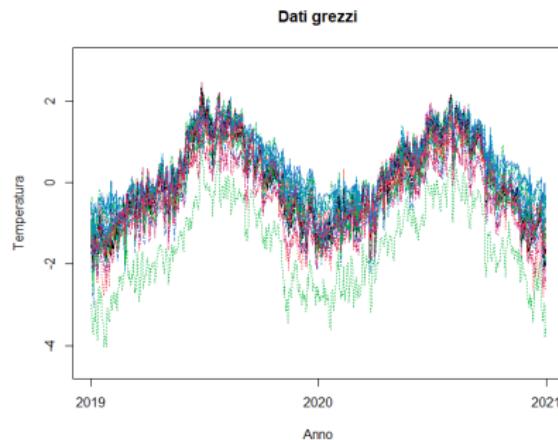
- Rappresentazione tramite *B-spline* con $K = 80$ funzioni di base



Lisciamento - Variabili metereologiche

Temperatura

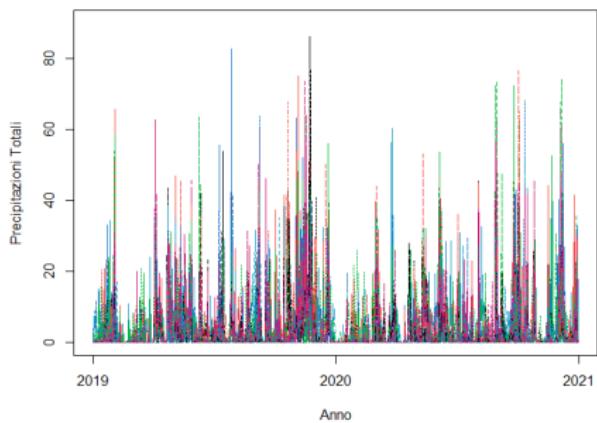
- Rappresentazione tramite *Basi di Fourier* con $K = 25$ funzioni di base e periodo $T = 365$



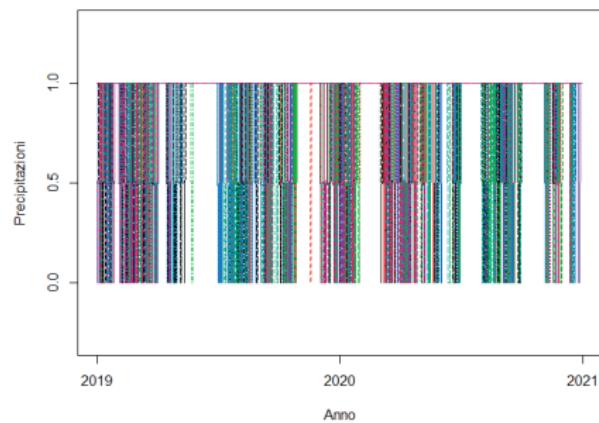
Lisciamento - Variabili metereologiche

Precipitazioni totali

Dati grezzi



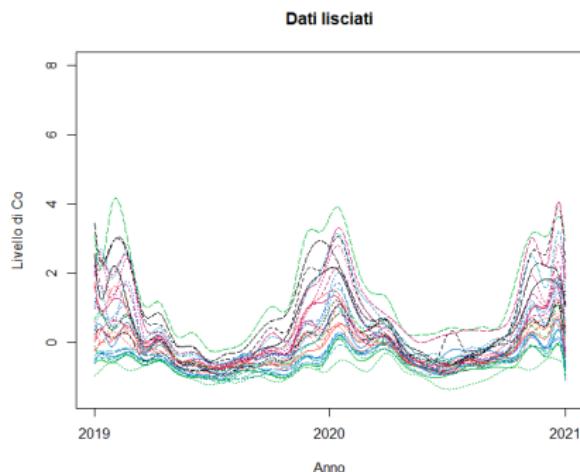
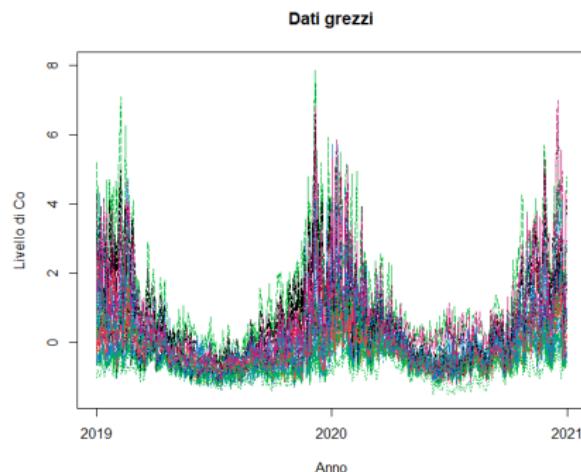
Dati Trasformati



Lisciamento - Altri inquinanti

Livello di Co

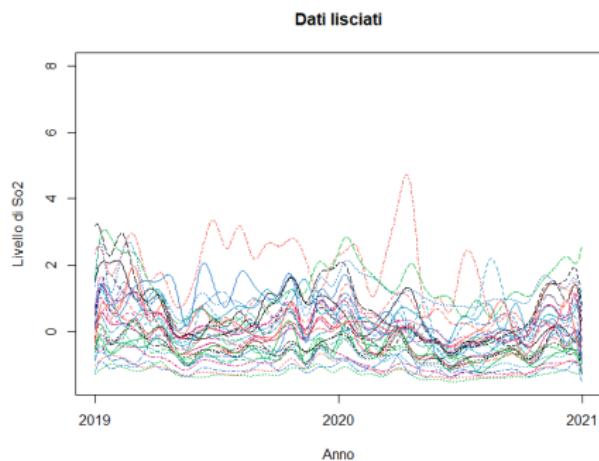
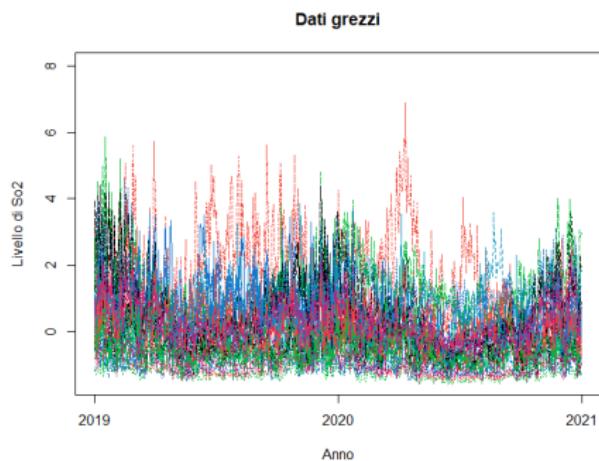
- Rappresentazione tramite *B-spline* con $K = 32$ funzioni di base



