PM10 and PM2.5 in an urban area of the Adriatic coast: Trends and forecast using a recursive neural network model

Fabio Biancofiore[1], Marcella Busilacchio[1], Marco Verdecchia[1], Barbara Tomassetti[1], Eleonora Aruffo[1], Sebastiano Bianco[2], Sinibaldo Di Tommaso[2], Carlo Colangeli[2] and Piero Di Carlo[1]

[1]Centro di eccellenza CETEMPS - Dipartimento di Scienze Fisiche e Chimiche Universita' dell'Aquila, Via Vetoio, 67010 Coppito, L'Aquila,

[2]ARTA, Agenzia Regionale per l'Ambiente, viale Marconi, 65100, Pescara

Email: piero.dicarlo@aquila.infn.it

Sommario

- Modelli utilizzati (MLRM, ANNF, ANNE) e loro architettura
- Analisi di 3 anni di dati in un’area urbana costiera
- Confronto tra i 3 modelli
- Performance dei modelli nella previsione di PM10 e PM2.5
- Conclusioni

Multiple linear regression model (MLRM)

Modello regressivo costruito con la tecnica Stepwise per escludere i parametri ridondanti e massimizzare la correlazione:

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \ldots + b_k X_{ik} + \varepsilon_i$$

dove: i sono le osservazioni, Y_i sono le variabili da simulare, calcolate come combinazione lineare dei coefficienti b_i e un set di variabili indipendenti X (i predittori) ed ε_i i residui.

In genere i predittori utilizzati sono i parametri meteo, ma a seconda dei casi anche modelli ibridi meteo + composti atmosferici

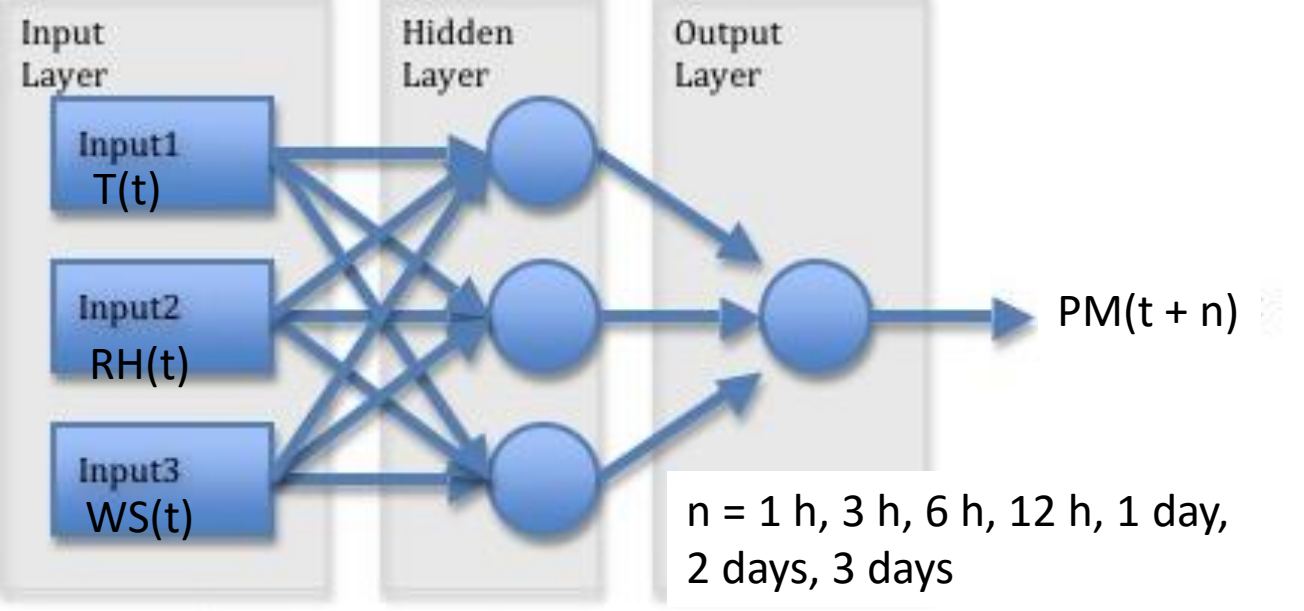
Sviluppato e testato con l'ozono:

