

Bologna, 18-20 Maggio 2022



Progetto PULVIRUS: Applicazione del Machine Learning per la Normalizzazione Meteorologica su scala nazionale nel 2020

**M. Antonietta Reatini, Fabiana Scotti, Ilaria D'Elia,
Massimo D'Isidoro, Andrea Algieri**



Agencia nazionale per le nuove tecnologie,
l'energia e lo sviluppo economico sostenibile



Istituto Nazionale per lo
Studio e la Ricerca
in Environmental Protection and Research



European Network
for the Environment
for the Environment
Protection



Nasce dalla collaborazione di **ENEA, SNPA, ISS** per approfondire il legame tra inquinamento atmosferico e diffusione della pandemia, le interazioni fisico-chimiche-biologiche fra polveri atmosferiche e virus, per valutare gli effetti del lockdown sulle concentrazioni atmosferiche degli inquinanti e dei gas serra.

Il progetto ha poi avuto ambizioni più ampie, strutturandosi in **6 obiettivi**.

Per approfondire:
<https://www.pulvirus.it/>



Il progetto pulvirus: obiettivo 1



SCOPO DELL'OBIETTIVO 1,

TASK 4:

quantificare l'entità della riduzione della concentrazione dei principali inquinanti determinata dai provvedimenti adottati per ridurre la diffusione del contagio, **al netto della variabilità indotta dalle condizioni meteorologiche**

METODOLOGIE UTILIZZATE

1. GAM
2. **Random forest**
3. modello Bayesiano:
approccio INLA



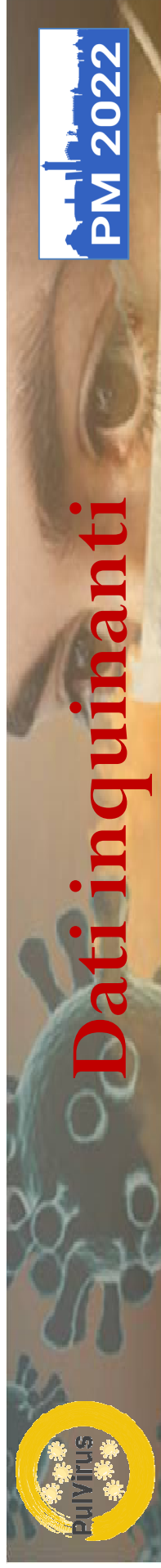
Il progetto pulvirus: obiettivo 1



GRUPPO DI LAVORO TASK4

- Coordinamento: Maria Antonietta Reatini (ISPRA)
- Andrea Algieri (ARPA Lombardia)
- Ilaria D'Elia, Massimo D'Isidoro (ENEA)
- Fabiana Scotto (Arpae Emilia-Romagna)

Il responsabile dell'obiettivo 1 nel suo complesso è
Giorgio Cattani (ISRPA)



Dati inquinanti

Sono stati analizzati di dati di **CO**, **C₆H₆**, **NO₂**, **O₃**, **PM₁₀** e **PM_{2.5}** delle centraline di tutt'Italia → **più di 1000 serie di dati validi**, divise per inquinante e distribuite nelle varie regioni.

- Periodo dal 2013 al 2019: dati archiviati in ISPRA nel database InfoARIA.
- Per l'anno 2020:
 - ❖ I dati relativi al primo semestre 2020 sono stati anticipati dalle agenzie
 - ❖ Nel mese di ottobre 2021 sono poi stati raccolti e completati i dataset con i dati relativi all'intero 2020 utilizzando il database Info-ARIA



