



Napoli, 27 Maggio 2026

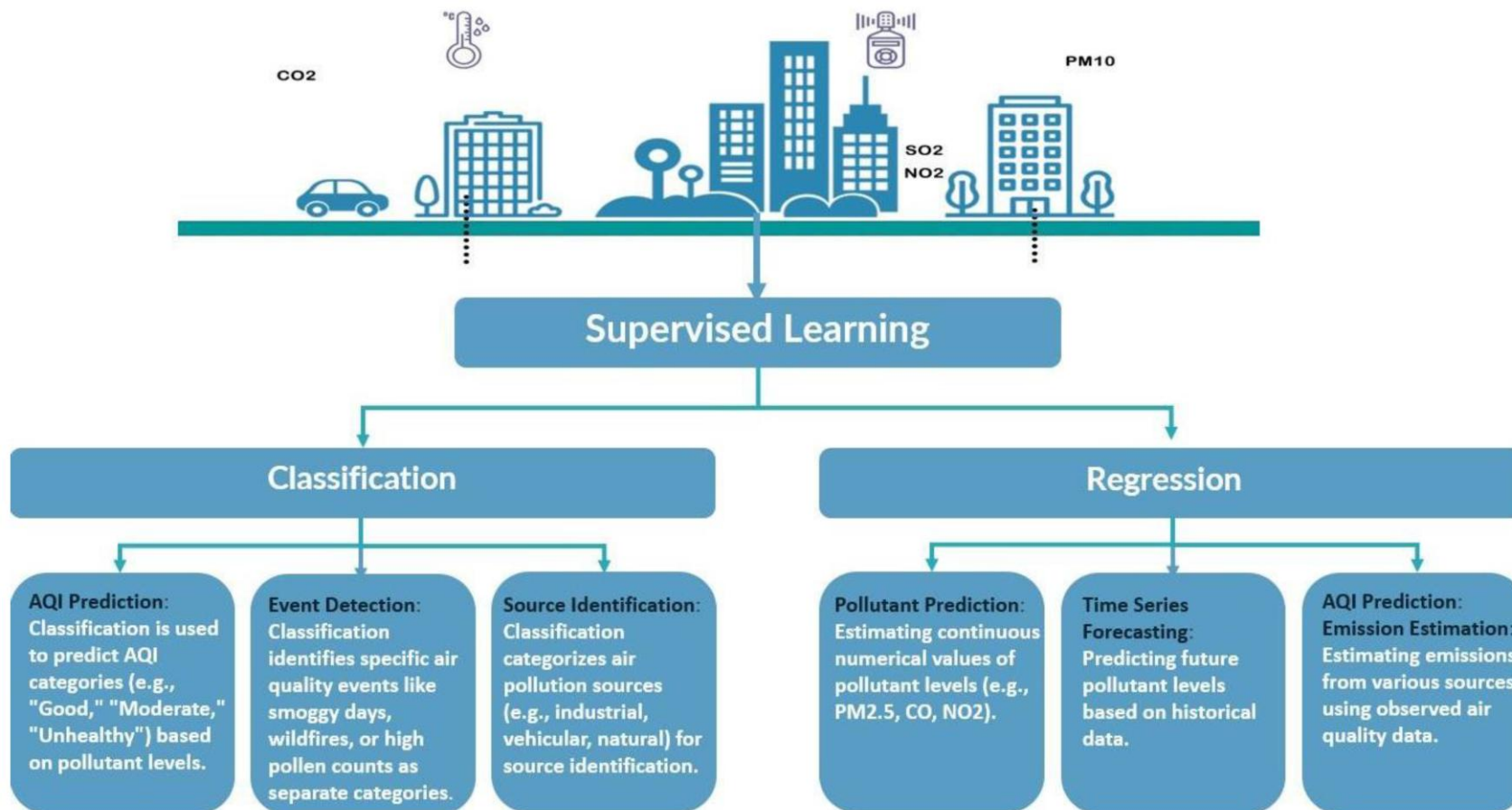
Integrazione di tecniche di Machine Learning in un sistema modellistico previsionale della qualità dell'aria

Andrea Bisignano, Chiara Righi

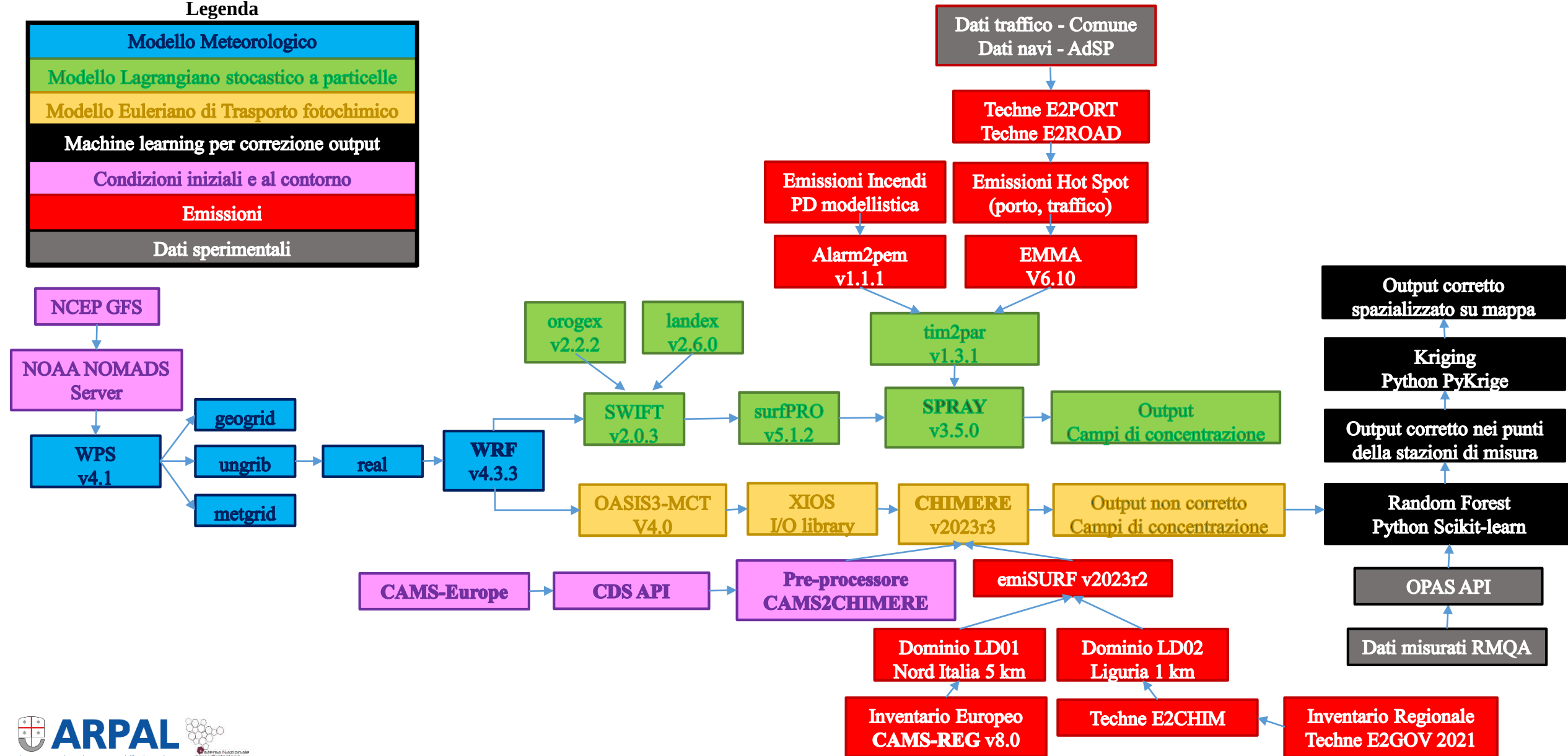
UO Stato Qualità Aria, ARPA Liguria



- I modelli di trasporto fotochimico come CHIMERE presentano errori sistematici rispetto alle osservazioni.
- Tali errori non sono costanti, ma dipendono dalle condizioni meteorologiche, dalla stagione, dalla variabilità locale ecc.
- Le osservazioni disponibili sono accurate, ma spazialmente sparse.
- **Nella nuova normativa la modellistica non è solo un supporto, ma uno strumento obbligatorio per estendere le misure.**
- **Lo scopo è costruire un campo di concentrazione più realistico, combinando entrambe le fonti (simulazioni e misure)**



Legenda





Outer (Ld01)

- Risoluzione 5 km, 133 x 73 punti griglia. 45 (WRF) e 17 (CHIMERE) livelli verticali
- Emissioni: CAMS-reg v8.0
- Condizioni iniziali ed al contorno QA: CAMS Europe – CAMS Globe
- Condizioni iniziali meteo: GFS



Inner (Ld02)

- Risoluzione 1 km, 211 x 111 punti griglia, 45 (WRF) e 17 (CHIMERE) livelli verticali.
- Emissioni: Inventario regionale E2GOV
- Condizioni iniziali ed al contorno: Ld01
- Condizioni iniziali meteo: da Ld01

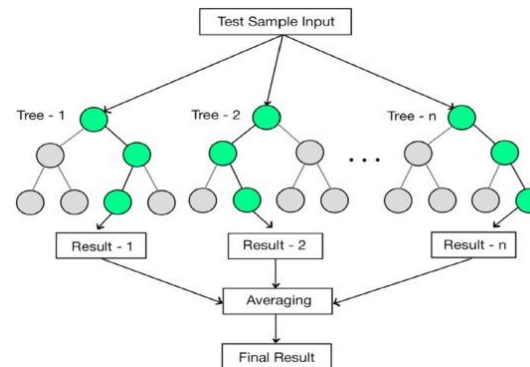
➤ **Due simulazioni giornaliere** con orizzonte di 48 ore: configurazione mattutina con condizioni al contorno GFS 00z e CAMS-Global e configurazione pomeridiana con condizioni al contorno GFS 12z e CAMS-Europe.

➤ **Configurazione Hardware:** 4 nodi Dell R760, ciascuno equipaggiato con CPU Intel Xeon Gold 6548Y (5a gen), 64 core fisici e 256 GB di RAM

Il metodo presentato è un approccio standard in meteorologia statistica, ormai ampiamente usato anche per la qualità dell'aria; il sistema è basato su due livelli: **correzione locale del bias e interpolazione geostatistica**

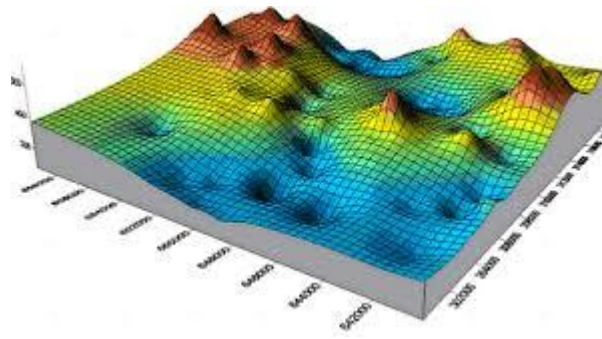
Livello 1

- Il primo livello utilizza un modello di machine learning (i.e. **Random Forest**) per correggere il bias locale del modello
- In pratica: il modello viene corretto nelle celle dove sono presenti stazione della rete di monitoraggio



Livello 2

- Il secondo livello utilizza il **Kriging universale con external drift** per ricostruire la struttura spaziale degli errori residui
- In pratica: la correzione del livello precedente viene propagata nel domino in modo continuo e coerente



- **Cos'è:** è un modello supervisionato di regressione
- **Come funziona:** costruisce **ensemble di alberi decisionali (Forest) su campioni bootstrap dei dati (Random)**; ogni albero predice un valore numerico; la predizione finale è la media delle singole predizioni albero vota
- **Input:** campo di concentrazione Chimere + variabili meteorologiche e di dispersione WRF (**predittori o variabili esplicative**), osservazioni stazioni
- **Cosa apprende:** la relazione tra output grezzo del modello fisico e **l'errore sistematico**
- **Output:** correzione locale nelle sole celle con osservazioni
- **Libreria:** Python **scikit-learn**(modulo **RandomForestRegressor**)

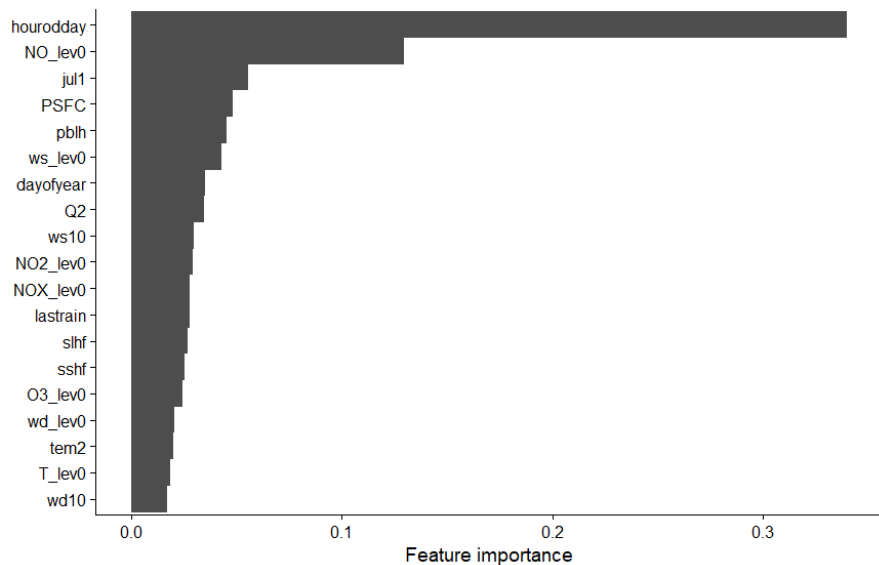
Il modello viene corretto operativamente ad ogni previsione con i dati disponibili dal 1° gennaio 2024 fino all'ultimo giorno del mese precedente alla previsione corrente; il training viene aggiornato quindi con frequenza mensile.