Lisciamento - Altri inquinanti

Livello di So₂

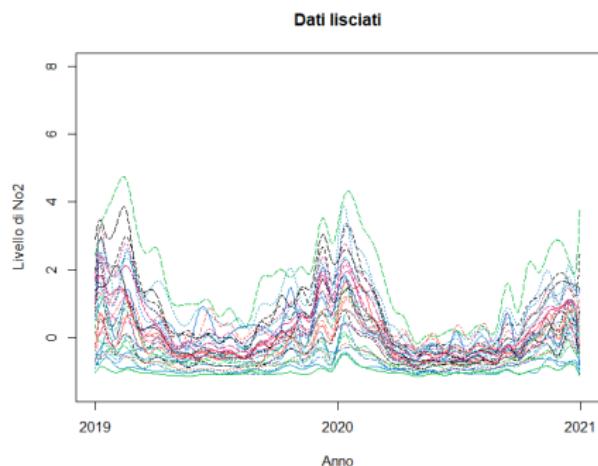
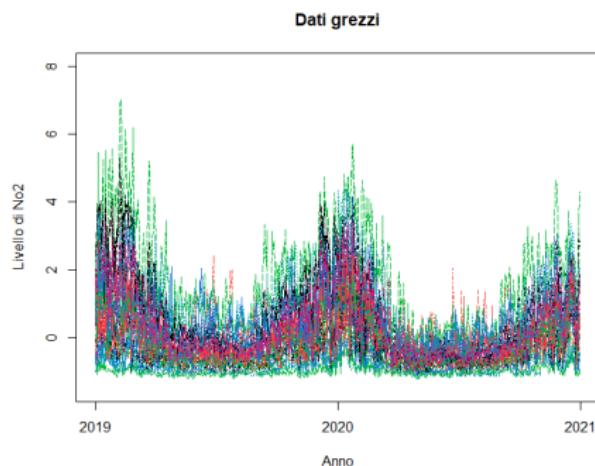
- Rappresentazione tramite *B-spline* con $K = 40$ funzioni di base



Lisciamento - Altri inquinanti

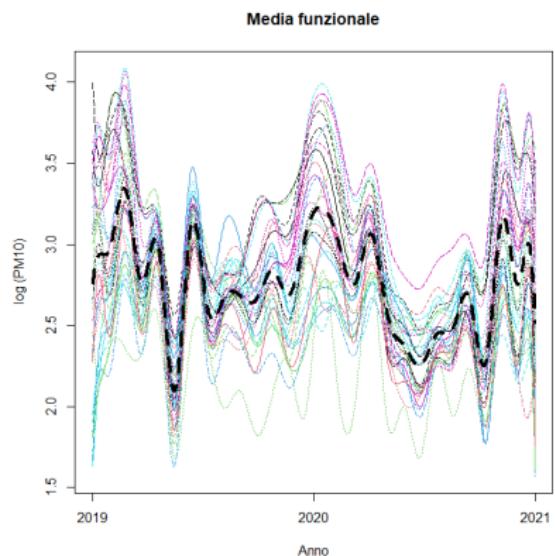
Livello di No₂

- Rappresentazione tramite *B-spline* con $K = 50$ funzioni di base



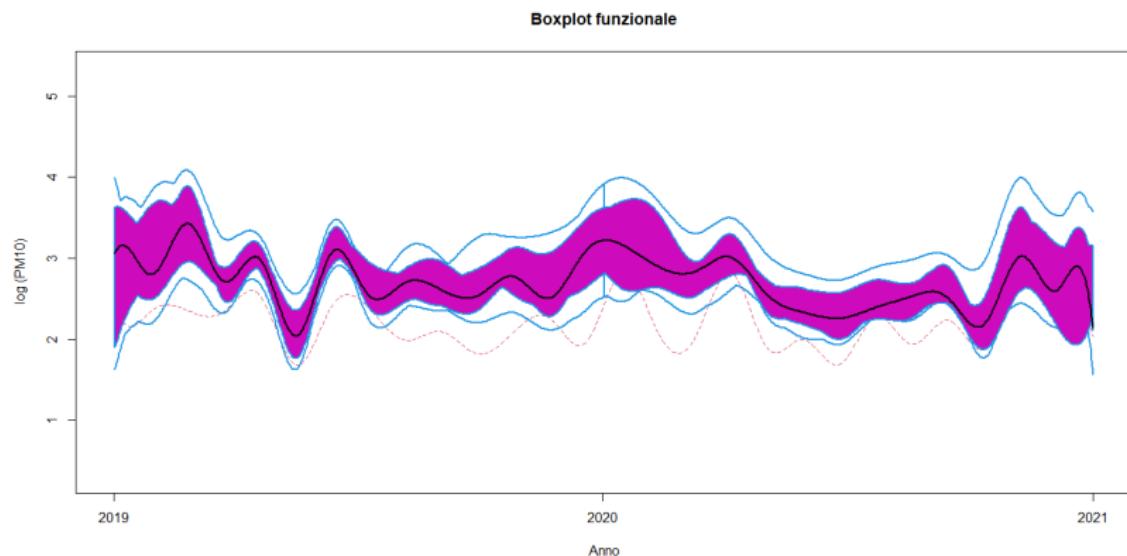
Analisi Esplorativa - Variabile risposta

Livello di PM10 - Media e Standard Error funzionale



Analisi Esplorativa - Variabile risposta

Livello di PM10 - Boxplot funzionale



Obiettivi dell'analisi

- 1 Valutazione del livello di PM10 per area geografica
- 2 Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti
- 3 Valutazione dell'effetto del lockdown sul livello di PM10