Dati meteorologici

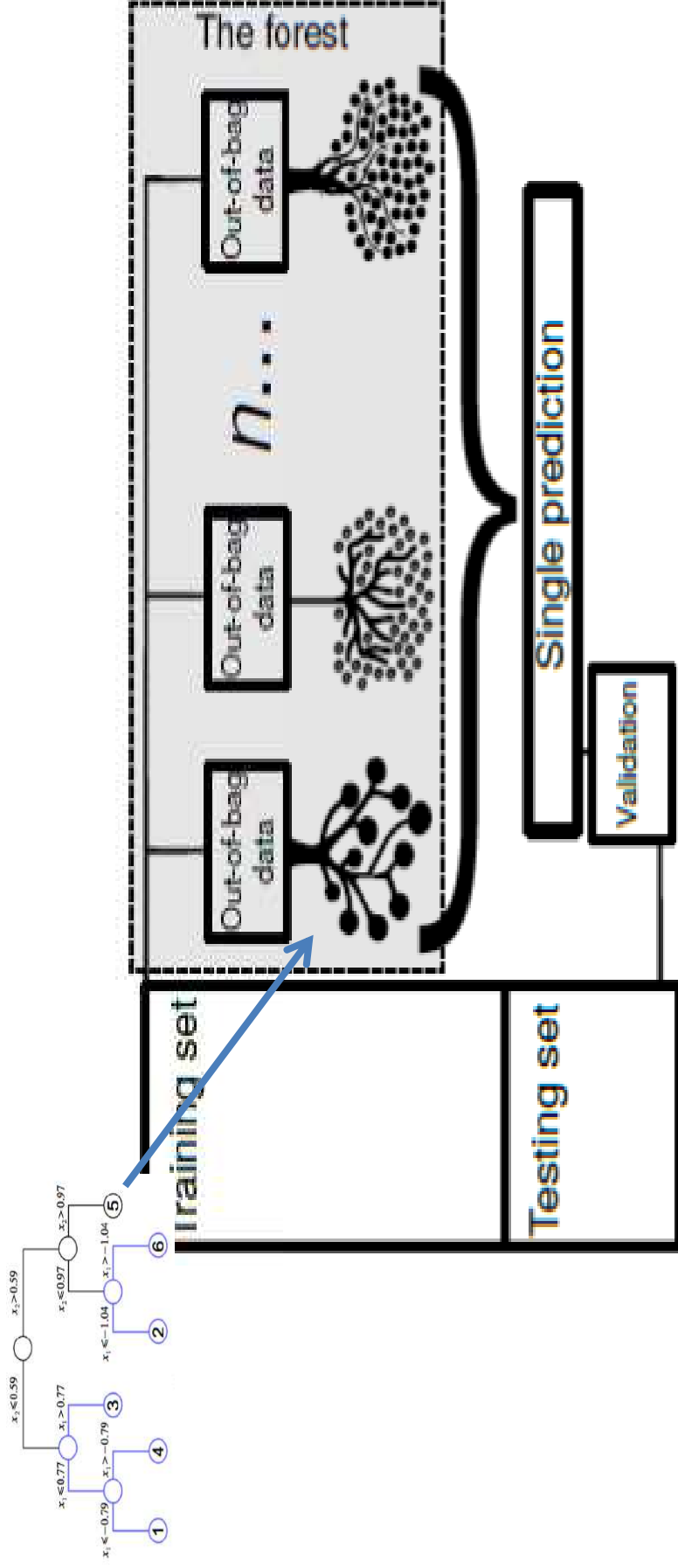


I dati meteorologici sono stati estratti dal set di dati ERA5 generato dal servizio Copernicus sui cambiamenti climatici, con risoluzione orizzontale (latitudine x longitudine) pari a $0.1^\circ \times 0.1^\circ$ ($0.25^\circ \times 0.25^\circ$ per la sola hpbl).



Metodologia: la random forest

PM 2022





Metodologia: la random forest



- Tecnica non parametrica di machine learning.
- Ha il grande vantaggio di non dover fare assunzioni come la normalità del campione, l'omoschedasticità, l'indipendenza, l'aderenza ad altre rigorose ipotesi parametriche e l'attenta gestione degli effetti di interazione che sono invece richieste per i modelli statistici.