Per ridurre il rischio di leakage, la validazione viene effettuata con cross-validation spaziale e non con leave-one-station-out completa (buon compromesso tra robustezza statistica e sostenibilità computazionale)

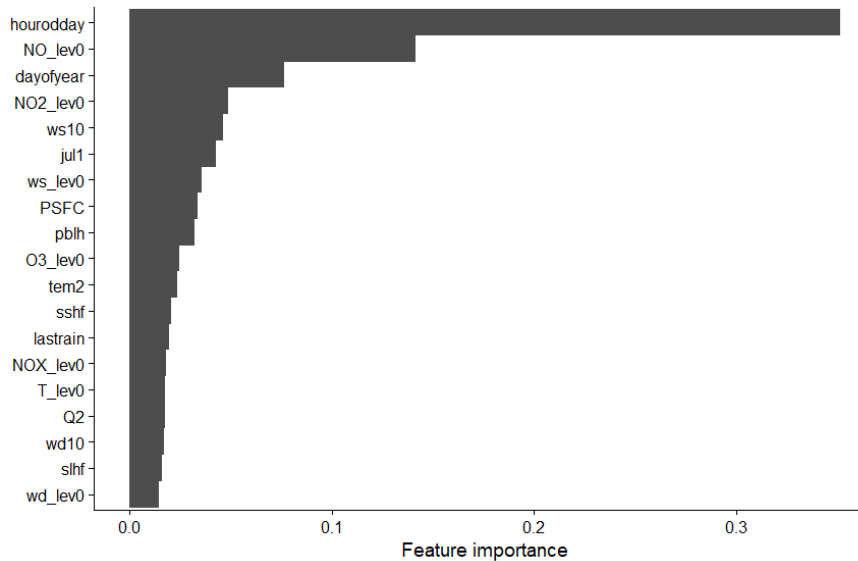
- **Cos'è:** è un metodo di interpolazione geostatistica che permette di stimare valori in punti non misurati, sfruttando la correlazione spaziale
- **Come funziona:** assume dipendenza spaziale tra le osservazioni e modella la struttura tramite variogramma; separa componente deterministica (trend) e componente stocastica (variabilità residua)
- **Tipologia:** include trend spaziale non costante e valore medio non fisso; la parte deterministica viene forzata da una variabile esterna (si parla di **kriging universale con drift esterno**)
- **Input:** residui Random Forest nelle celle con stazioni e output grezzo CHIMERE come drift
- **Output:** Campo continuo dei residui e quindi della concentrazione
- **Libreria:** Python **PyKrige**, metodo: **Universal Kriging, External Drift**

La validazione viene effettuata coerentemente con la cross-validation spaziale adottata per la Random Forest