In un sito di background: proxy oltre ai parametri meteo usato il radon (Di Carlo, P, et al, *J. Geoph. Res*., 2007)

Nel caso di plume dovuti ad incendio: proxy parametri meteo + PM (Di Carlo et al., *Atmo. Pollution Research, 2015)*

Multi-layer Neural Network feed-forward (ANNF)

Feed-forward: Without the recurrent architecture (ANNF)



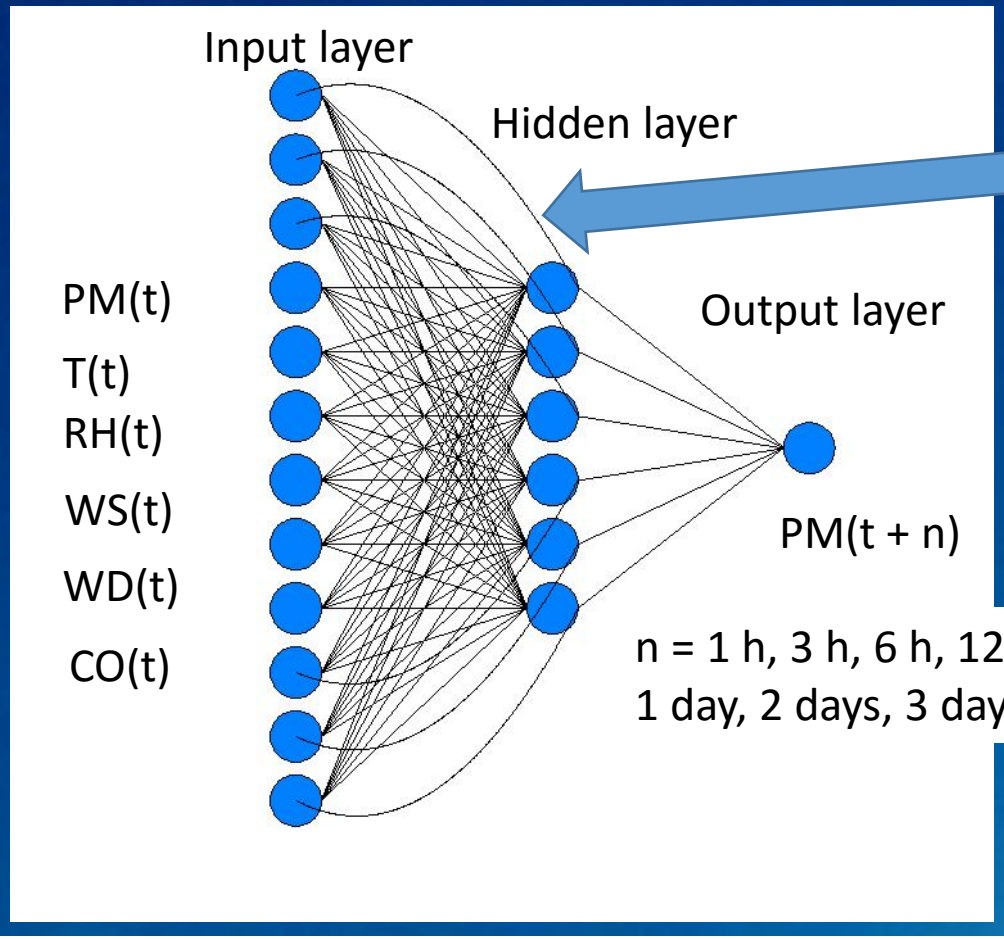
The phase of calibration is carried out by minimizing the error function E at each training step the error function E is calculated as the mean squared differences between the observed and predicted PM concentrations

Sviluppato e testato con l'ozono:

Biancofiore F., et al., *Science of the Total Environment*, 2015.
Putero et al. *ACP*, 2015.