Metodologia: la random forest



- Parametri RF:
 - ❖ 1000 alberi
 - ❖ Mtry = 5
- Le variabili sono state selezionate in base al contributo in termini di aumento di R^2 apportato al modello escludendo progressivamente le variabili molto correlate



Variabili utilizzate



Le **variabili meteorologiche** selezionate:

- Pressione
- Temperatura media(a 2m.)
- Precipitazione cumulata
- Precipitazione cumulata del giorno precedente
- Umidità relativa
- Irradianza solare netta
- Altezza dello strato limite planetario (minimo e massimo giornaliero)
- Direzione e velocità del vento



Variabili utilizzate



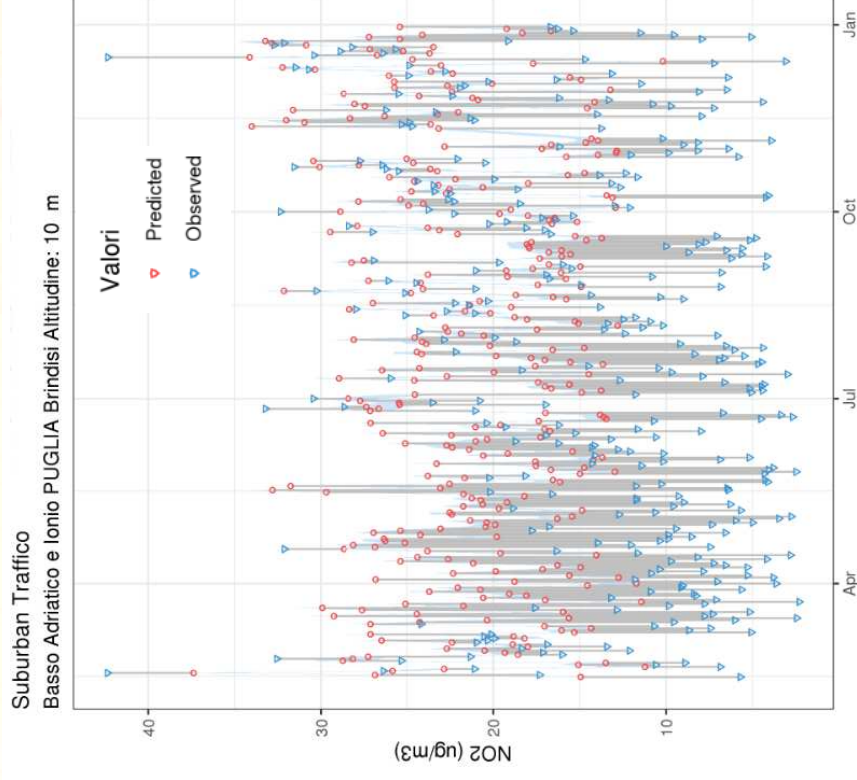
Inoltre sono stati utilizzati un termine di tendenza sotto forma di data Unix (numero di secondi dal 1° gennaio 1970), e due termini stagionali sotto forma di giorno giuliano (giorno dell'anno) e giorno della settimana.



Previsione Business As Usual



La previsione è stata effettuata sul **modello addestrato negli anni fino al 2019**: in questo modo, si ottiene una previsione che non tiene conto degli effetti della pandemia (“Business As Usual”), ma che tiene conto della situazione meteorologica verificatasi nel 2020 e la si confronta con il dato osservato. La (eventuale) **differenza tra dato osservato e stimato** viene intesa come **effetto della pandemia**, al netto dell’incertezza (e dell’eventuale distorsione) della stima del modello





Previsione Business As Usual



Periodo di **addestramento** del modello:
dal 1° gennaio 2013 al 31 dicembre 2019

Periodo di **validazione** del modello:
dal 1° gennaio 2020 al 14 febbraio 2020

Stime "**Business As Usual**":
dal 15 febbraio 2020 e fino a dicembre 2020



Stime di fitting e validazione



Le centraline superano la **fase di validazione** solo se vengono verificati i seguenti criteri:

- $NMSE_v \leq 0.5$ (Normalized Mean Square Error del periodo di validazione)
- $NMSE_t_{oob} \leq 0.5$ (Normalized Mean Square Error sui dati di training out of bag)
- $tr_r_squared_{oob} \geq 0.5$ (R^2 sui dati di training out of bag)
- $r_squared_test \geq 0.5$ (R^2 sui dati di testing)



Stime di fitting validazione



Le percentuali di modelli che non hanno superato tutti i criteri di validazione e di valutazione delle performance sono di seguito riportate suddivise per inquinante:

- NO_2 : 2,30%
- $\text{PM}_{2.5}$: 15,72%
- PM_{10} : 33,93%
- Benzene: 7,09%
- CO: 6.91%
- O_3 : tutti i modelli hanno superato i controlli.

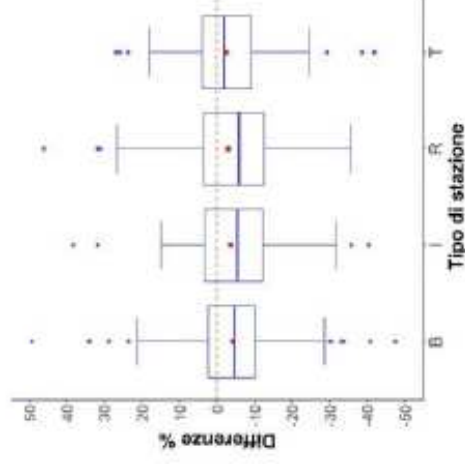
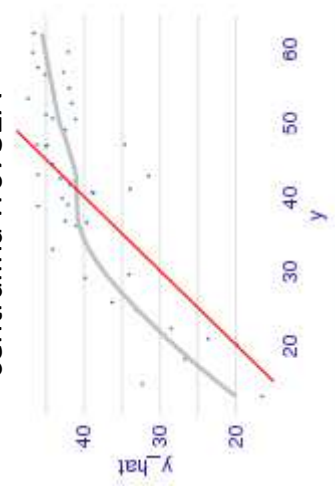


Stime di fitting e validazione

PM 2022

Per ogni inquinante sembra che la tendenza del modello sia quella di sovrastimare il dato per valori di concentrazioni bassi e di sottostimare per valori di concentrazioni alti; questo comportamento è abbastanza comune in tutti gli algoritmi di ML che tendono a cogliere bene soprattutto i valori modal

Esempio: stima dell' NO_2
centralina IT0732A



ossido di azoto. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello BAU e concentrazione media osservata (periodo 1 gennaio 2020 - 14 febbraio 2020).



Risultati

PM 2022

PREMESSA:

6 diversi inquinanti, 8 diversi mesi e tantissime centraline e risultati “poco coerenti” → difficile fare una sintesi!