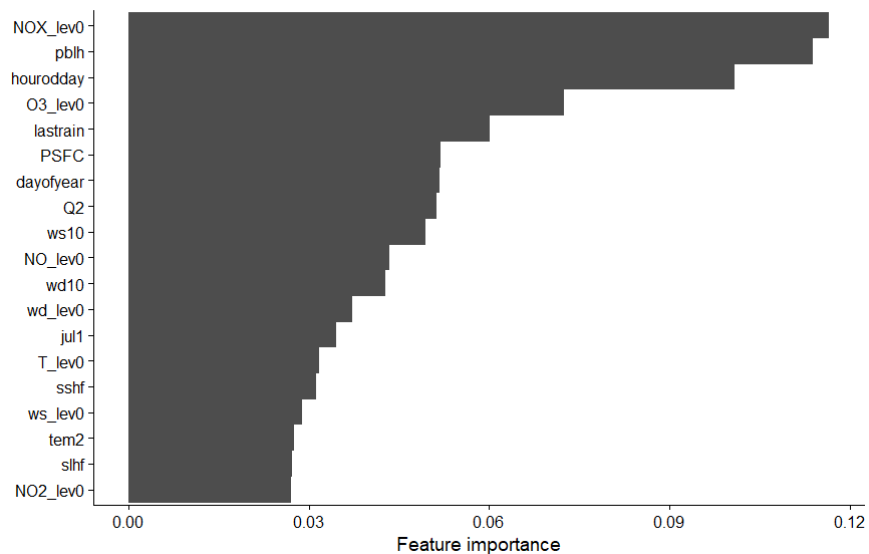
Genova Via Buoizzi NO2



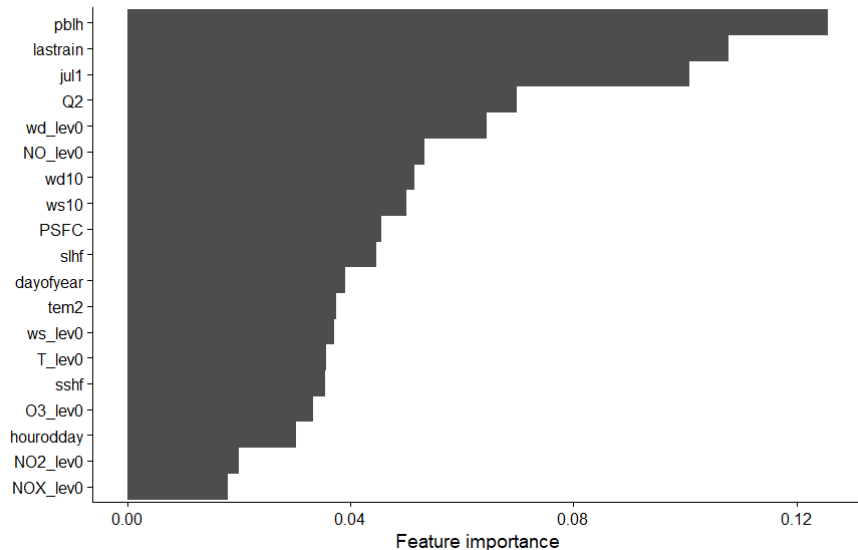
Genova Europa Via San Martino NO2



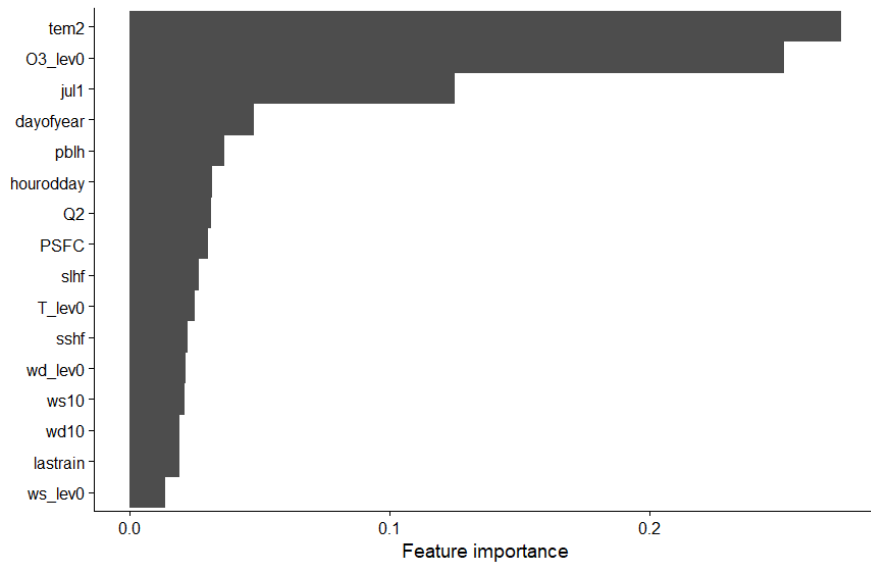
Genova Quarto NO2



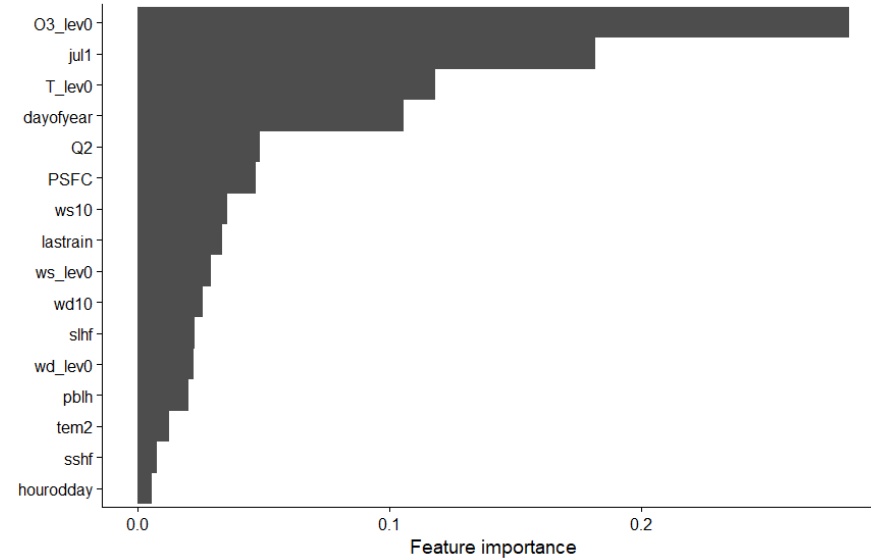
Genova Propata NO2



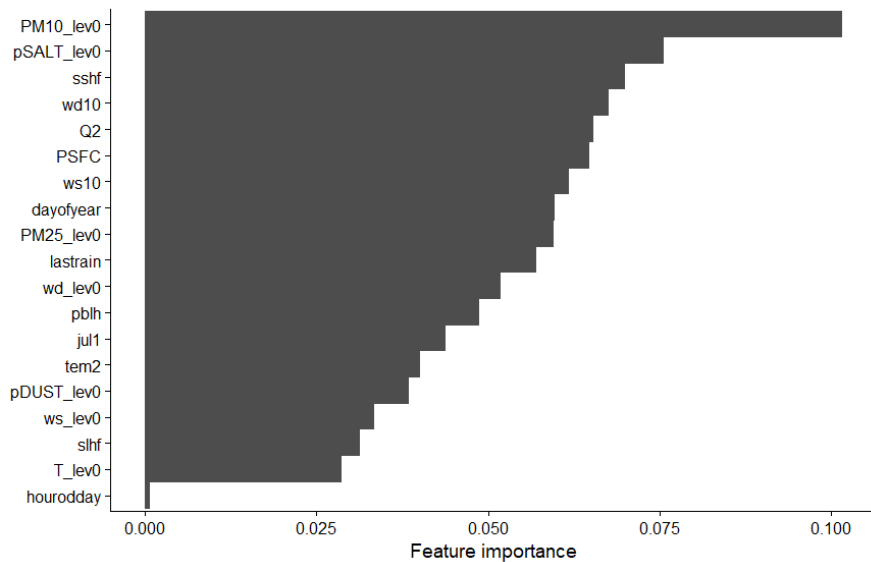
Genova Quarto O3



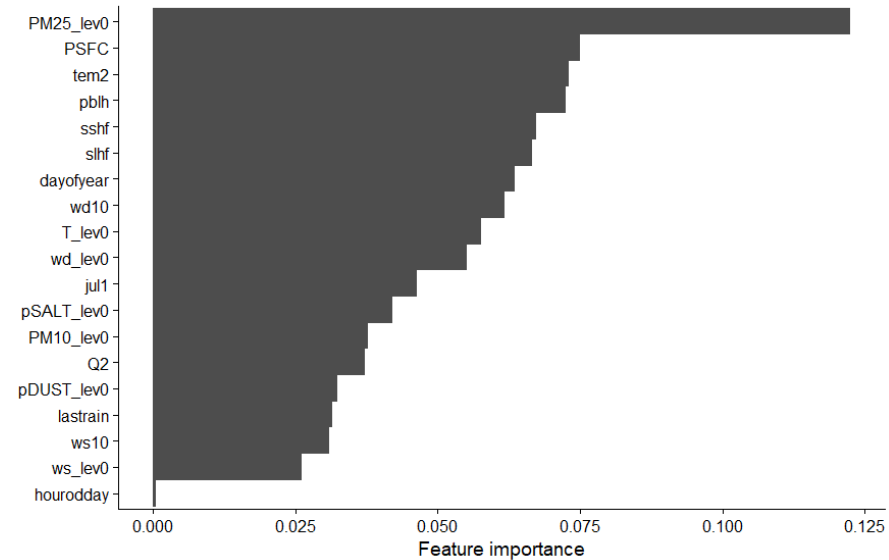
Genova Propata O3



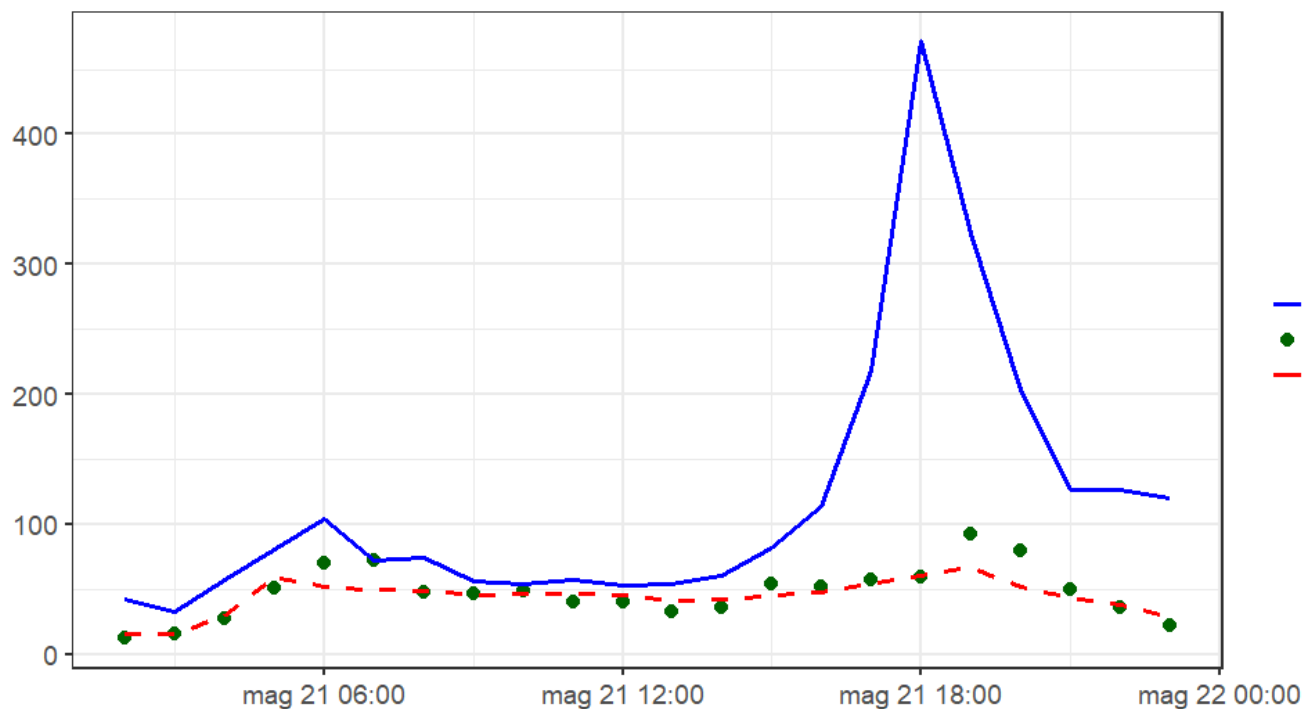
Genova Cso Buenos Aires PM10



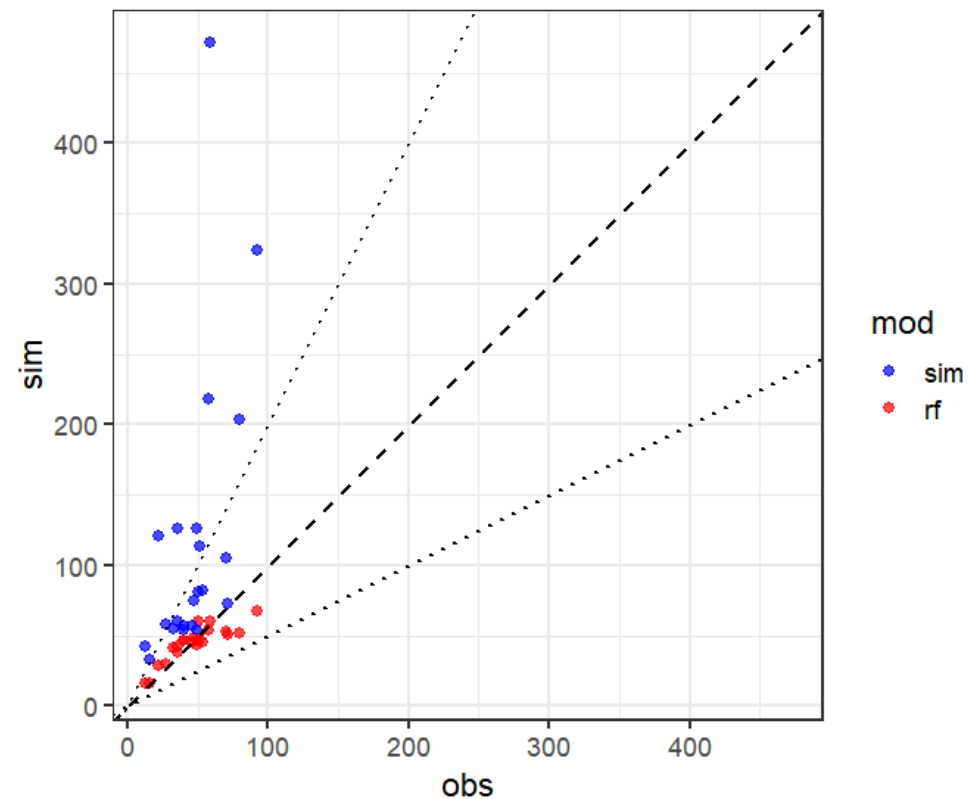
Cengio SV PM10



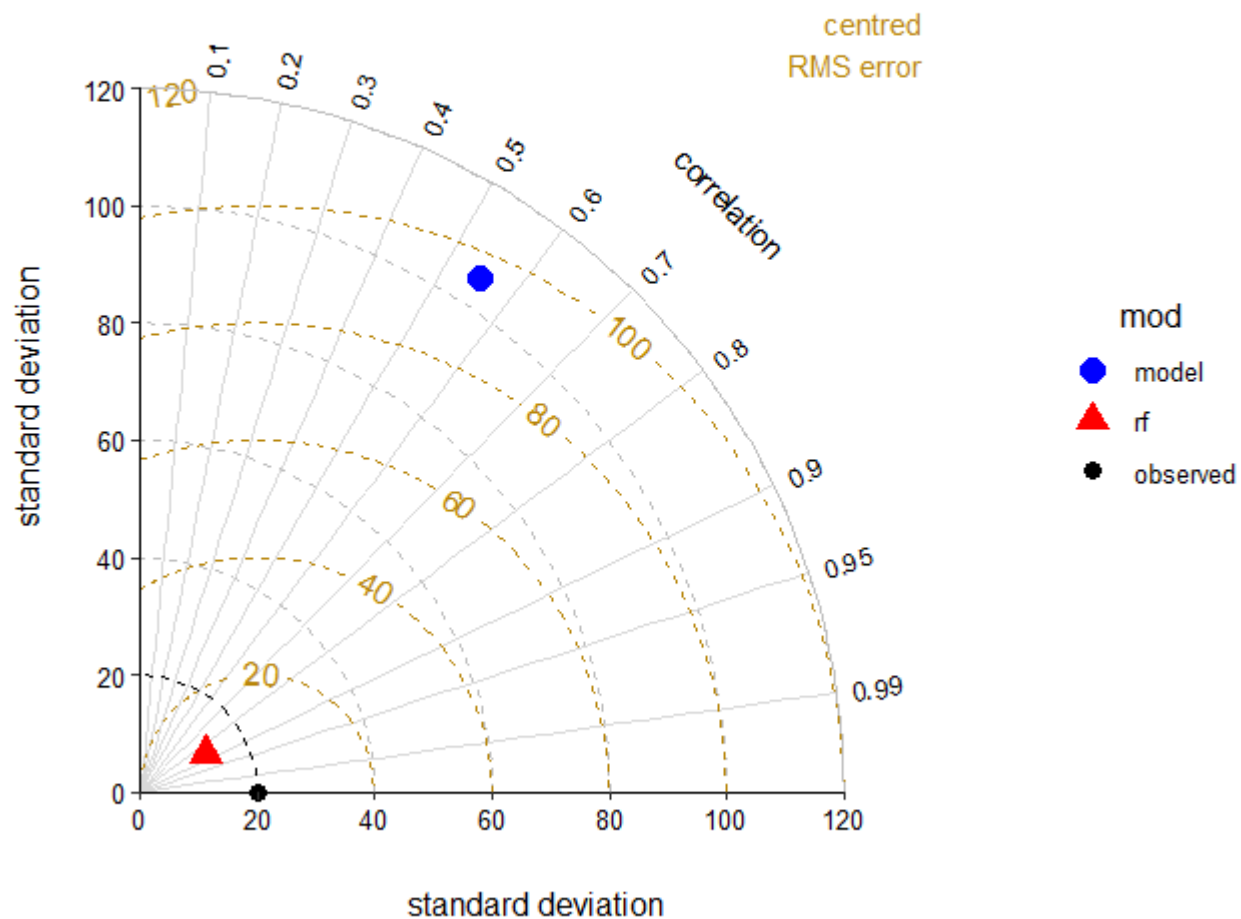
IT1698A Genova - Europa/Via San Martino Urbana



IT1698A Genova - Europa/Via San Martino Urbana

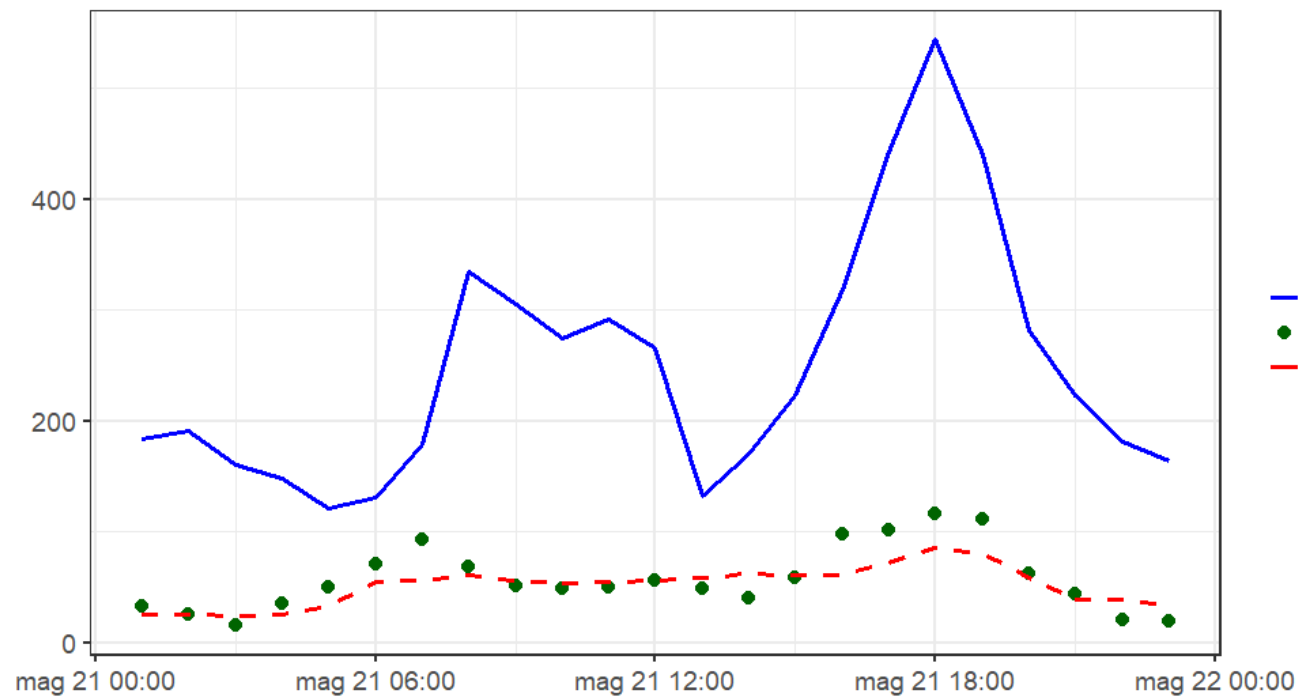


IT1698A Genova - Europa/Via San Martino Urbana

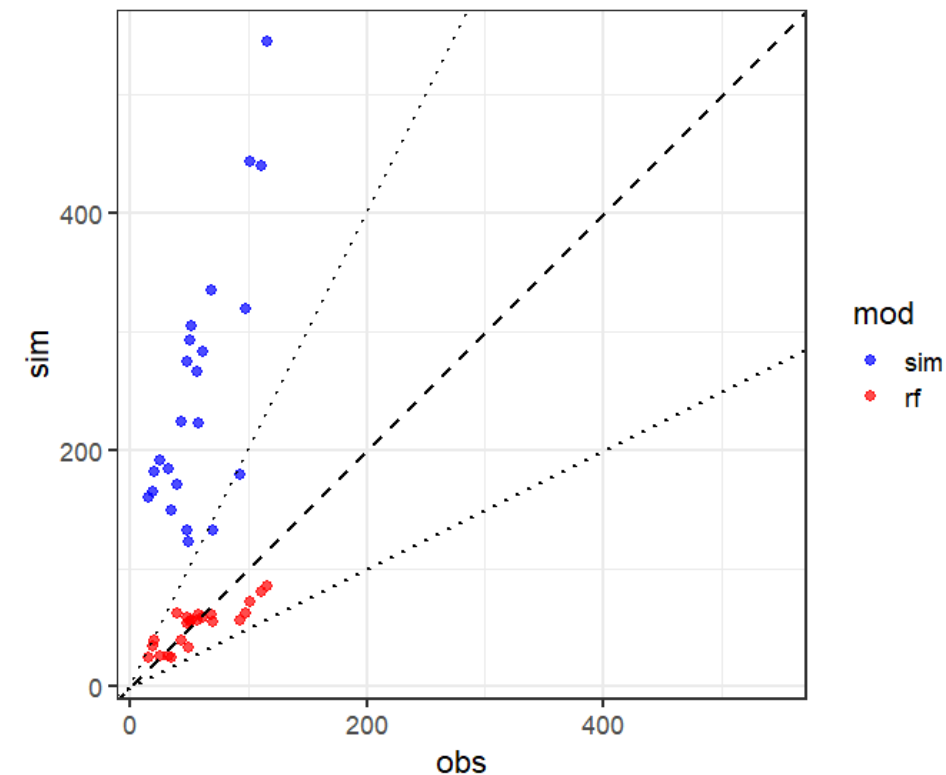


	model	rf
n	22	22
FAC2	0.5	1.0
MB	70.26	-2.89
MGE	70.26	7.86
NMB	1.49	-0.06
NMGE	1.49	0.17
RMSE	116.74	11.08
r	0.55	0.87