Valutazione del livello di PM10 per area geografica

Variabile *area*



Valutazione del livello di PM10 per area geografica

ANOVA funzionale - Specificazione del modello

- Si prende in considerazione un modello di ANOVA funzionale:

$$\log PM10_{ij}(t) = \mu(t) + area_j(t) + \varepsilon_{ij}(t), \quad i = 1, \dots, 30, \quad j = 1, \dots, 4,$$

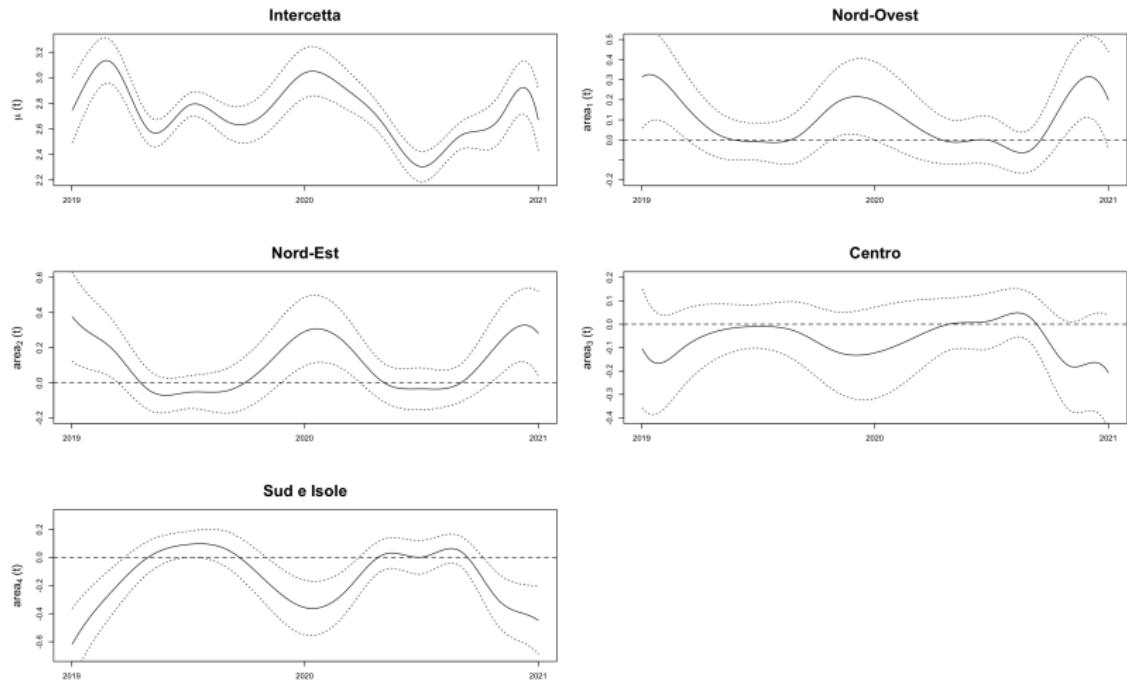
sotto il vincolo $\sum_{j=1}^4 area_j(t) = 0$

- I coefficienti funzionali vengono rappresentati tramite B-spline con $K = 15$ funzioni di base;
- I coefficienti sono penalizzati e il parametro di regolazione λ viene selezionato tramite convalida incrociata *leave-one-out*.



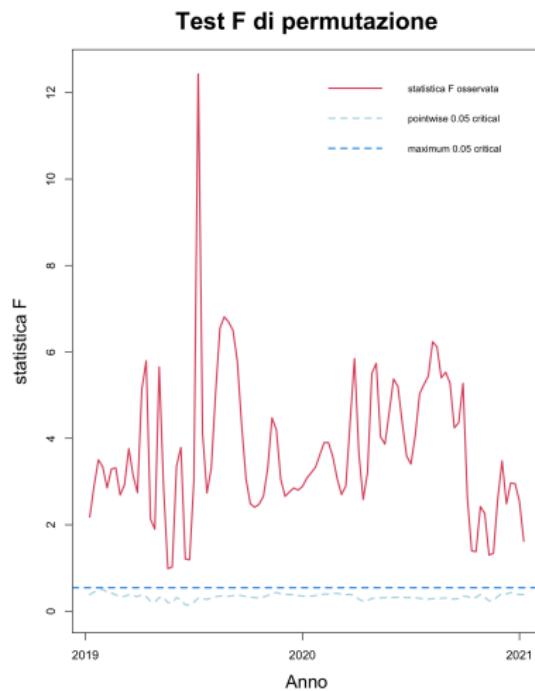
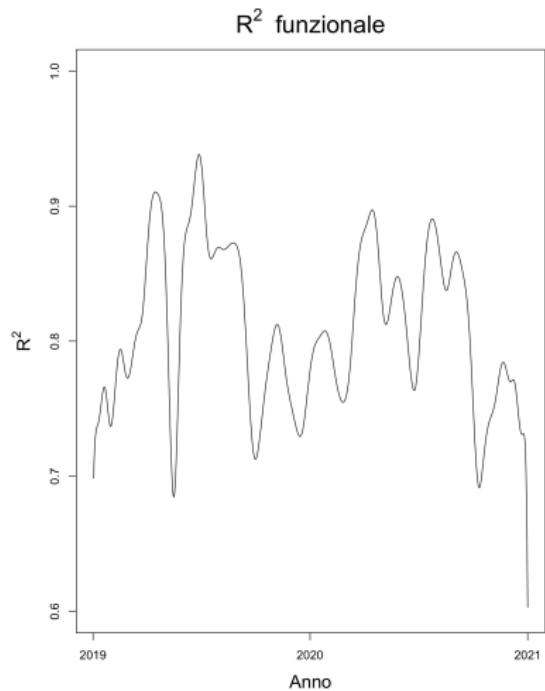
Valutazione del livello di PM10 per area geografica