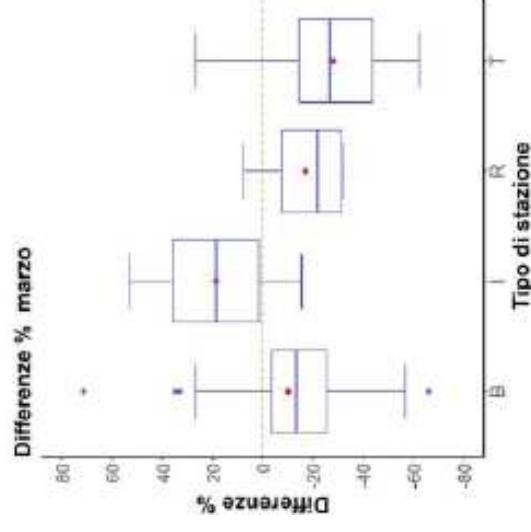
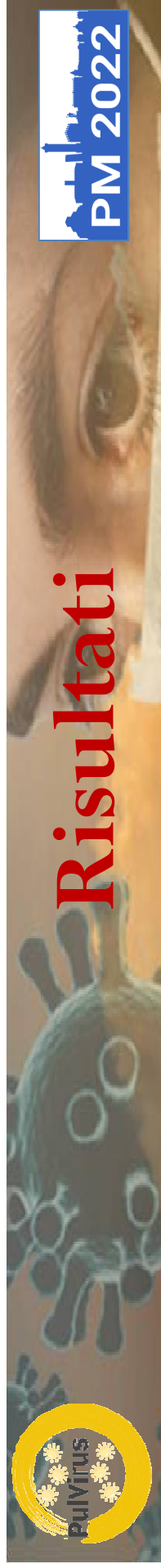
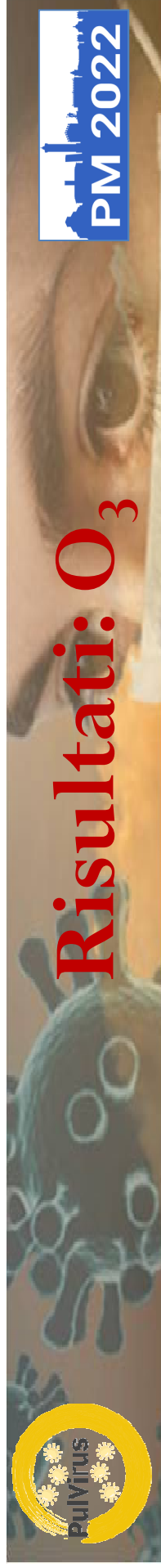


Fig. 3-4 Benzene, marzo 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello BAU e concentrazione media osservata

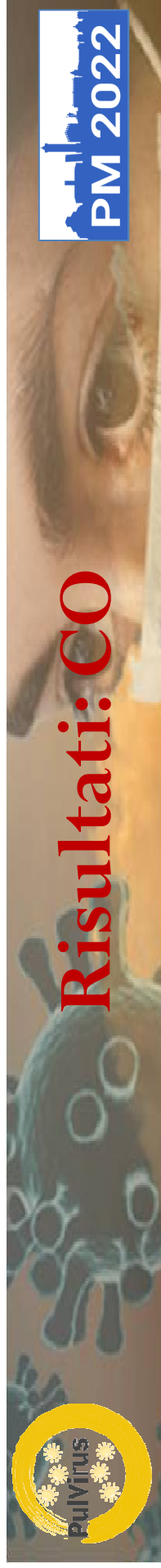


Risultati

- ❖ A seconda dell'inquinante si è osservata una diminuzione delle concentrazioni diversa e un diverso comportamento nei mesi.
- ❖ I gas "primari" presentano un comportamento eterogeneo tra le tipologie di stazioni che non si osserva per NO_2 , O_3 e particolato



- ❖ L'ozono **non** presenta **variazione**, positiva o negativa, in alcun periodo dell'anno



Il **CO** presenta un calo, durante i mesi primaverili di lockdown, nelle sole stazioni da traffico, industriali e di background urbano (non per le centraline rurali); il calo si osserva ancora solo per il mese di giugno.