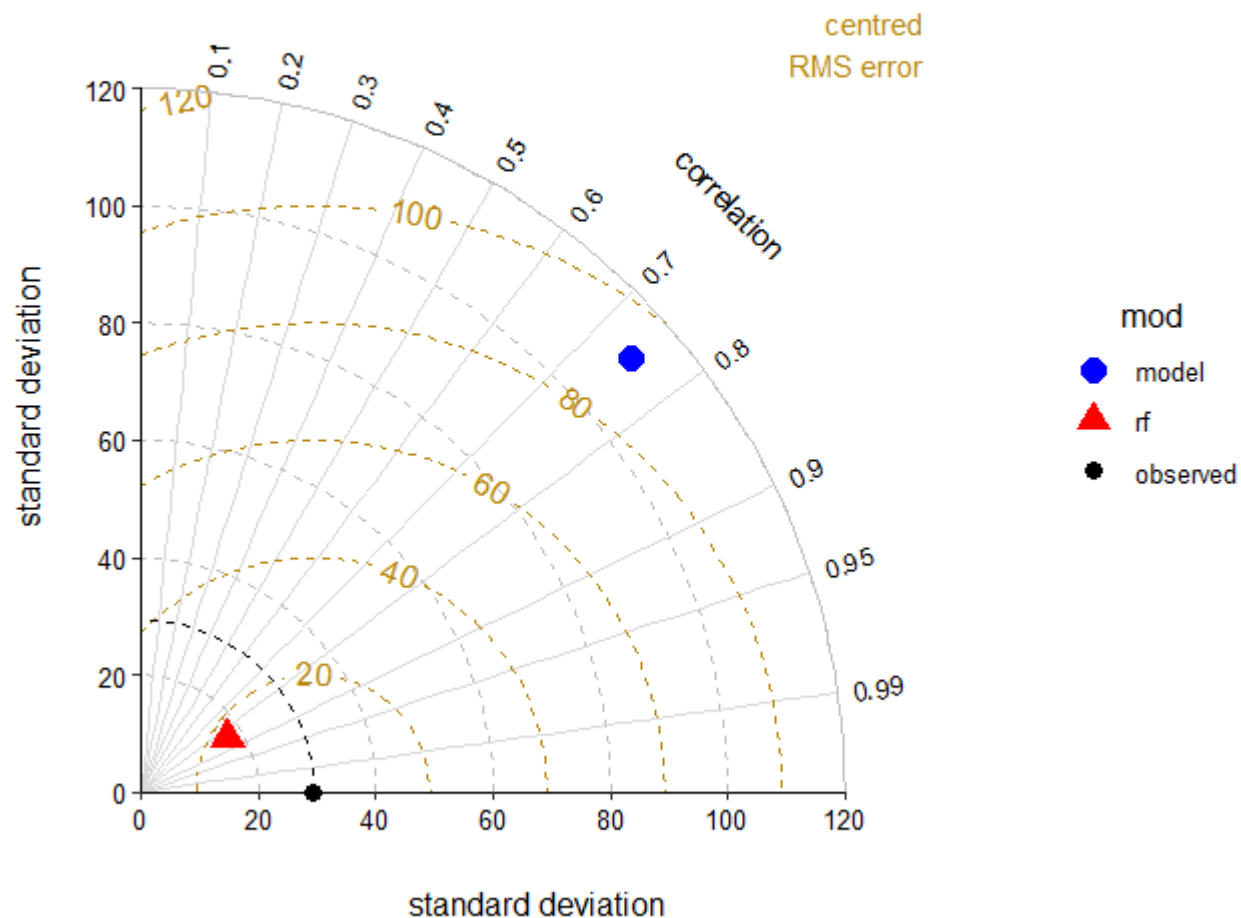
IT1887A Genova - Buoizzi Urbana



IT1887A Genova - Buoizzi Urbana

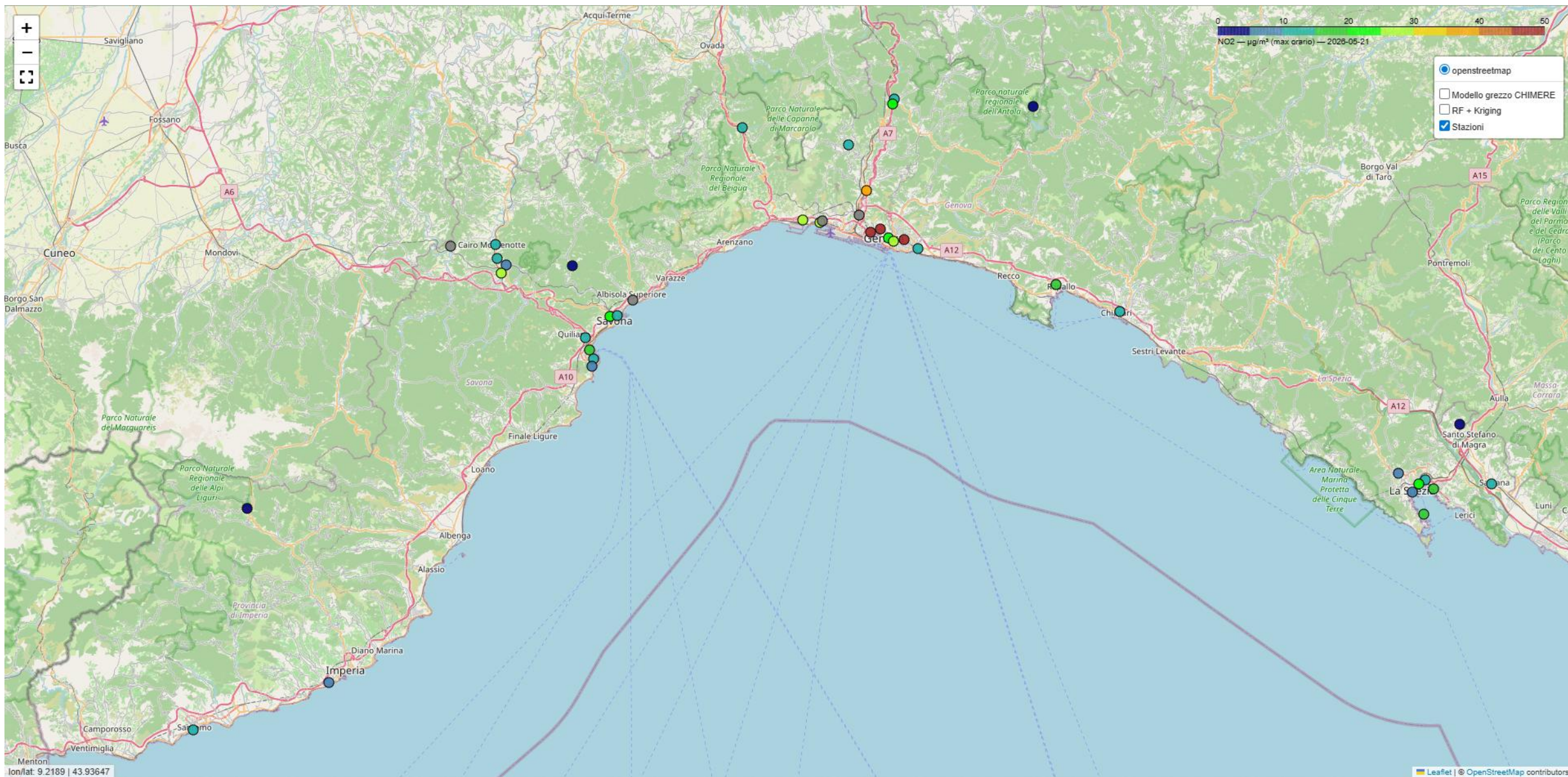


IT1887A Genova - Buoizzi Urbana

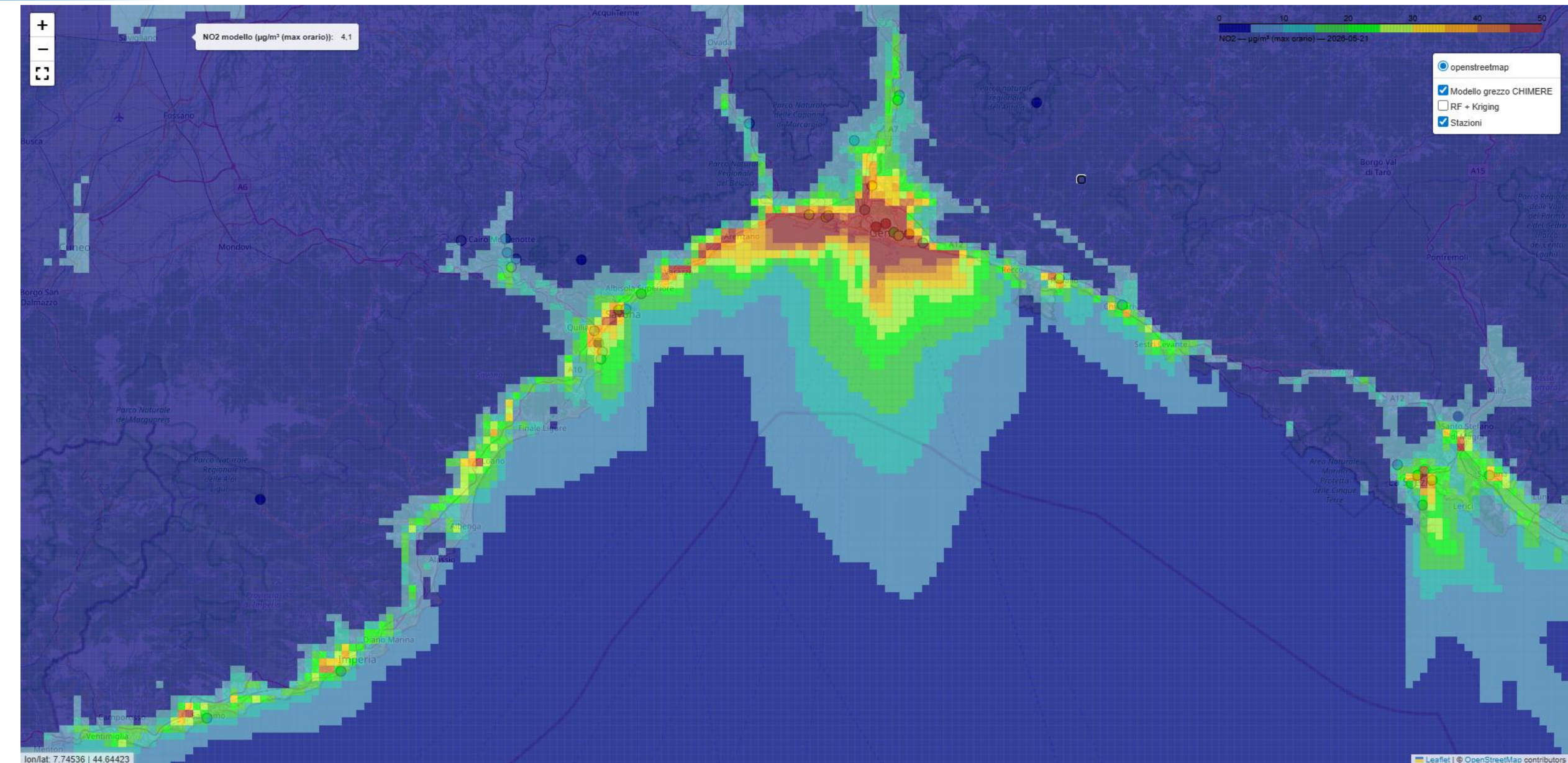


	model	rf
n	23	23
FAC2	0.09	1.00
MB	190.77	-6.30
MGE	190.77	13.94
NMB	3.32	-0.11
NMGE	3.32	0.24
RMSE	210.78	18.08
r	0.75	0.84

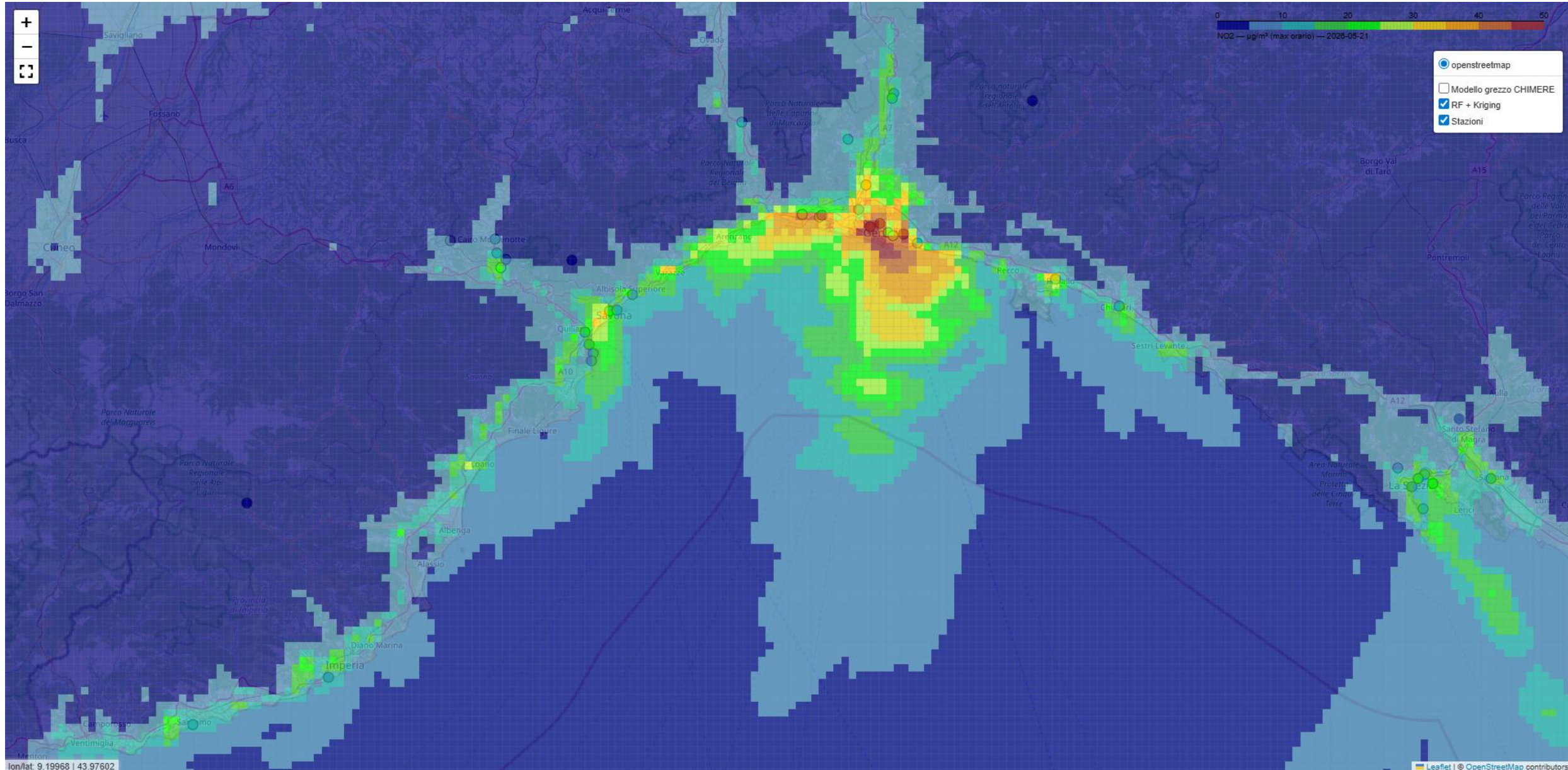
Risultati NO₂



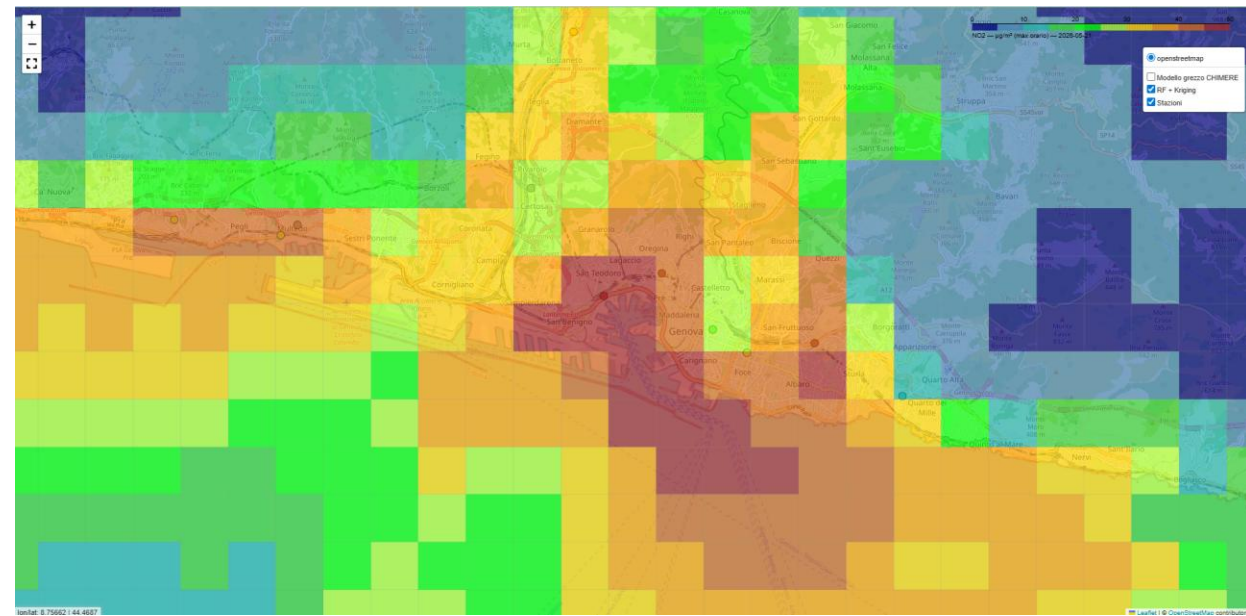
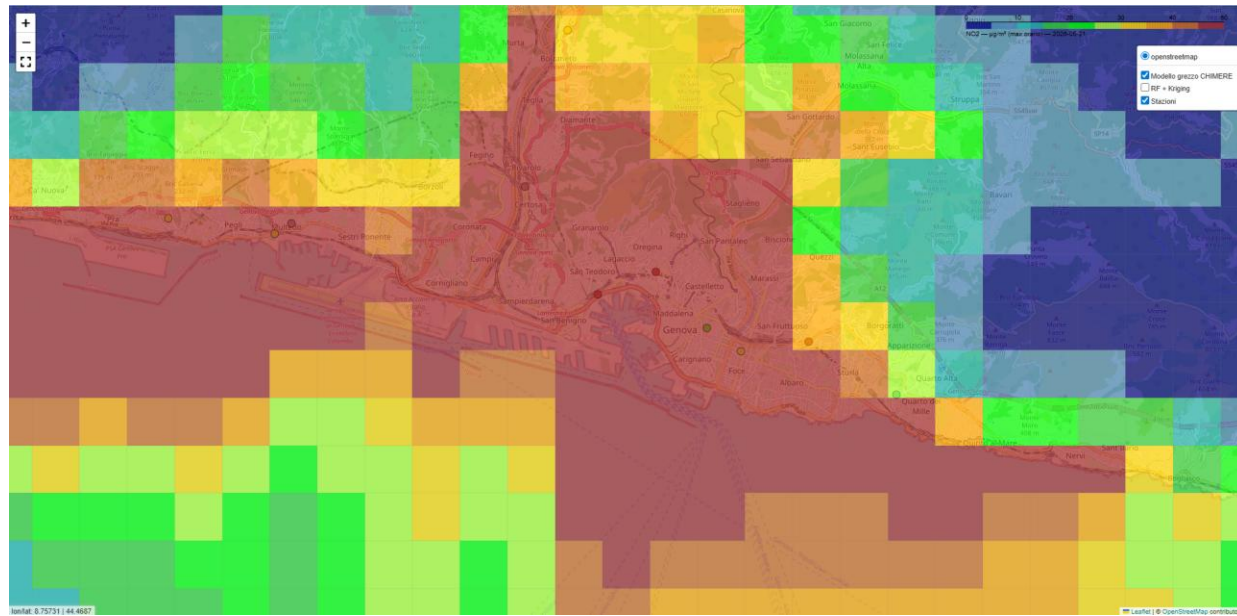
Risultati NO₂