Multi-layer Neural Network Recurrent (ANNE)

Uses the Elman recurrent architecture

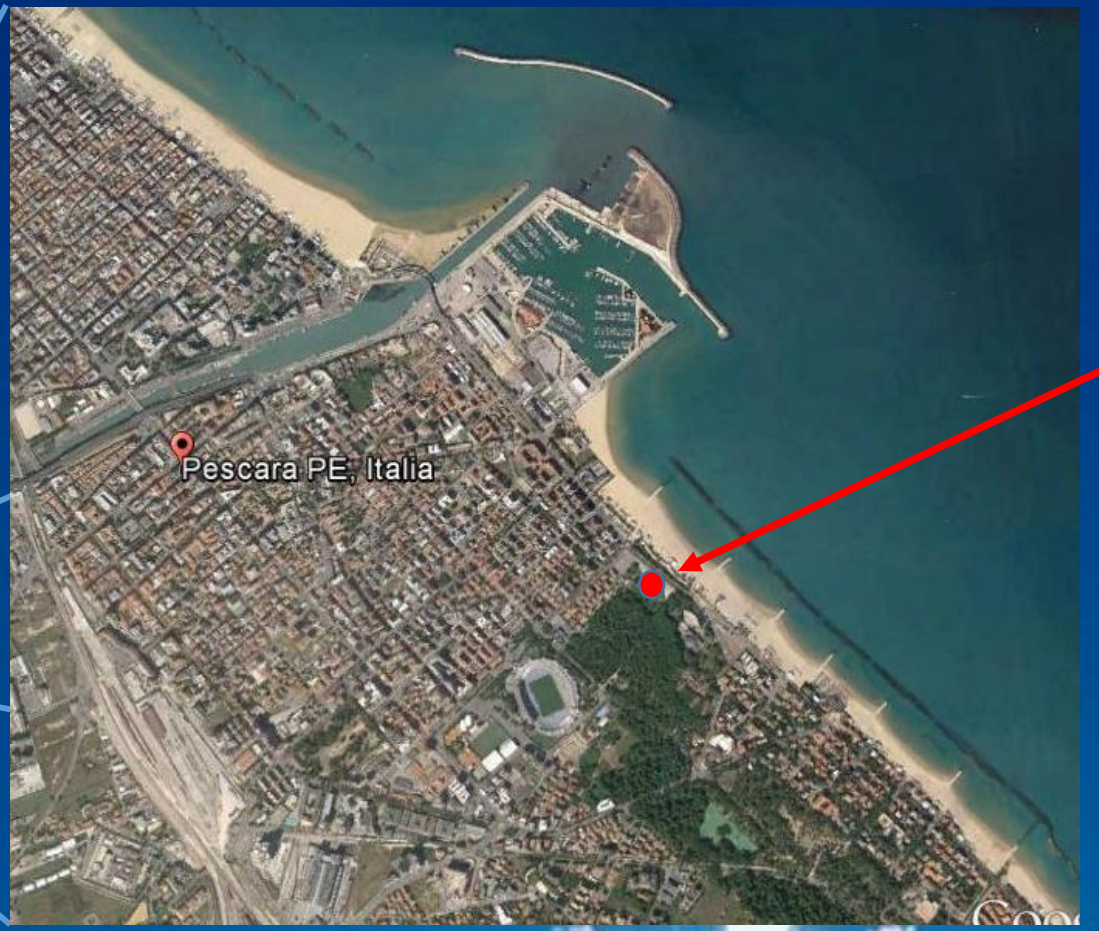


Gli output non vengono trasferiti solo ai nodi degli strati successivi ma anche agli strati precedenti

Sviluppato e testato con l'ozono:

Biancofiore F., et al., *Science of the Total Environment*, 2015.
Putero et al. *ACP*, 2015.