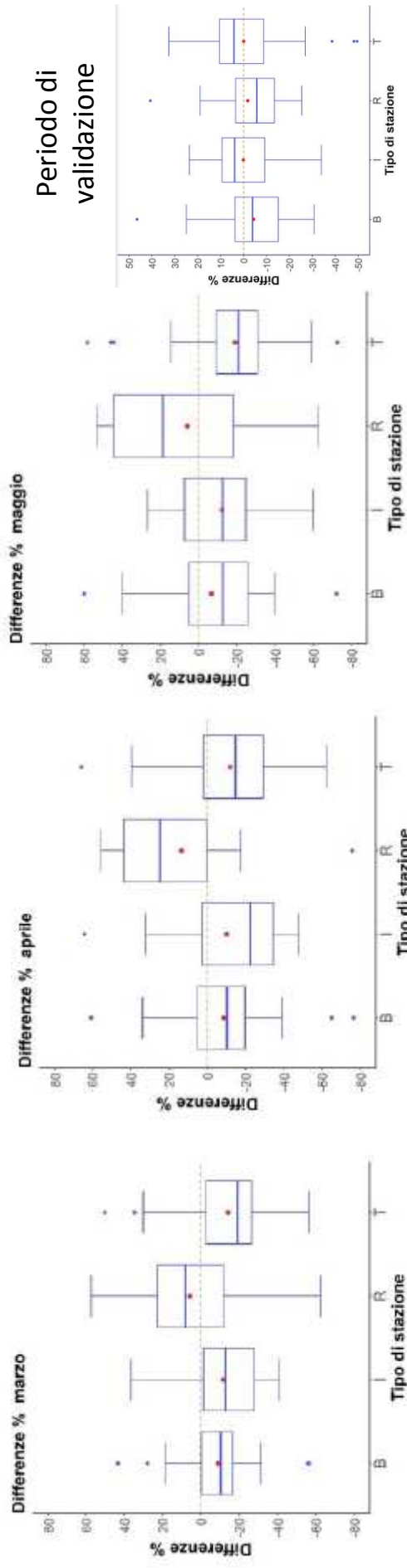


Fig. 47 CO, marzo 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello EAU e concentrazione media osservata

Fig. 48 CO, aprile 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello EAU e concentrazione media osservata

Fig. 49 CO, maggio 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello EAU e concentrazione media osservata



Risultati: C₆H₆

• Decremento durante tutto l'anno e non solo per il periodo di lockdown di marzo-maggio.

• Stazioni industriali: andamento variabile durante l'anno, con differenze percentuali anche inaspettatamente positive in alcuni mesi.

• Stazioni rurali: calo importante durante molti mesi dell'anno e in particolar modo durante i mesi estivi (contrariamente al CO!), non nei mesi di novembre e dicembre

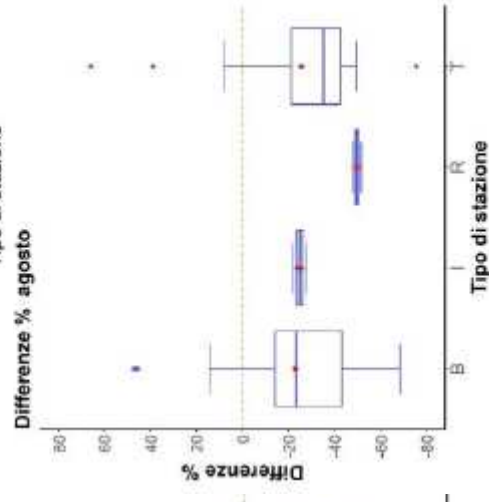
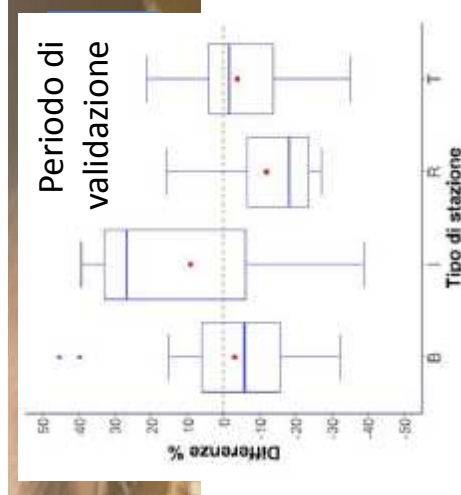


Fig. 35 Benzene, aprile 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello BAU e concentrazione media osservata