ANOVA funzionale - Effetti stimati



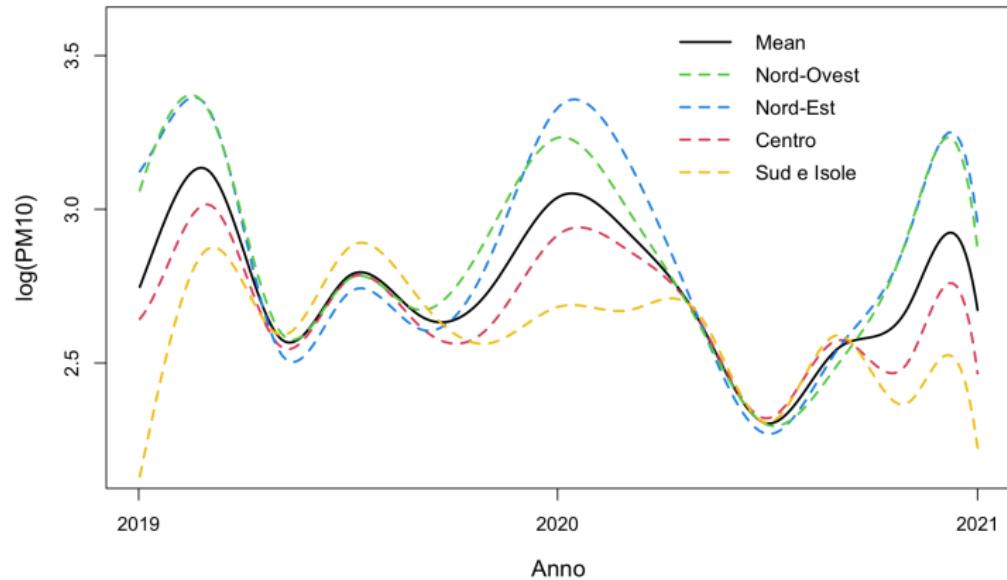
Valutazione del livello di PM10 per area geografica

ANOVA funzionale - Diagnostica



Valutazione del livello di PM10 per area geografica

ANOVA funzionale - Interpretazione



Obiettivi dell'analisi

- 1 Valutazione del livello di PM10 per area geografica
- 2 Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti
- 3 Valutazione dell'effetto del lockdown sul livello di PM10



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

Modello con effetti simultanei - Specificazione del modello completo

- Modello con effetti simultanei

$$\begin{aligned}\log PM10_i(t) = & \mu_i(t) + temp_i(t)\beta_{temp}(t) + wind_i(t)\beta_{wind}(t) + \\ & + prec_i(t)\beta_{prec}(t) + no2_i(t)\beta_{no2}(t) + \\ & + so2_i(t)\beta_{so2}(t) + co_i(t)\beta_{co}(t) + \varepsilon_i(t);\end{aligned}$$



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

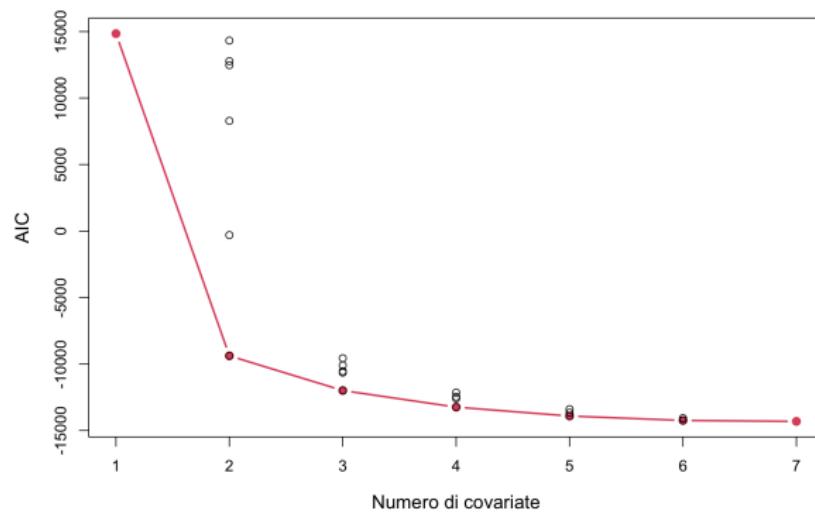
Selezione delle variabili

- Passo passo in avanti:
 - ▶ Criterio di selezione: AIC,
 - ▶ Si parte dal modello nullo,
 - ▶ Ad ogni passo viene fissato il parametro di regolazione delle variabili già inserite nel modello;
- Passo passo all'indietro:
 - ▶ Criterio di selezione: AIC,
 - ▶ Si parte dal modello completo,
 - ▶ Il parametro di regolazione delle variabili rimane fissato per tutta la durata della procedura;



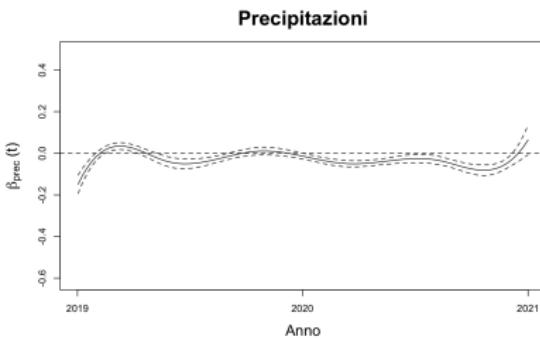
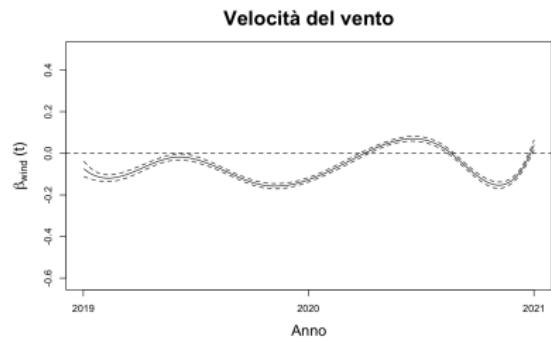
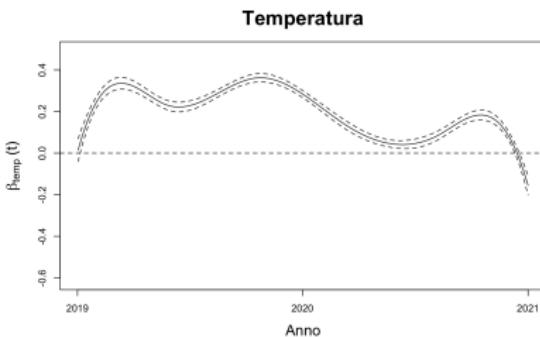
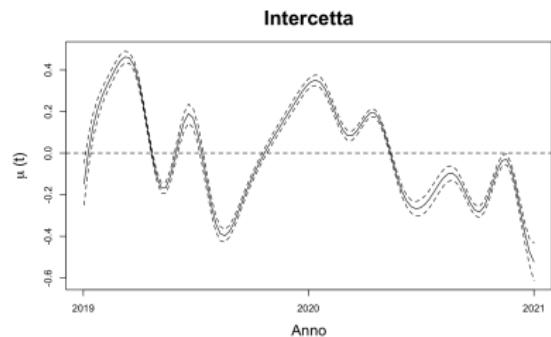
Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti