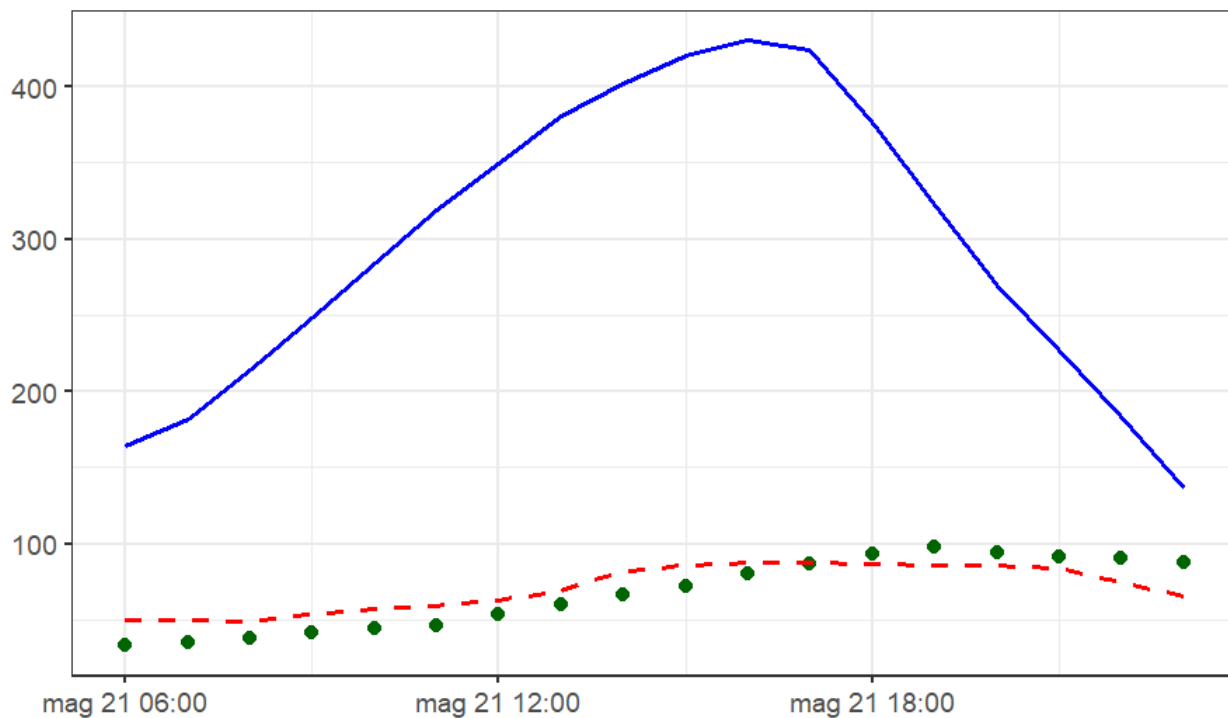
Risultati NO₂



Risultati NO₂

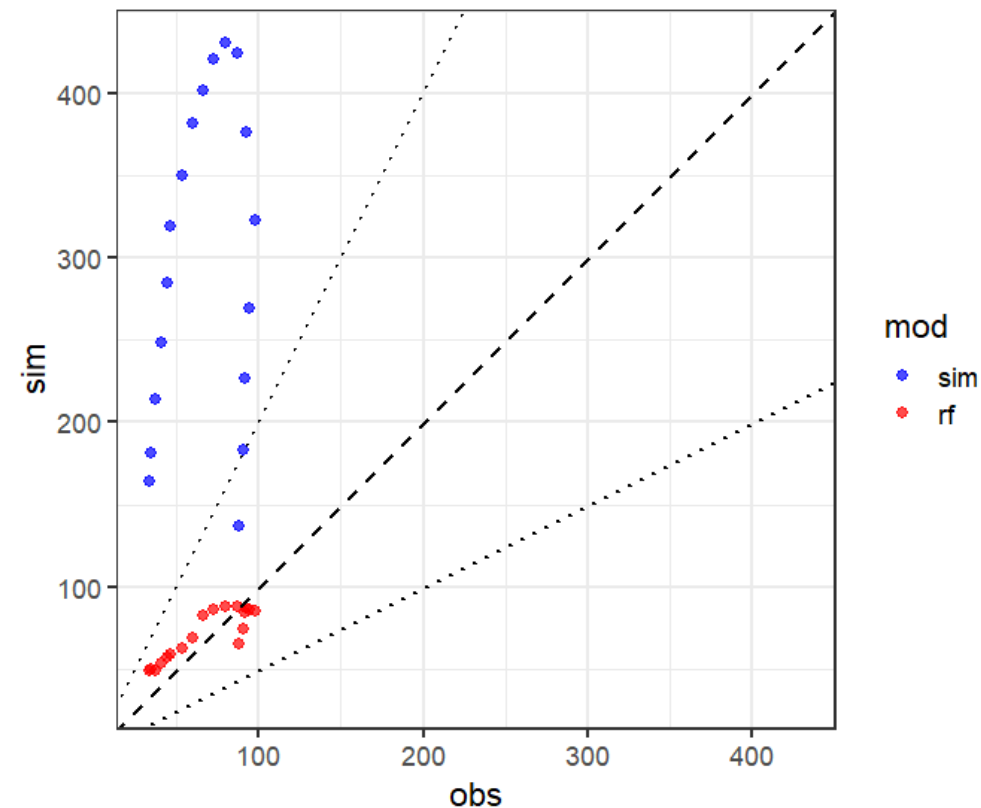


IT0858A Genova - Quarto Urbana

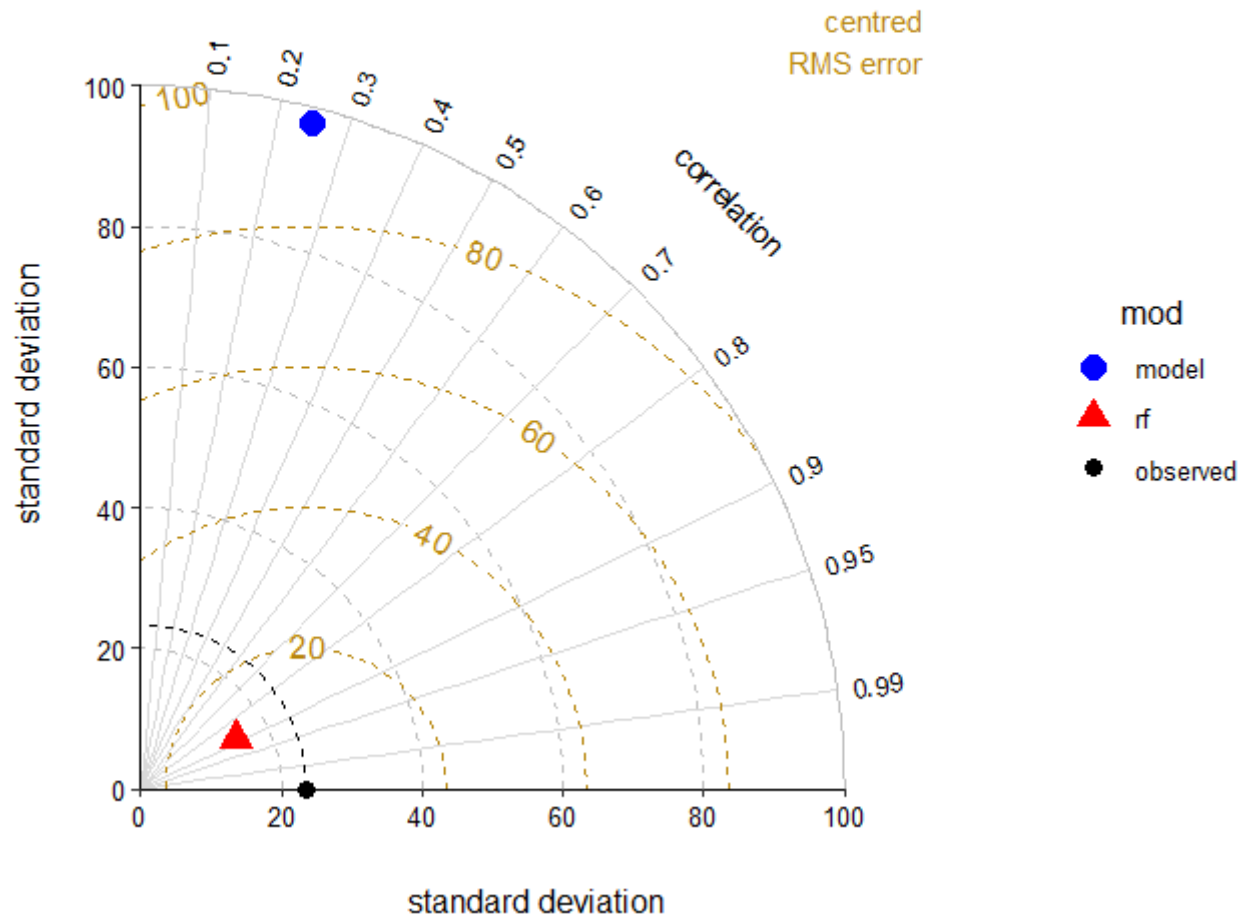


— sim
● obs
- - rf

IT0858A Genova - Quarto Urbana

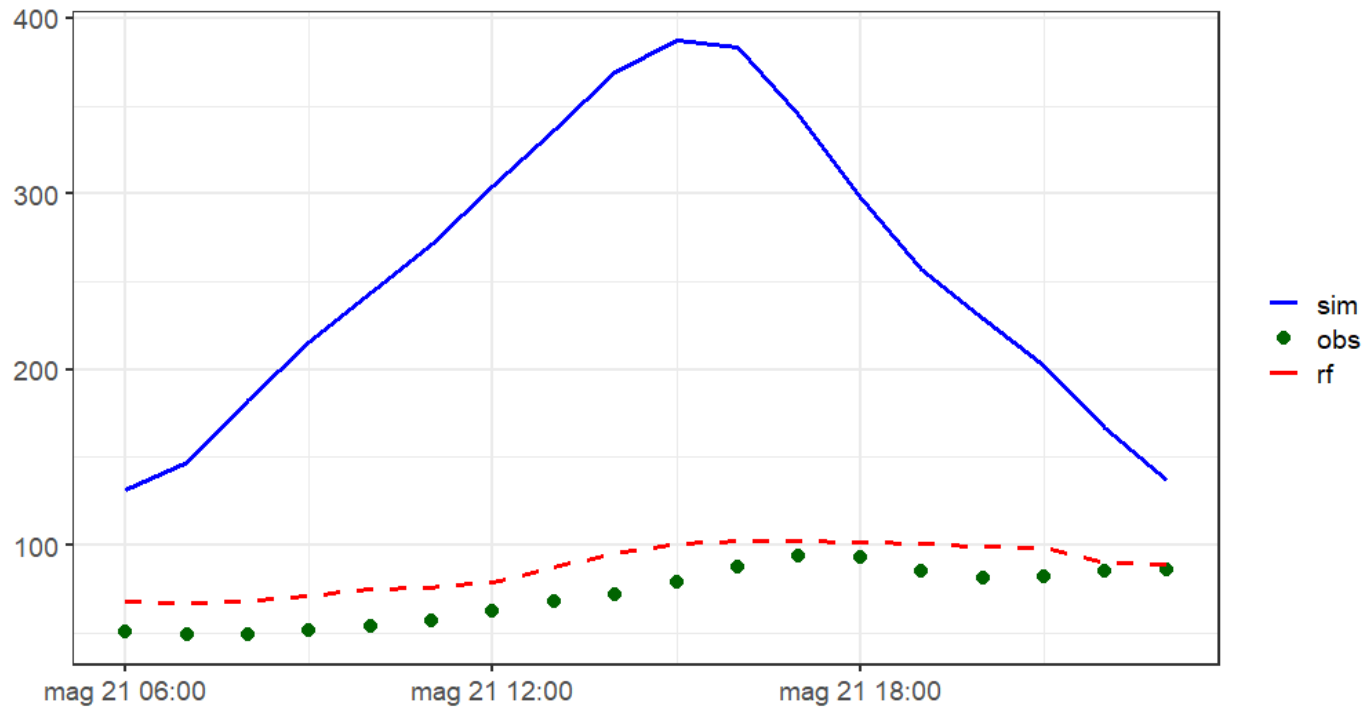


IT0858A Genova - Quarto Urbana

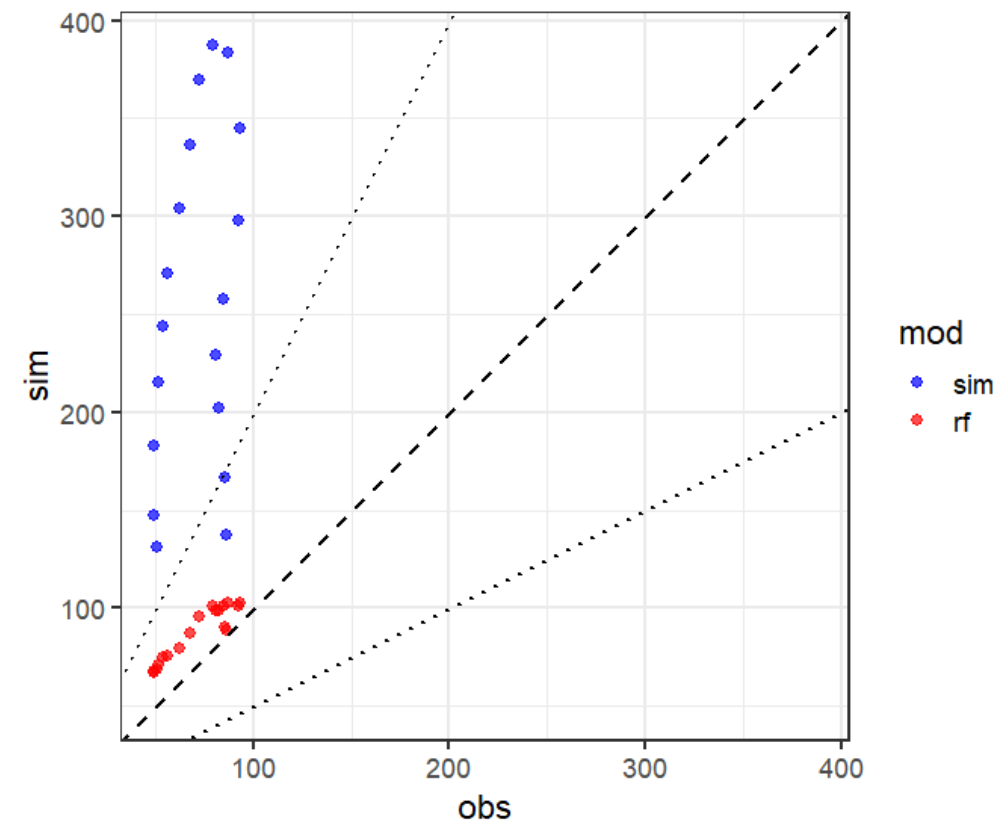


	model	rf
n	18	18
FAC2	0.06	1.00
MB	228.97	3.67
MGE	228.97	11.58
NMB	3.40	0.05
NMGE	3.40	0.17
RMSE	246.74	12.47
r	0.25	0.89