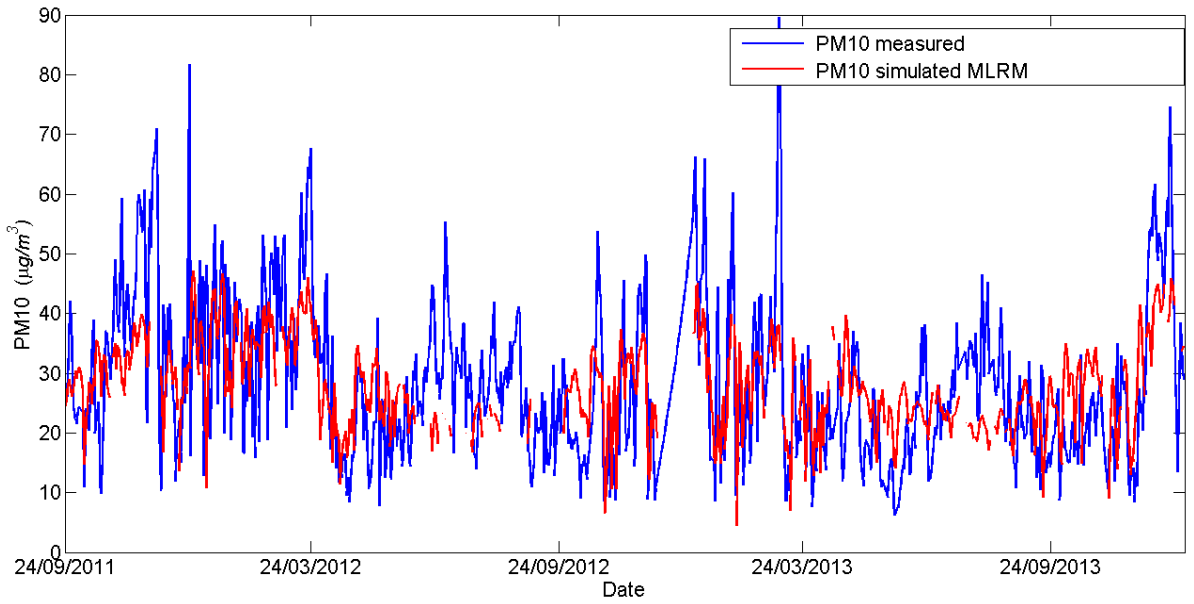
Sito osservativo

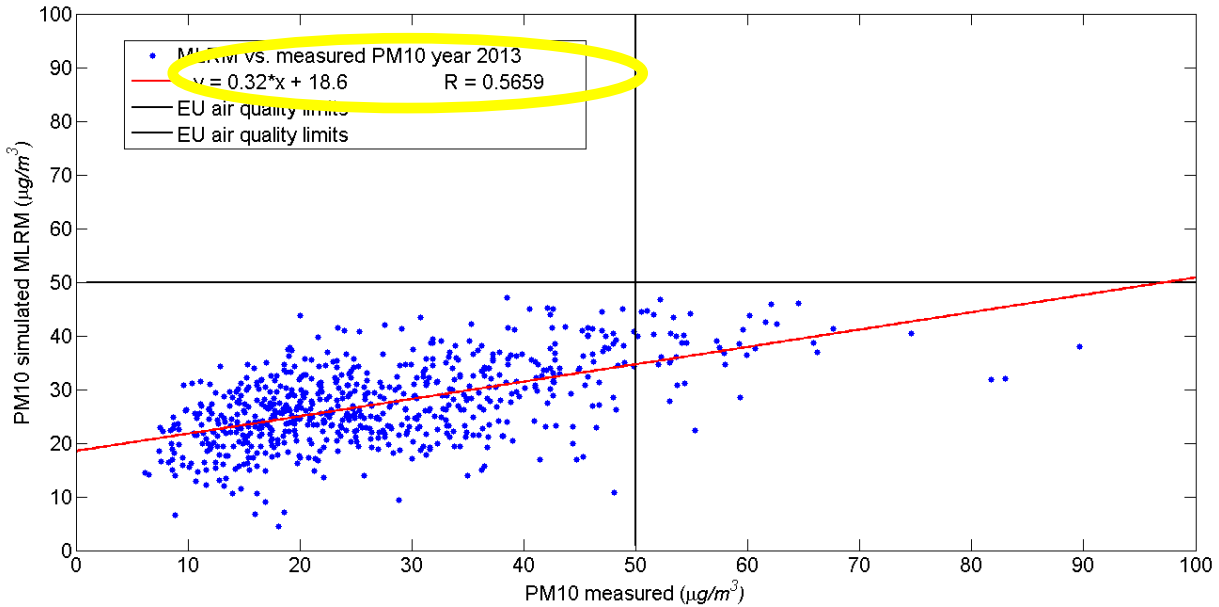


Stazione a circa 50 m dalla spiaggia nella pineta d'annunziana (classificata di background urbano)