Modello con effetti simultanei - risultati stepAIC forward



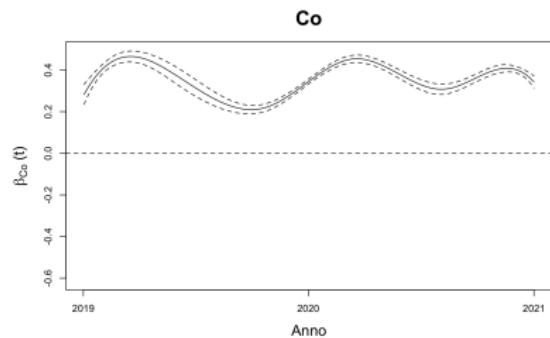
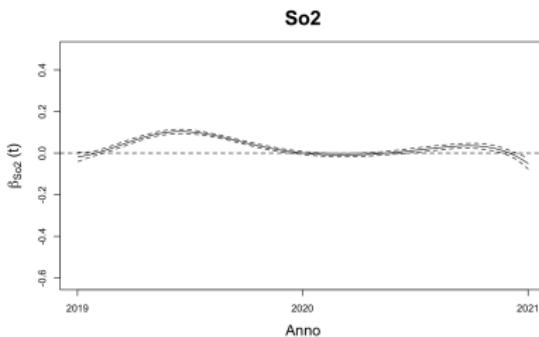
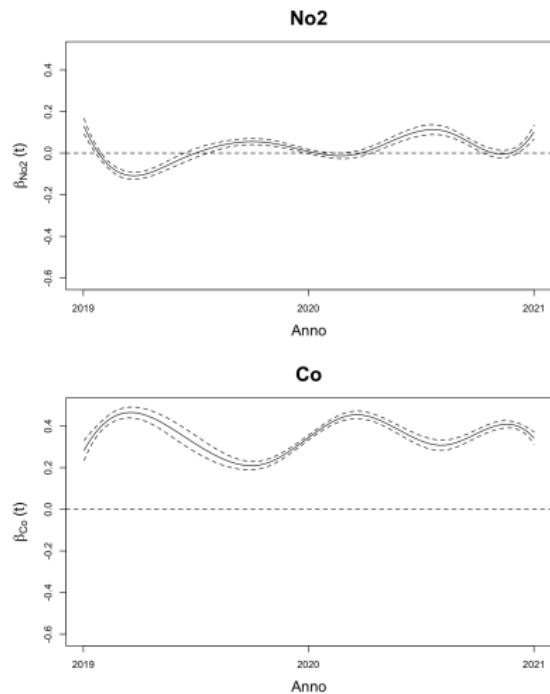
Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti

Modello con effetti simultanei - Effetti stimati



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

Modello con effetti simultanei - Effetti stimati



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

Modello con effetti bivariati - Specificazione del modello completo

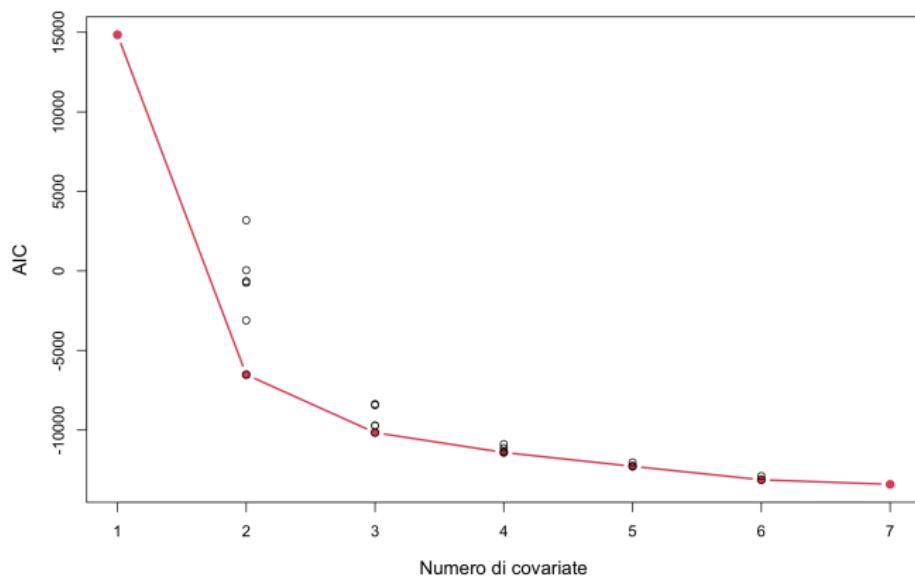
- Modello con effetti bivariati

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



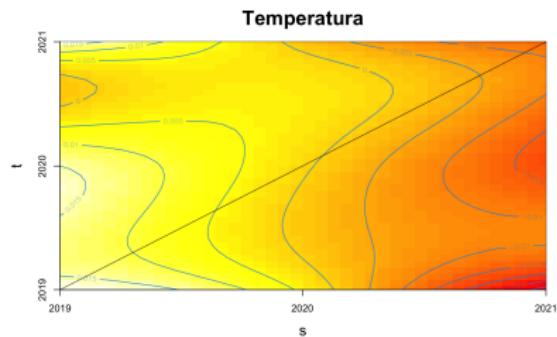
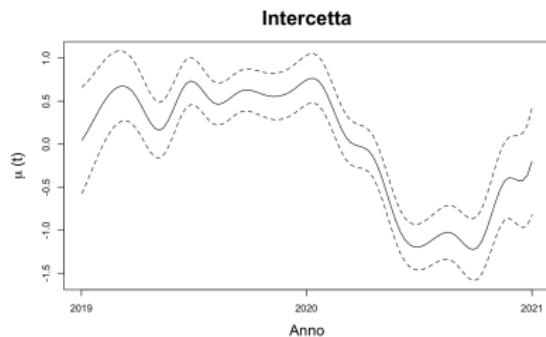
Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti