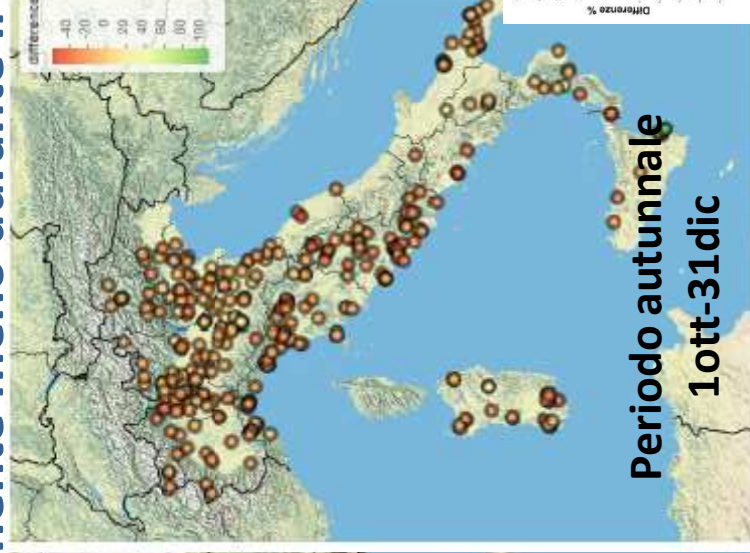
Fig. 39 Benzene, agosto 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello BAU e concentrazione media osservata



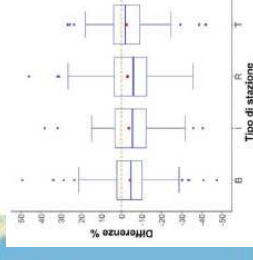
Risultati: NO₂

PM 2022

Decremento tra il 30% e 40% (x le stazioni rurali → decremento inferiore)
durante il lockdown e intorno al 20% o leggermente meno durante il resto
dell'anno



Periodo di
validazione





Risultati: PM

PM 2022

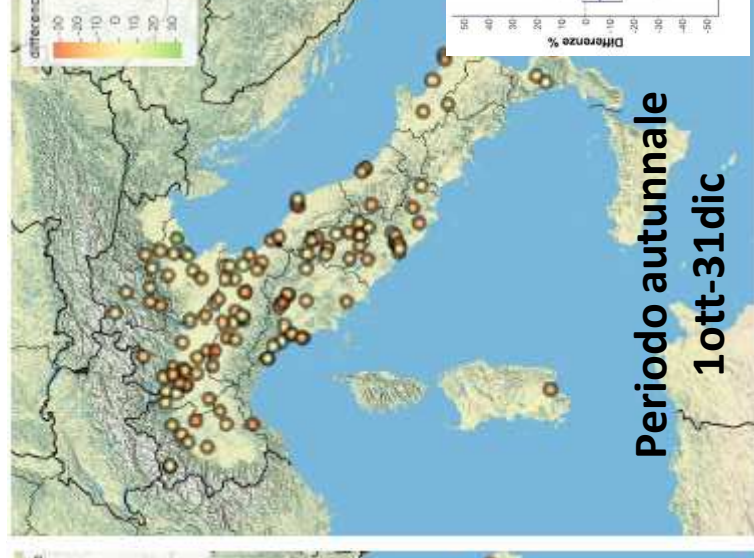
Diminuzioni limitate o nulle durante il lockdown; maggiori (entro il 20%) nel periodo estivo e discontinue nel periodo autunnale