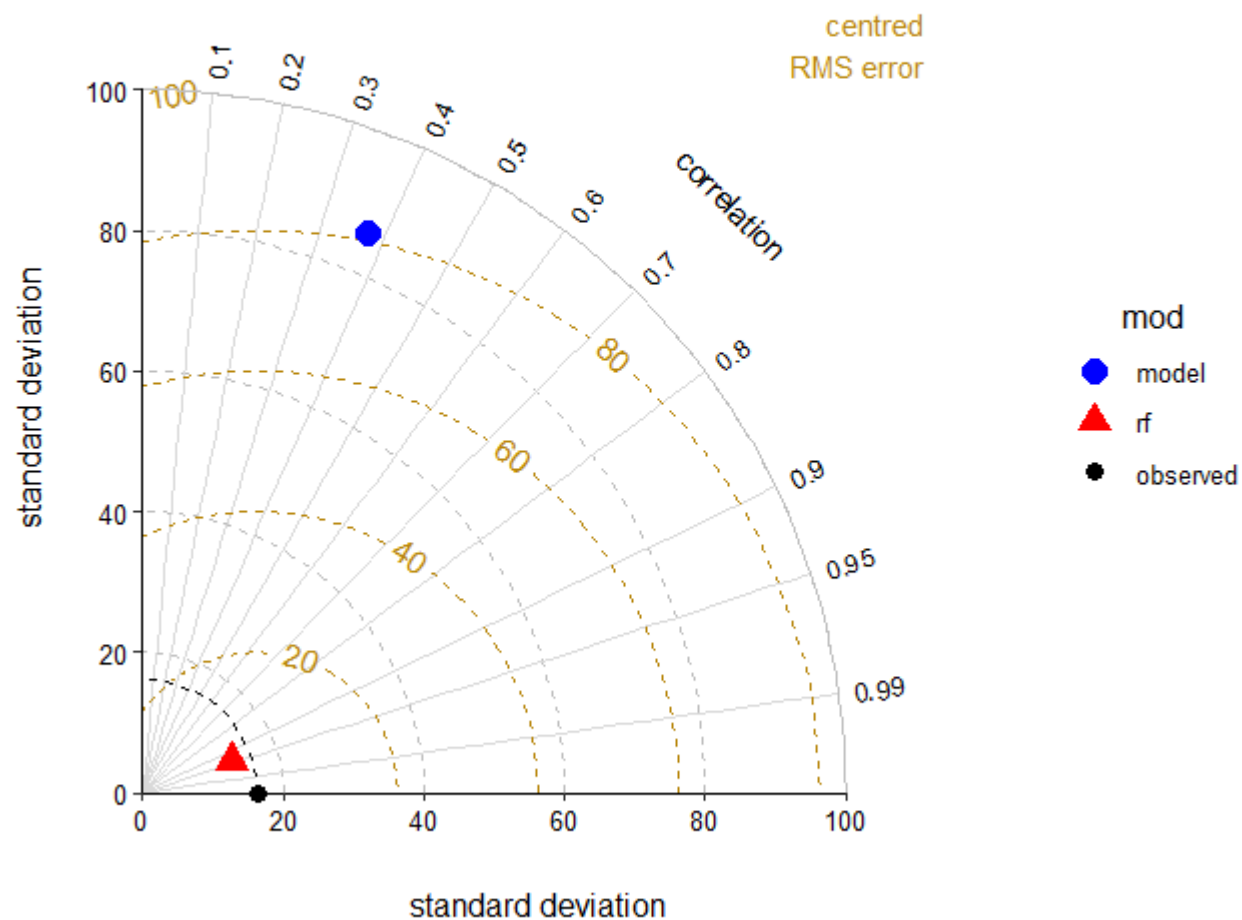
IT0856A Genova - Acquasola Urbana



IT0856A Genova - Acquasola Urbana

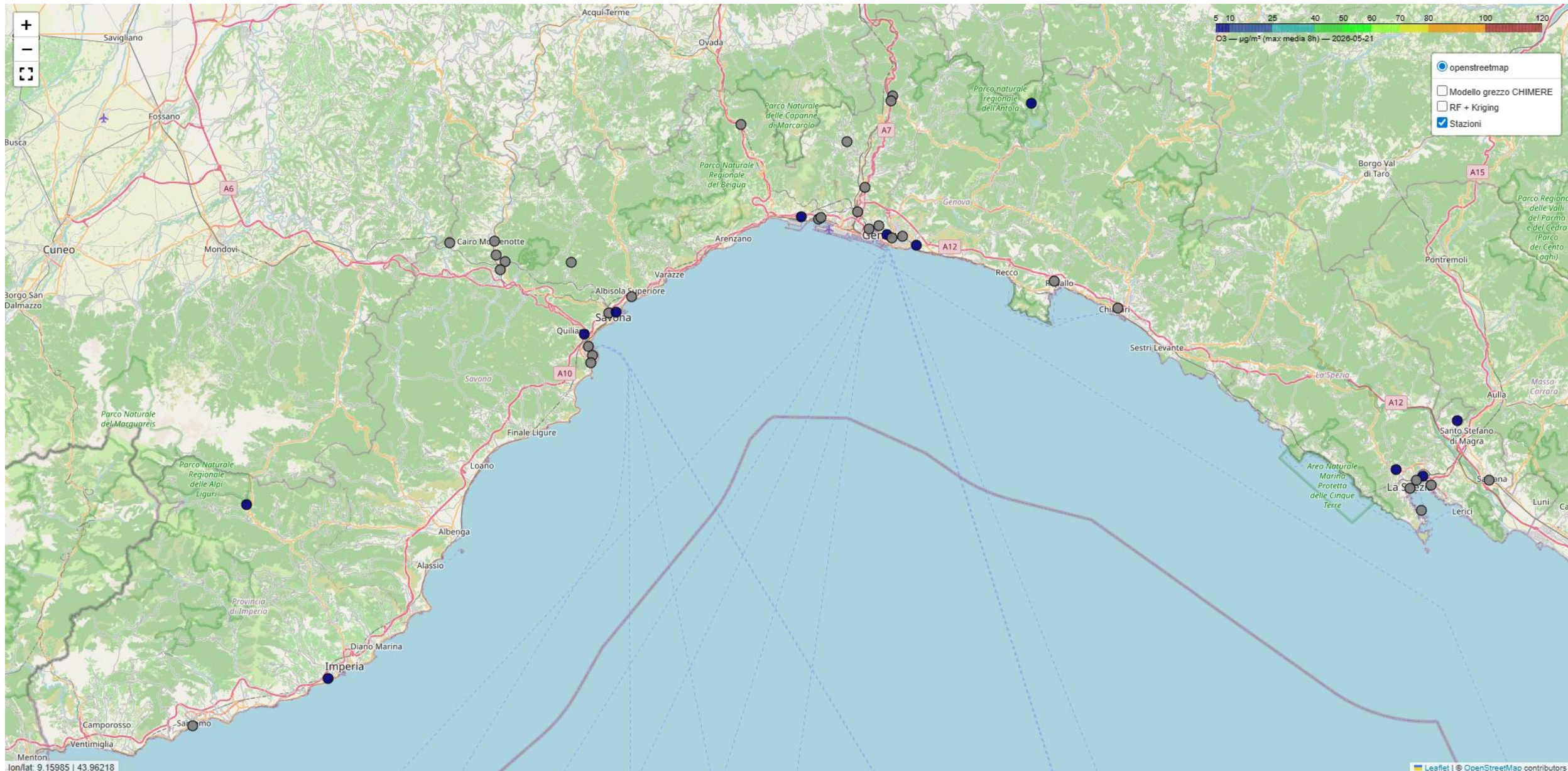


IT0856A Genova - Acquasola Urbana

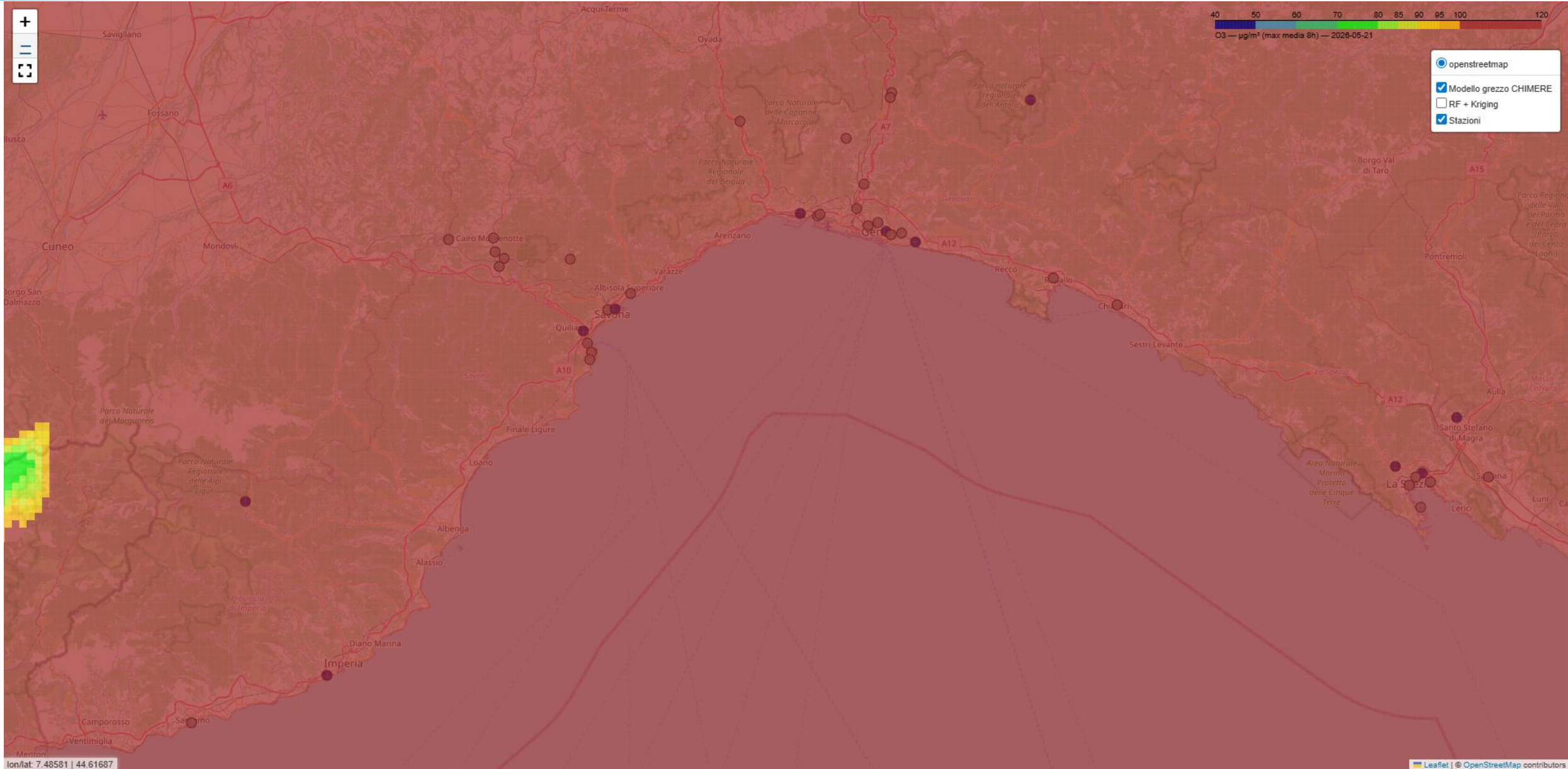


	model	rf
n	18	18
FAC2	0.11	1.00
MB	184.55	15.96
MGE	184.55	15.96
NMB	2.59	0.22
NMGE	2.59	0.22
RMSE	200.70	16.93
r	0.37	0.94