Modello con effetti bivariati - risultati stepAIC forward



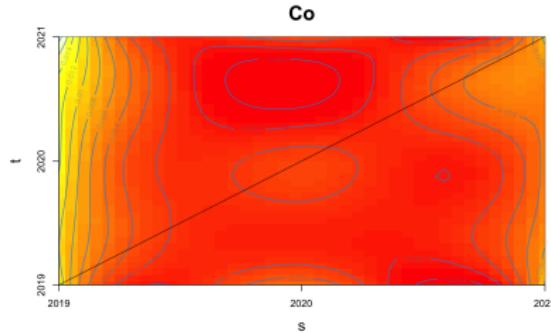
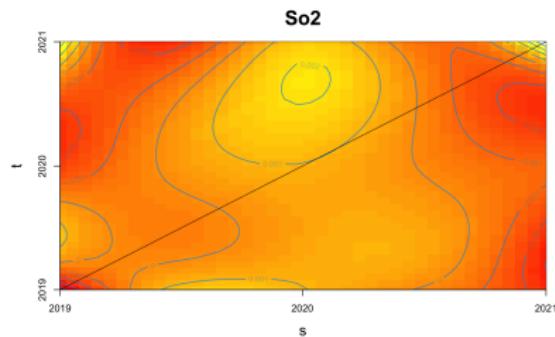
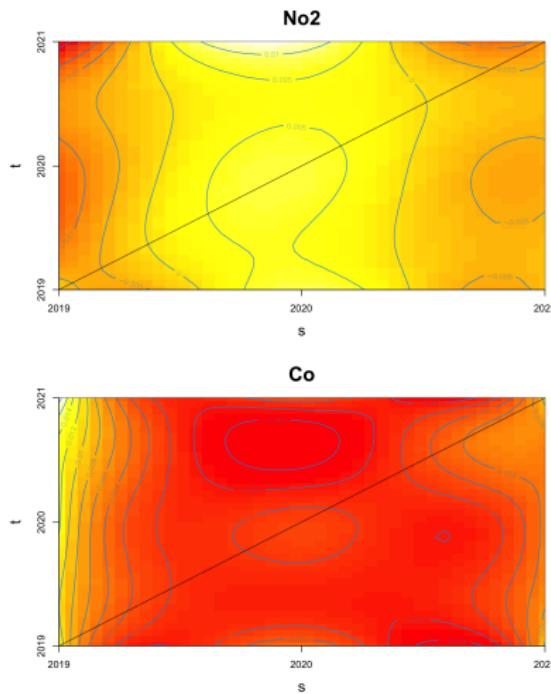
Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti

Modello con effetti simultanei - Effetti stimati



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

Modello con effetti simultanei - Effetti stimati



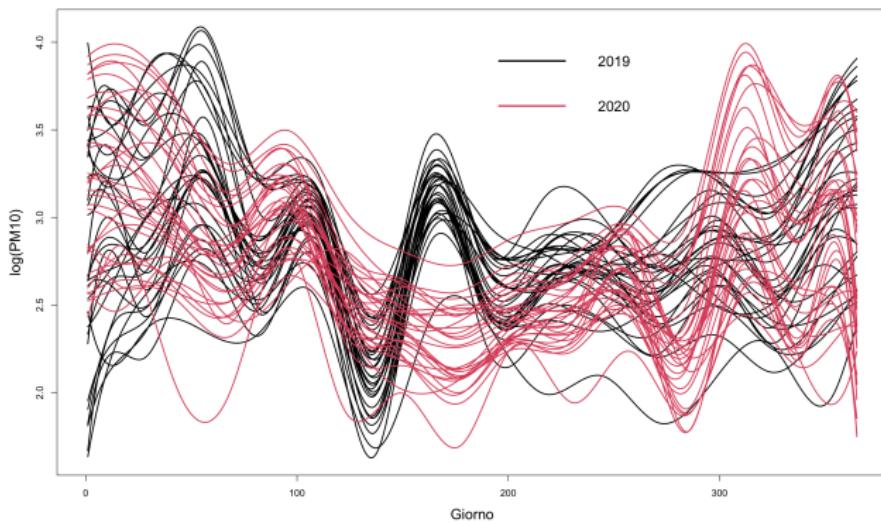
Obiettivi dell'analisi

- 1 Valutazione del livello di PM10 per area geografica
- 2 Valutazione della relazione tra il livello di PM10, alcuni fattori metereologici e altri inquinanti
- 3 Valutazione dell'effetto del lockdown sul livello di PM10



Valutazione dell'effetto del lockdown sul livello di PM10

- Per ognuna delle 30 città, si divide la funzione di ciascuna variabile in due: una funzione per il 2019 e una per il 2020 ($N = 60$)
- Si crea la variabile esplicativa scalare *Anno* (2019 = 0, 2020 = 1)