Periodo lockdown
1 mar-30 apr



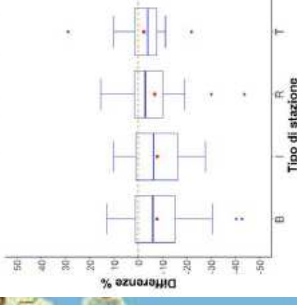
Periodo estivo
1 mag-30 sett



Periodo autunnale
1 ott-31 dic

PM_{2.5}

Periodo di
validazione





Risultati: PM



PM_{2.5}

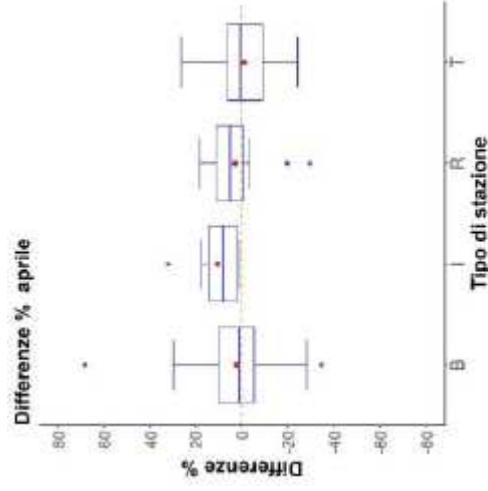


Fig. 100 PM_{2.5} aprile 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello BAU e concentrazione media osservata

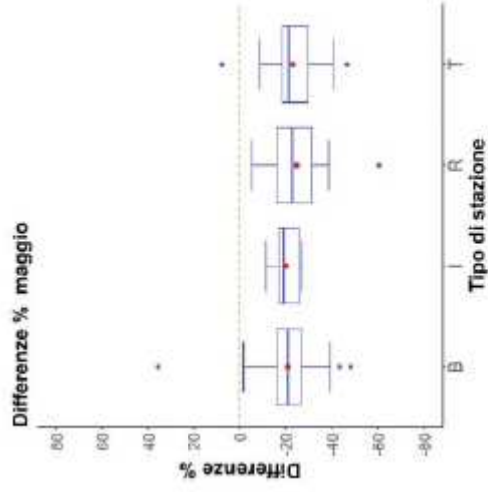


Fig. 101 PM_{2.5} maggio 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello BAU e concentrazione media osservata

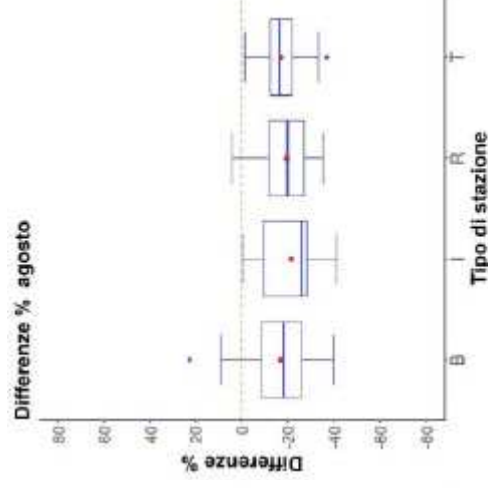
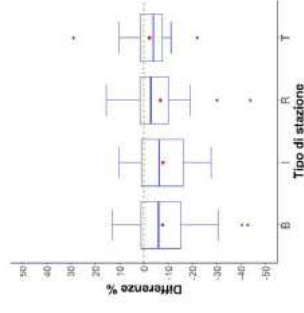


Fig. 104 PM_{2.5} agosto 2020. Box plot delle differenze percentuali, per stazione e tipologia di stazione, tra concentrazione media stimata dal modello BAU e concentrazione media osservata

Periodo di validazione





Conclusioni



Durante i mesi di lockdown la diminuzione della concentrazione degli inquinanti non è stata eclatante quanto ci si sarebbe potuti aspettare e soprattutto non è stata coerente e costante per tutti gli inquinanti (nemmeno i primari) e le tipologie di centraline; **l'NO₂ è l'inquinante che ha mostrato sia il calo maggiore, sia la maggior coerenza, come andamento nei mesi e uniformità tra le tipologie di stazioni, mentre l'O₃ non ha mostrato alcuna variazione.**

Al contempo, per alcuni inquinanti si sono osservate diminuzioni nelle concentrazioni anche durante i mesi estivi o autunnali, seppure, a volte, senza coerenza tra i diversi mesi o le diverse tipologie di stazione.



Fine

Grazie per l'attenzione!