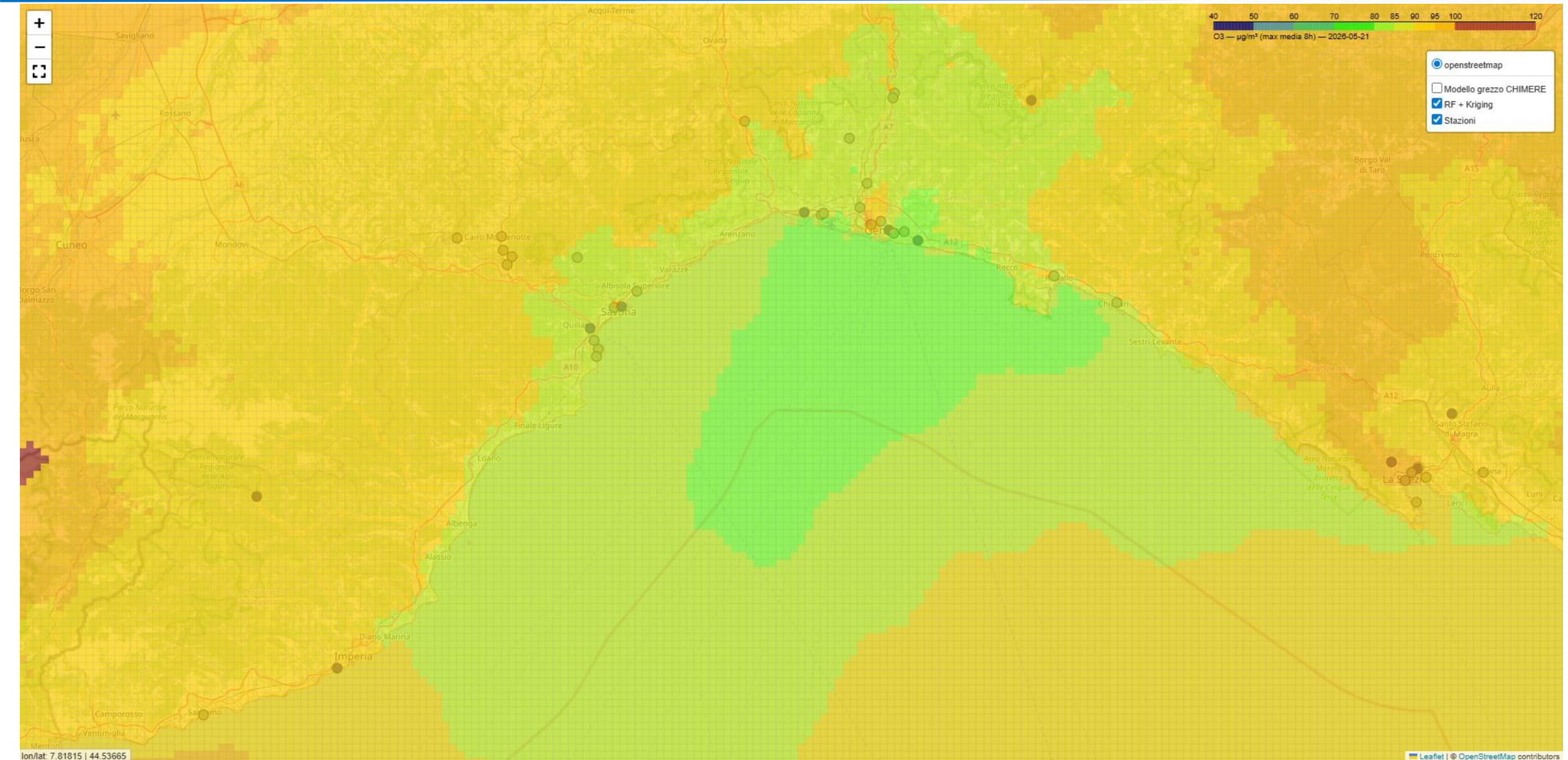
Risultati O₃



Risultati O₃

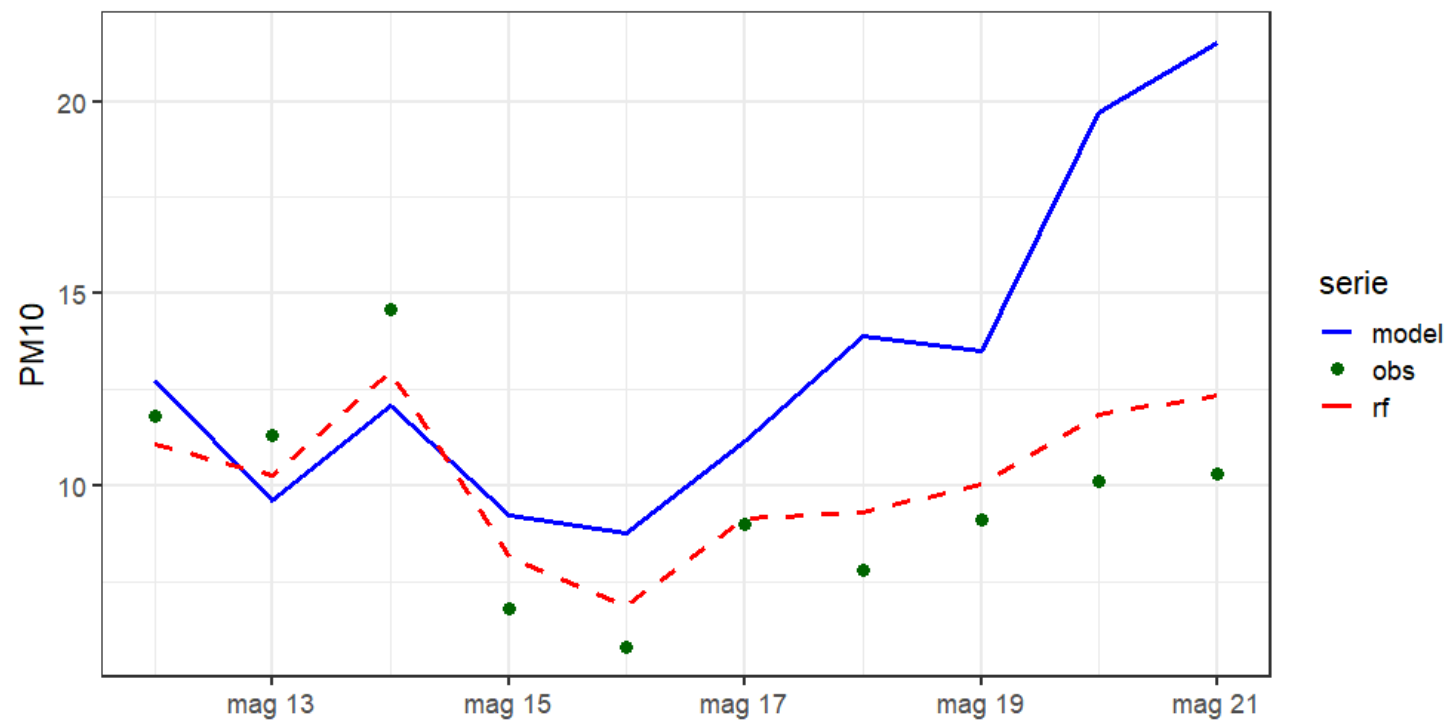


Risultati O₃

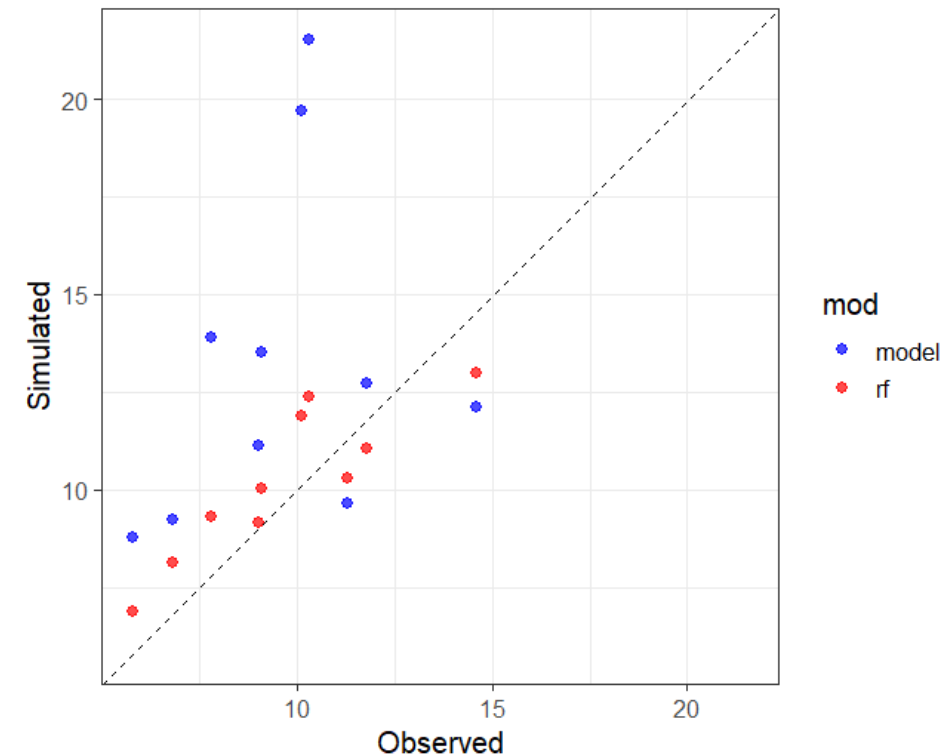


Campomorone GE - Campora

PM10 Time Series

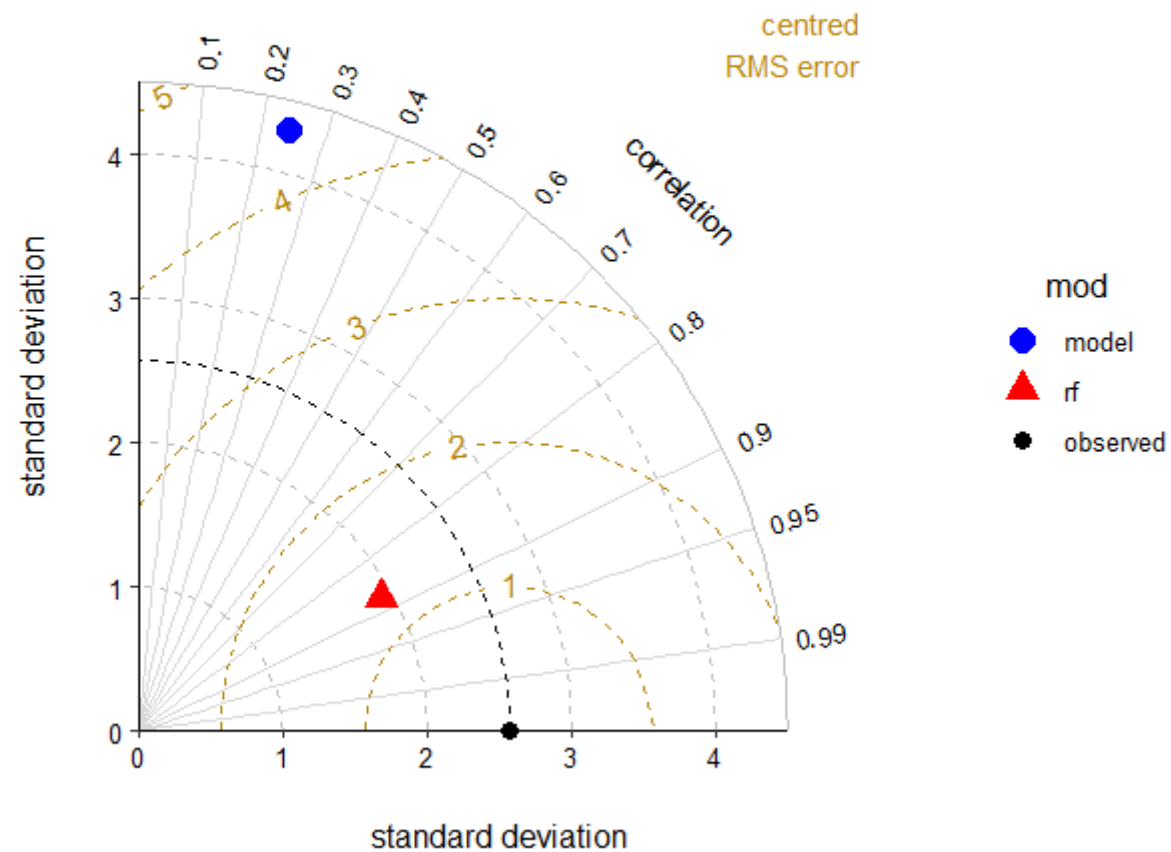


Obs vs Sim PM10



Campomorone GE - Campora

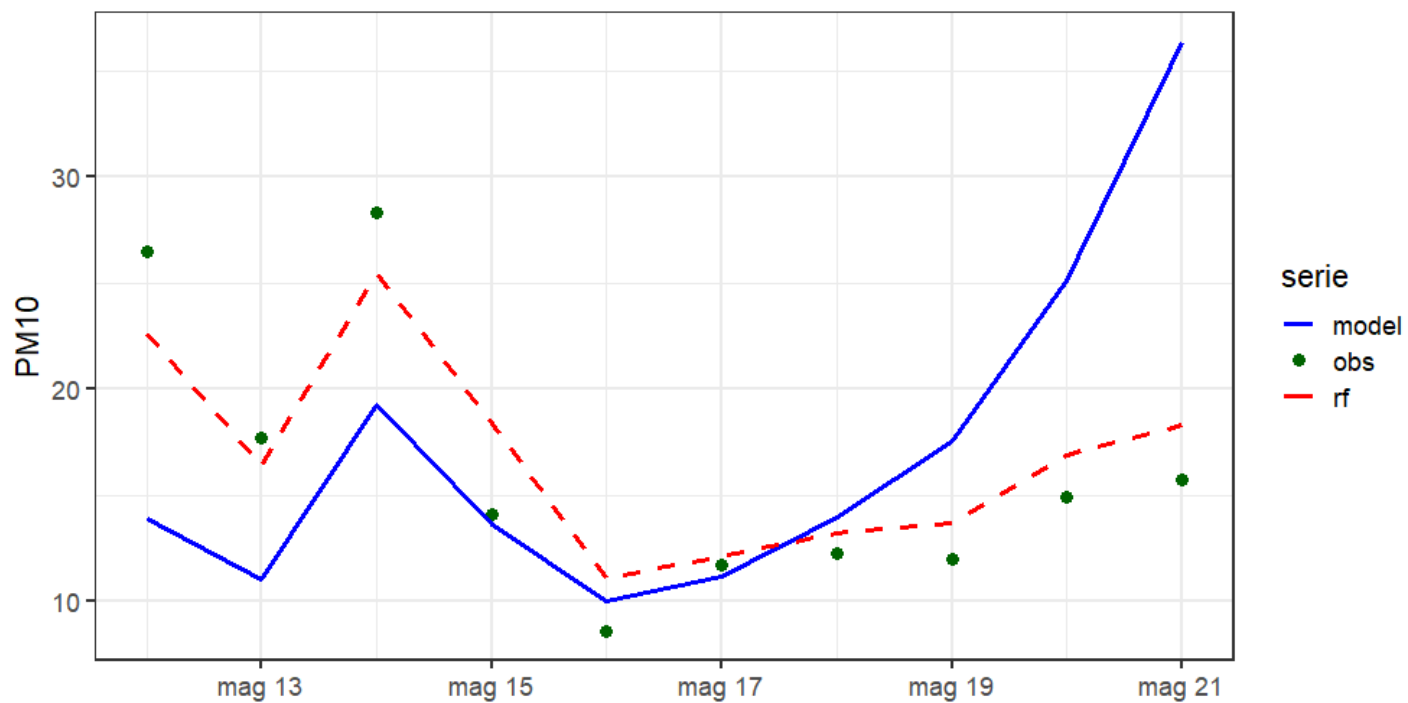
PM₁₀ Taylor Diagram



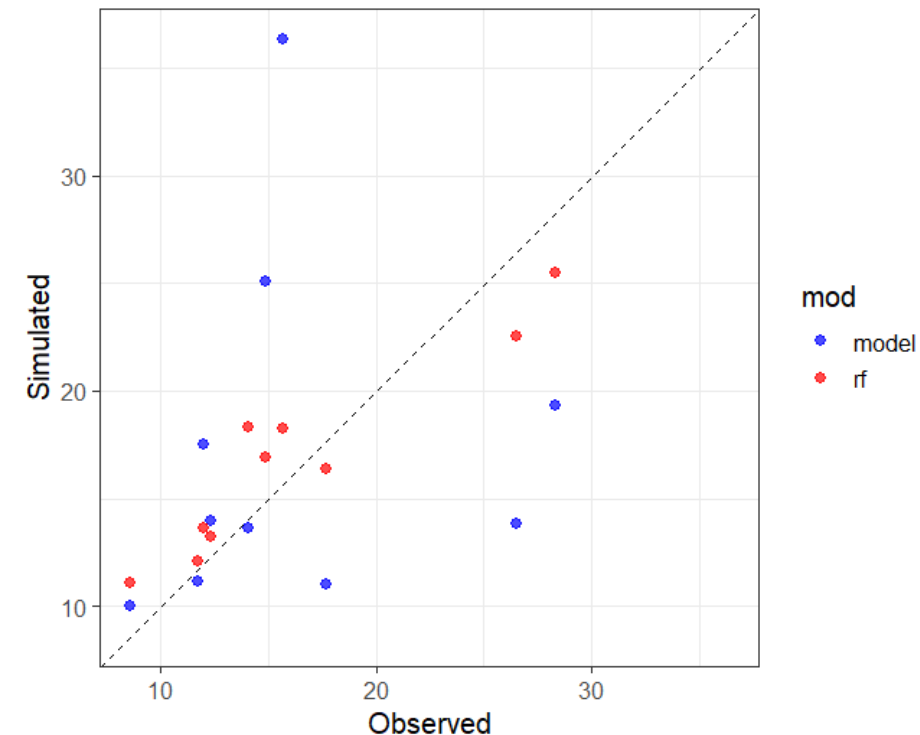
	MB	RMSE	r
model	3.57	5.52	0.24
rf	0.55	1.33	0.88