Valutazione dell'effetto del lockdown sul livello di PM10

Modello simultaneo - Specificazione del modello

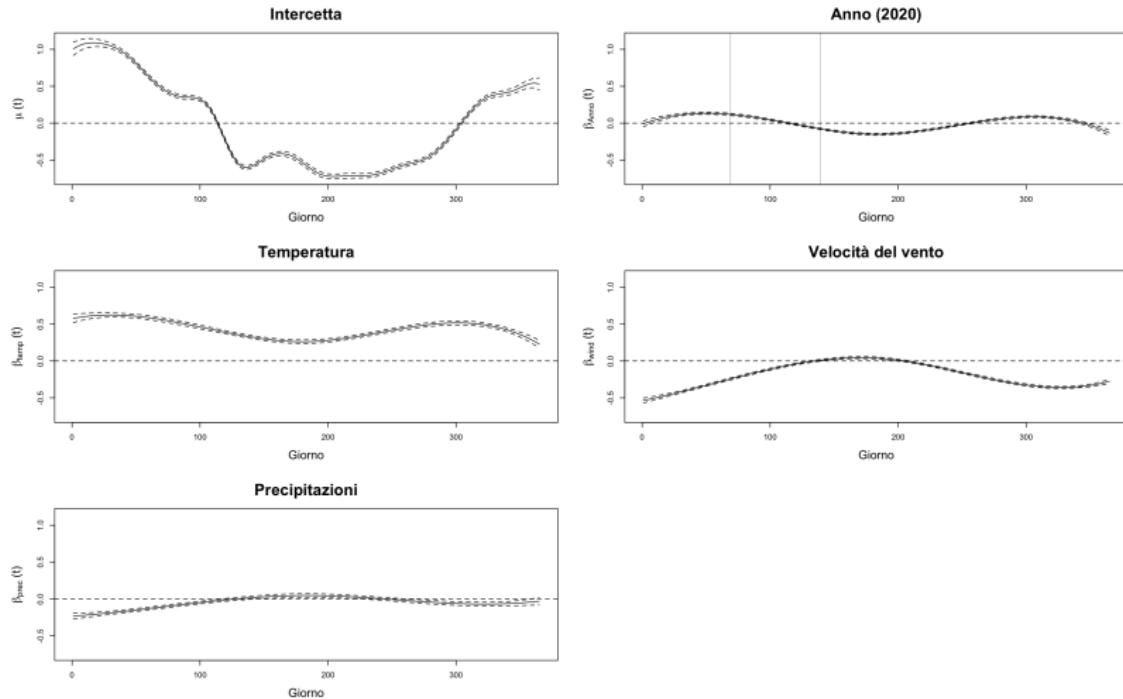
- Si prende in considerazione un modello simultaneo con l'inclusione delle tre variabili metereologiche (*Velocità del vento*, *Temperatura*, *Precipitazioni*) e della variabile *Anno*

$$\begin{aligned}\log PM10_i(t) = & \mu(t) + \text{anno}_i \beta_{\text{anno}}(t) + \text{temp}_i(t) \beta_{\text{temp}}(t) + \\ & + \text{wind}_i(t) \beta_{\text{wind}}(t) + \text{prec}_i(t) \beta_{\text{prec}}(t) + \varepsilon_i(t), \\ i = & 1, \dots, 60\end{aligned}$$



Valutazione dell'effetto del lockdown sul livello di PM10

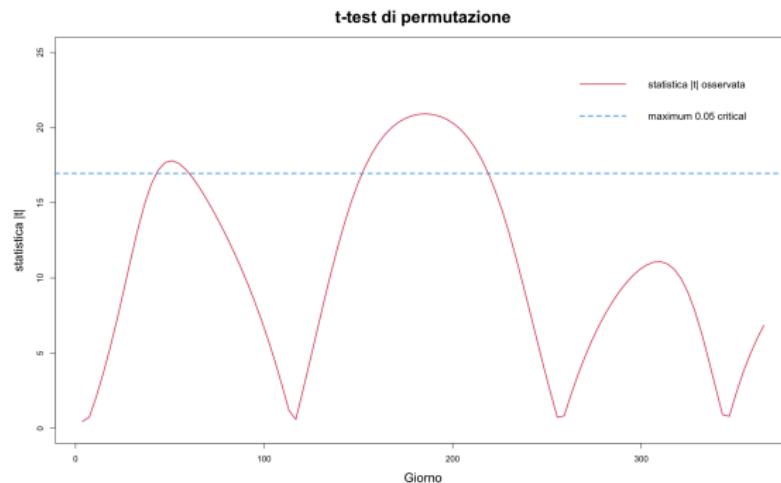
Modello simultaneo - Effetti Stimati



Valutazione dell'effetto del lockdown sul livello di PM10

Modello simultaneo - t-test di permutazione

- Si vuole valutare la significatività dell'effetto della variabile *Anno*:
- Si esegue un t-test di permutazione ($H_0: \beta_{anno}(t) = 0$ contro $H_1: \beta_{anno}(t) \neq 0, \forall t$) basato sulla procedura di Freedman-Lane³



³Freedman D., Lane D. (1983), *A nonstochastic interpretation of reported significance levels*



Valutazione dell'effetto del lockdown sul livello di PM10

Modello simultaneo con intercetta casuale - Specificazione del modello

- Si prende in considerazione l'aggiunta di un'intercetta casuale⁴ per ogni città

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

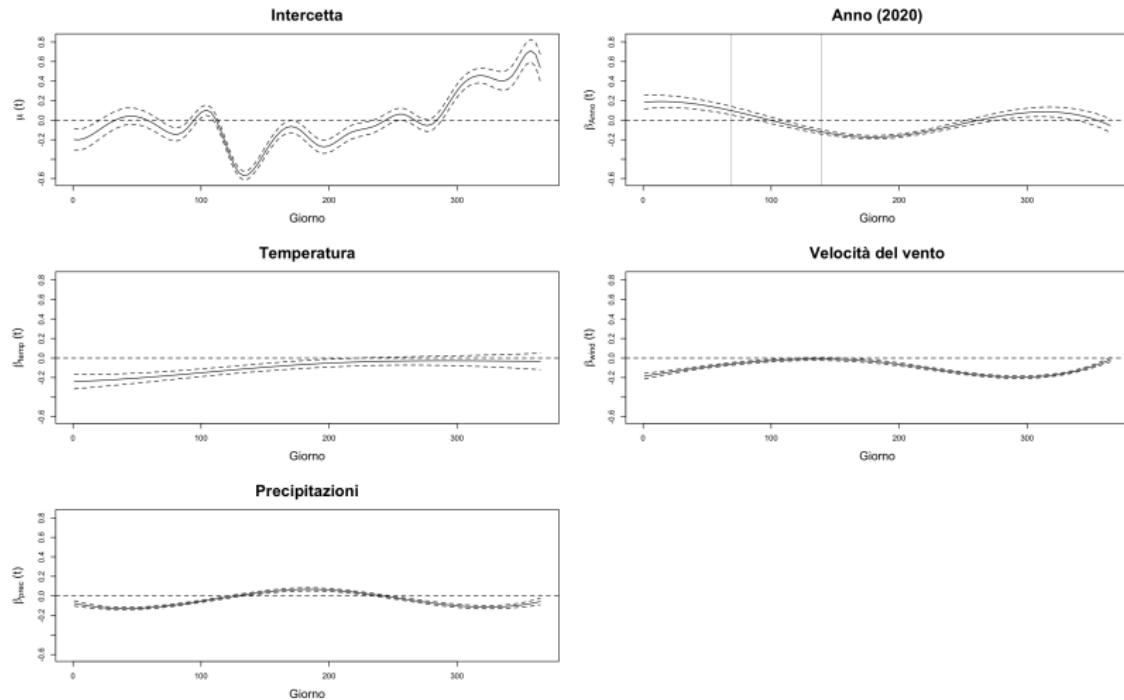
$$\text{con } città(t) = \mathbf{b}' \boldsymbol{\Phi}(t), \quad \mathbf{b} \sim N(0, \lambda R), \quad R = \int [L\boldsymbol{\Phi}(t)]' [L\boldsymbol{\Phi}(t)] dt$$