Genova - Buenos Aires

PM10 Time Series

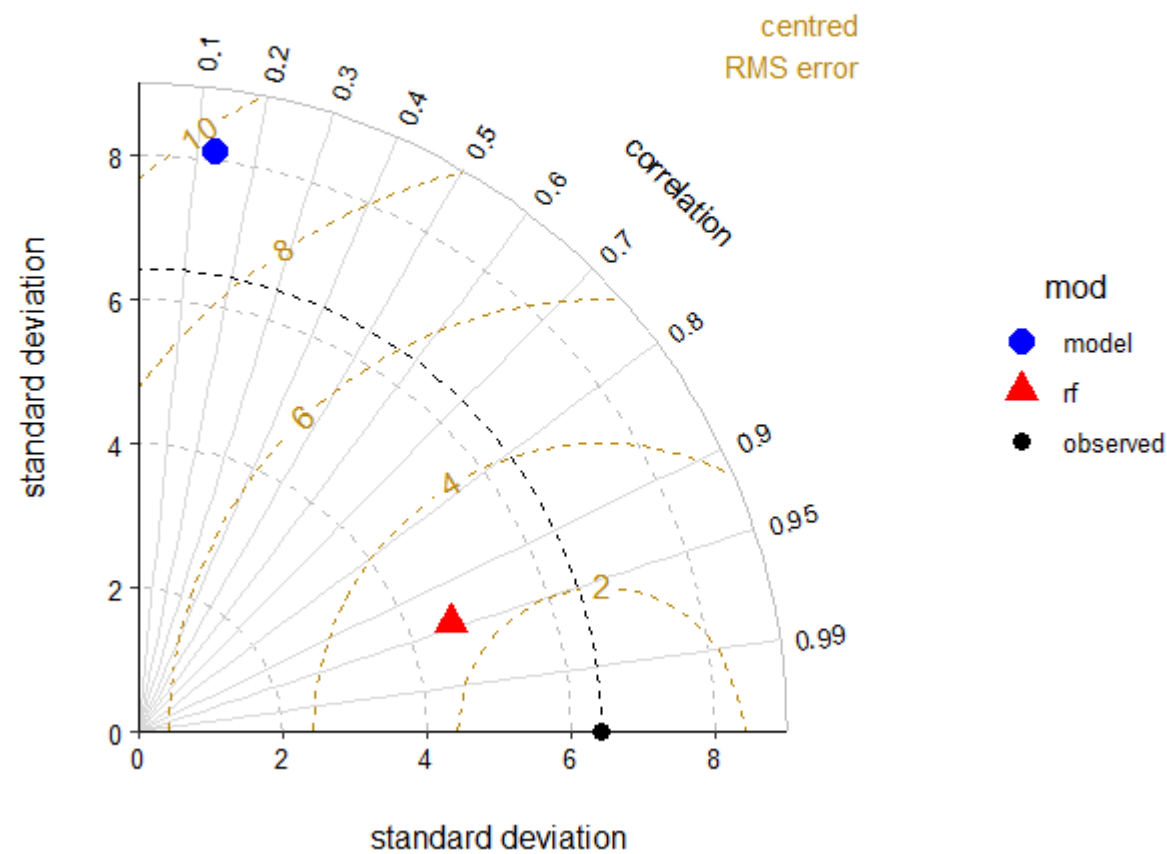


Obs vs Sim PM10



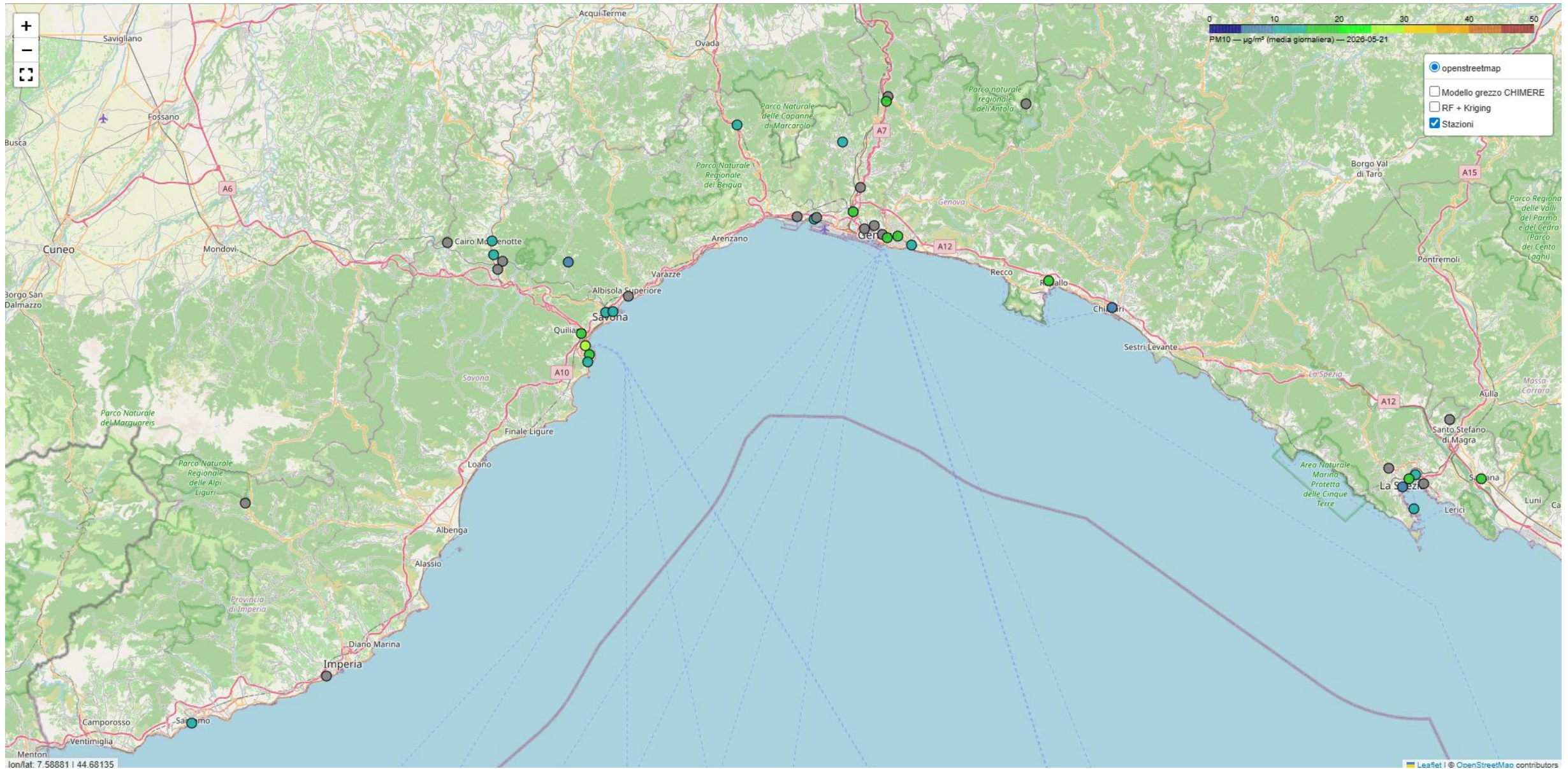
Genova - Buenos Aires

PM₁₀ Taylor Diagram

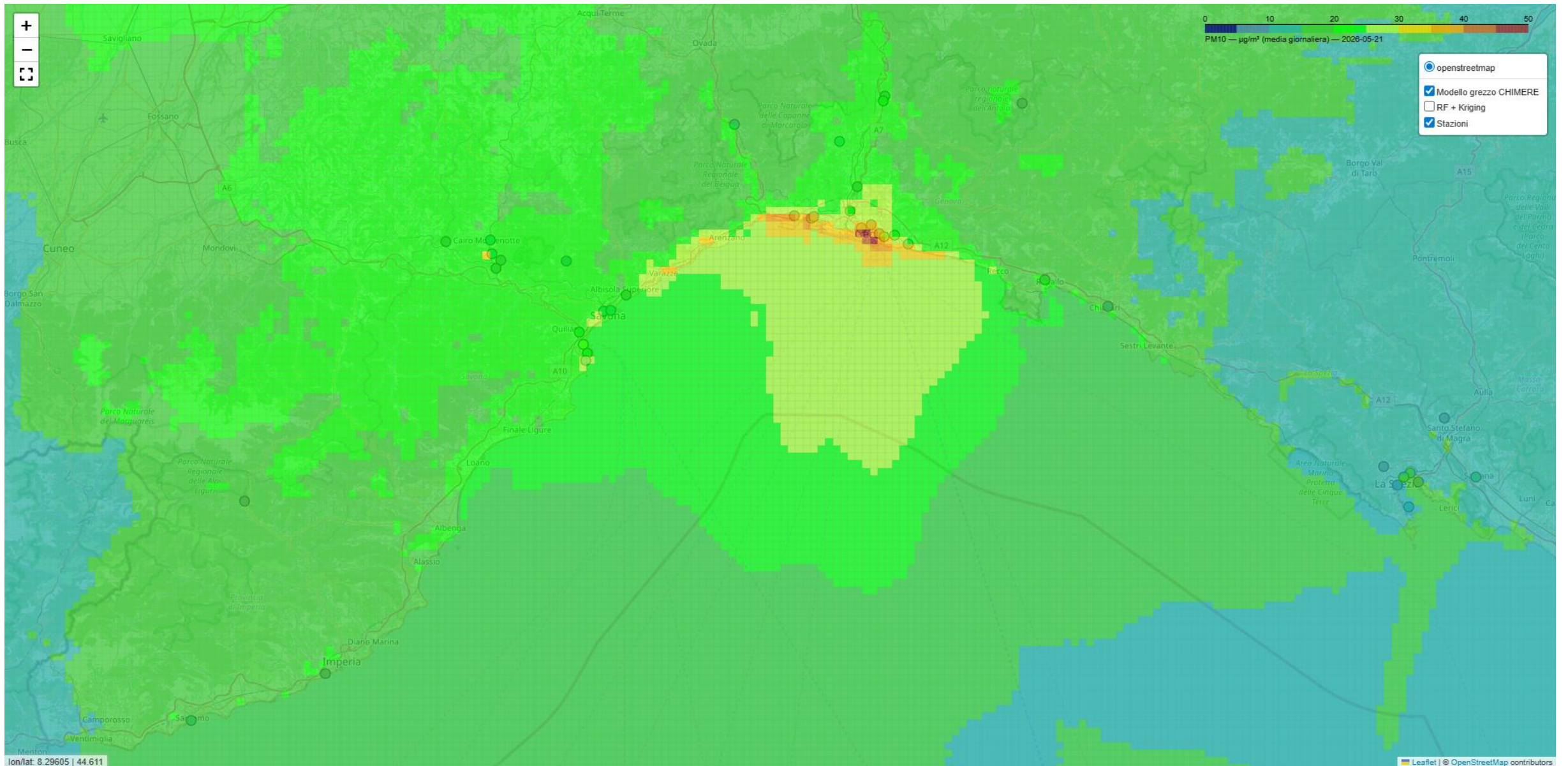


	MB	RMSE	r
model	1.02	9.23	0.13
rf	0.63	2.53	0.94

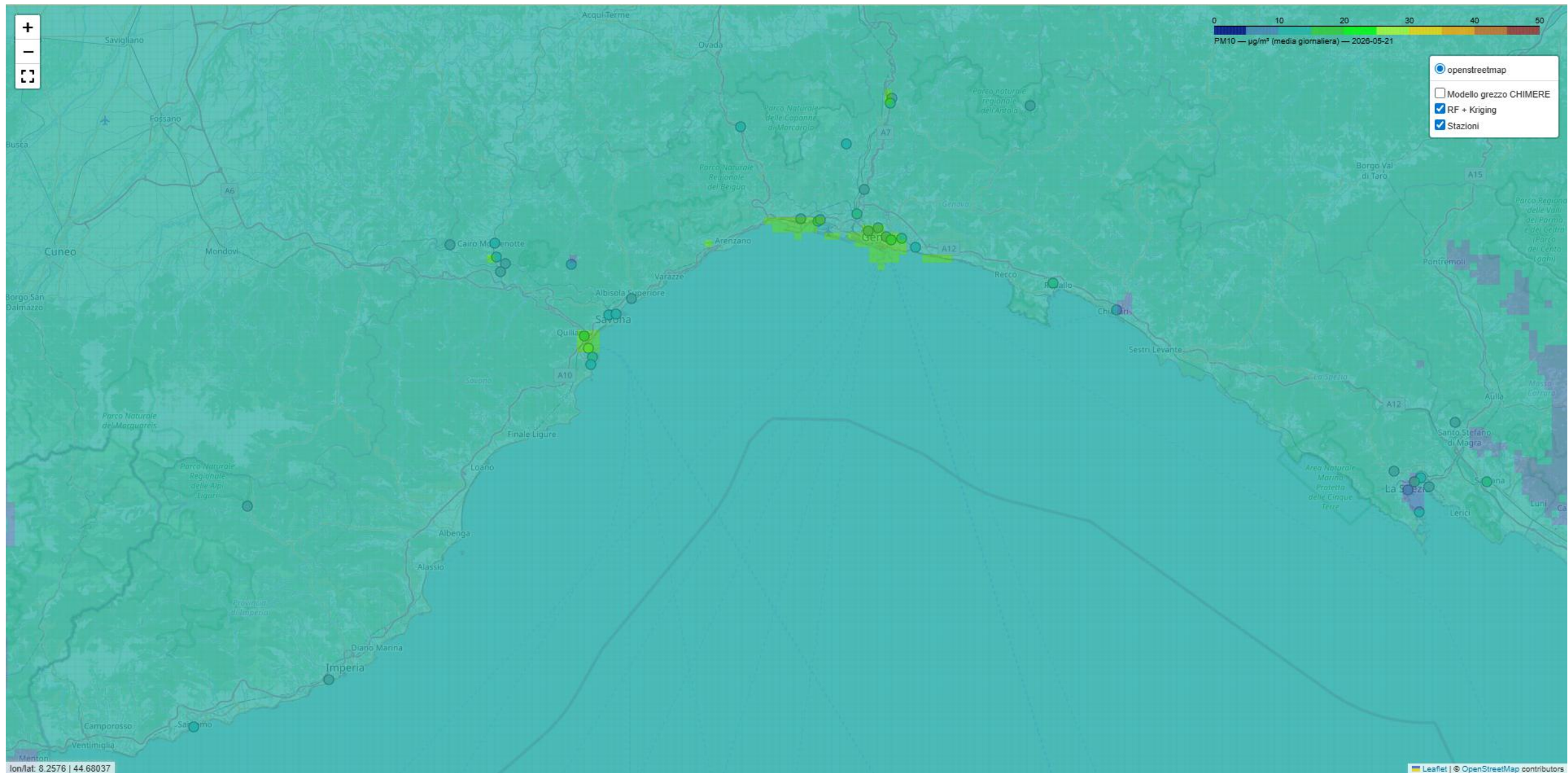
Risultati PM₁₀



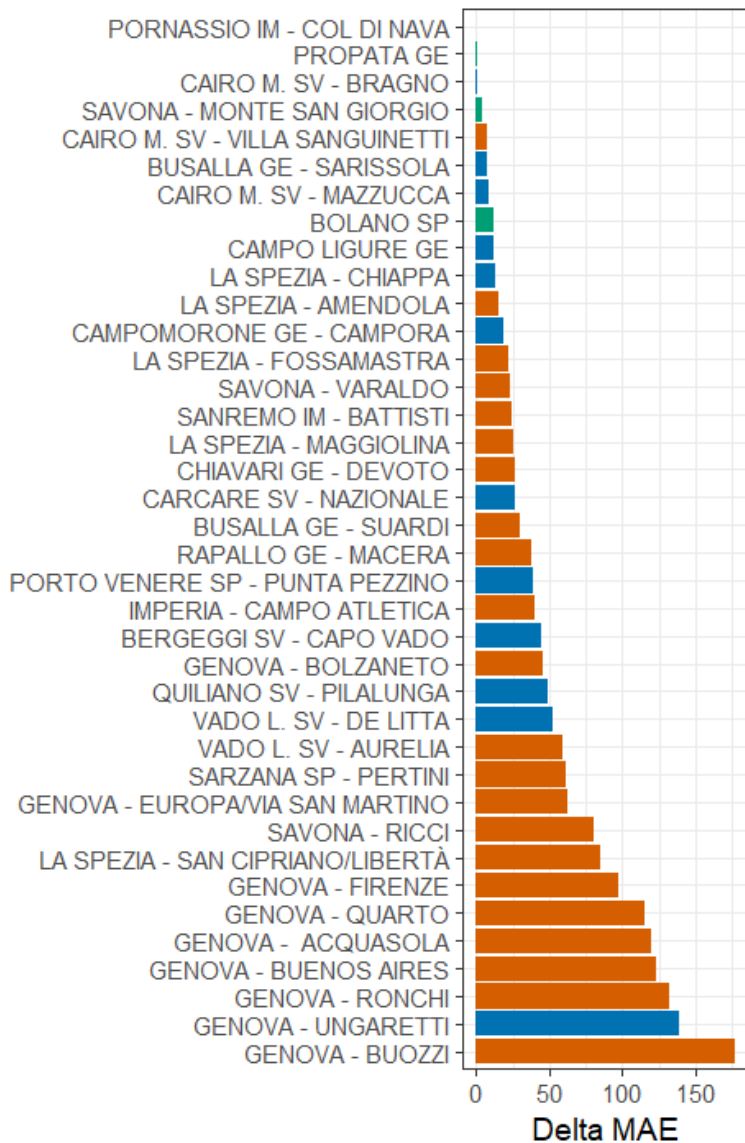
Risultati PM₁₀



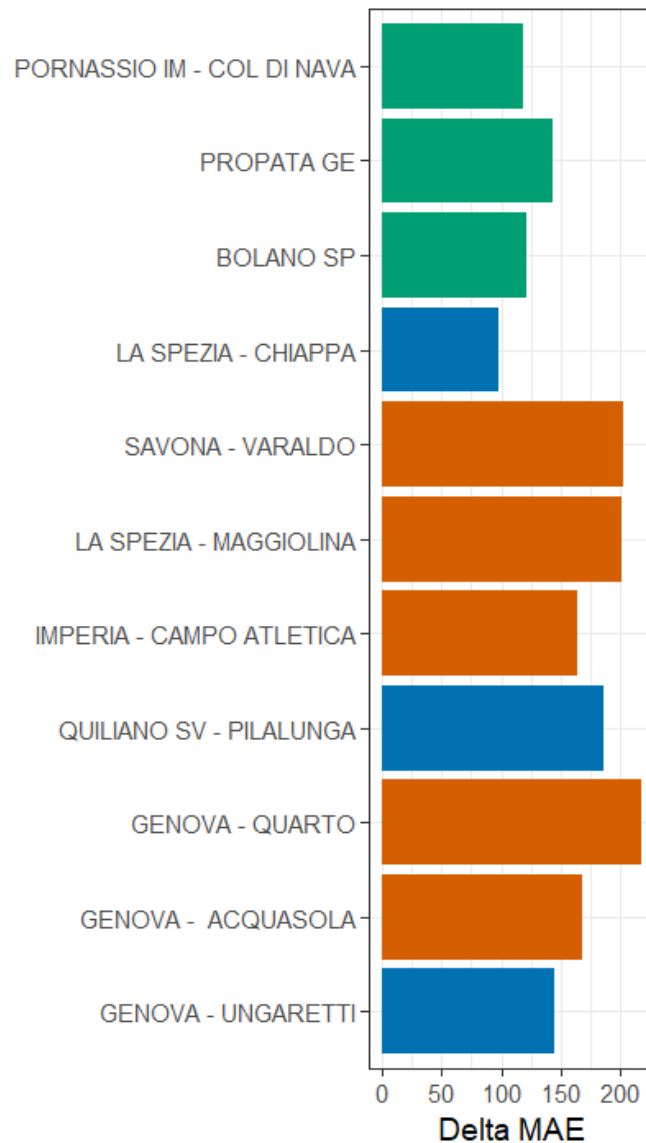
Risultati PM₁₀



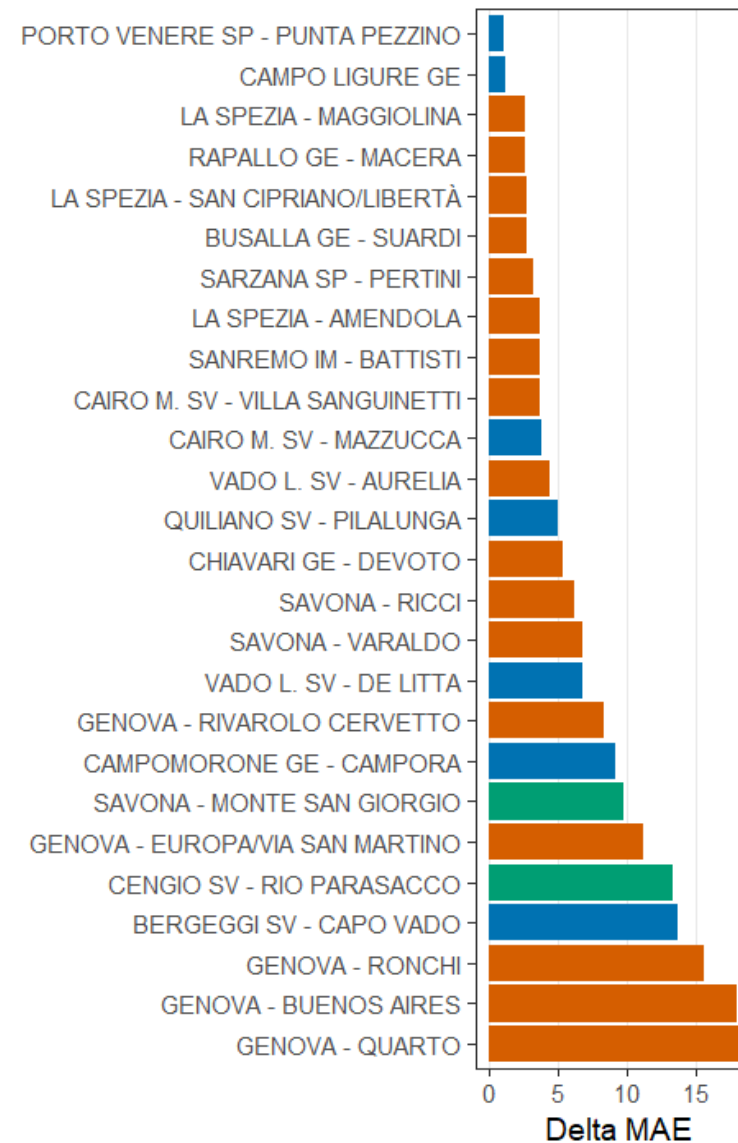
NO2



O3



PM10



area_classification

Rurale
Suburbana
Urbana

- Il Machine Learning può migliorare le previsioni di qualità dell'aria e ridurre il bias dei modelli fisici.
- Va usato con cautela: senza una validazione rigorosa si rischiano leakage temporali e spaziali e, quindi, risultati troppo ottimistici.
- Più complessità (es. reti neurali) non significa automaticamente più qualità: spesso si riduce l'interpretabilità fisica del sistema e può essere difficile distinguere cosa arriva dal modello fisico e cosa dal modello statistico

In sintesi il Machine Learning ...

- ... aiuta, ma non “sostituisce” la fisica!!!
- ... se usato male, può produrre risultati con indici di prestazione molto buoni, ma che non sono necessariamente affidabili in modalità previsionale.