⁴Scheipl F. et al. (2015), *Functional Additive Mixed Models*



Valutazione dell'effetto del lockdown sul livello di PM10

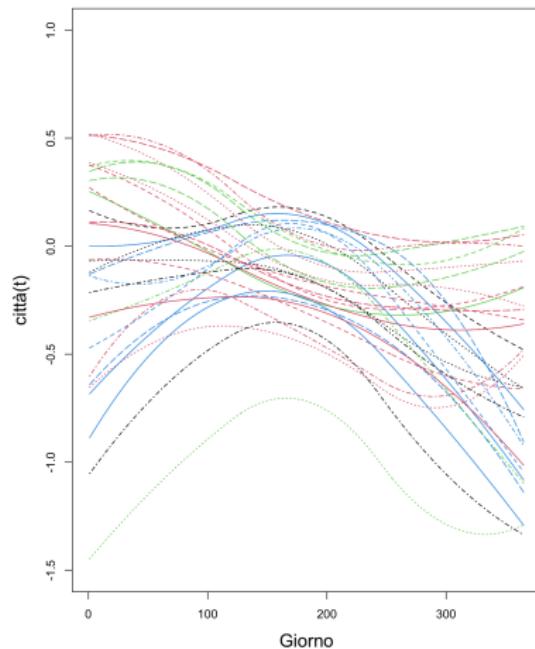
Modello simultaneo con intercetta casuale - Effetti Stimati



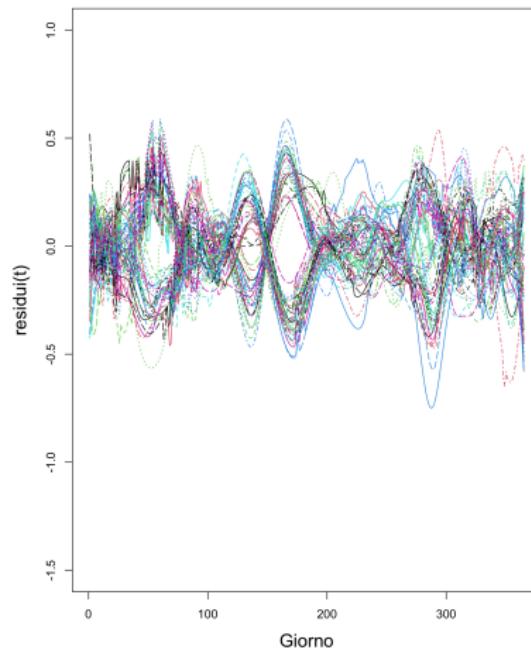
Valutazione dell'effetto del lockdown sul livello di PM10

Previsioni delle intercette casuali e Residui del modello

Intercette casuali

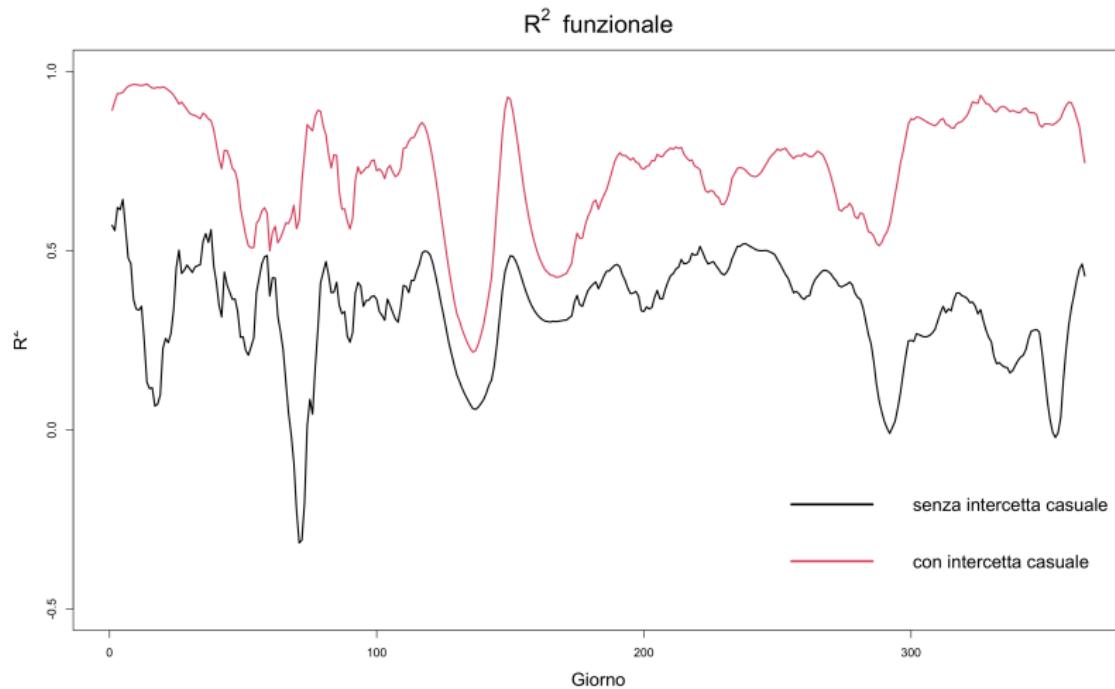


Residui



Valutazione dell'effetto del lockdown sul livello di PM10

Bontà di adattamento



Grazie per l'attenzione!