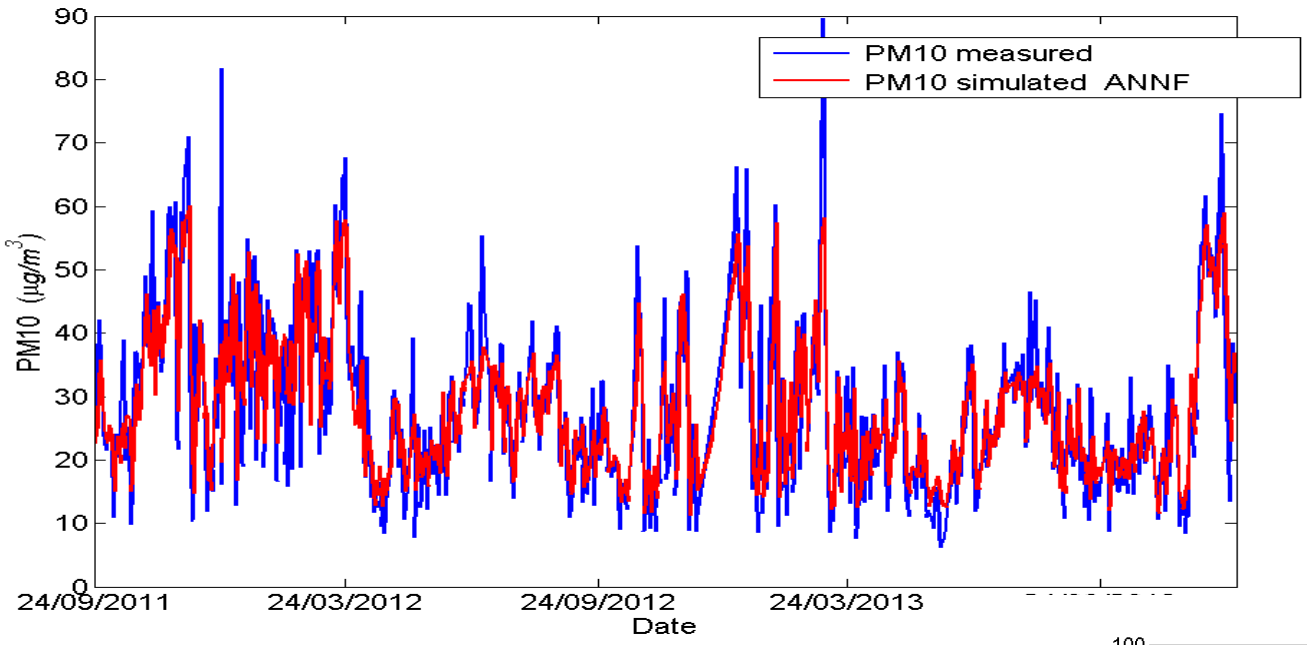
Stazione osservativa della rete Regionale ARTA (attiva da 20 anni: meteo, O3, CO, CO2, NOx, VOC, PM10, PM2.5)

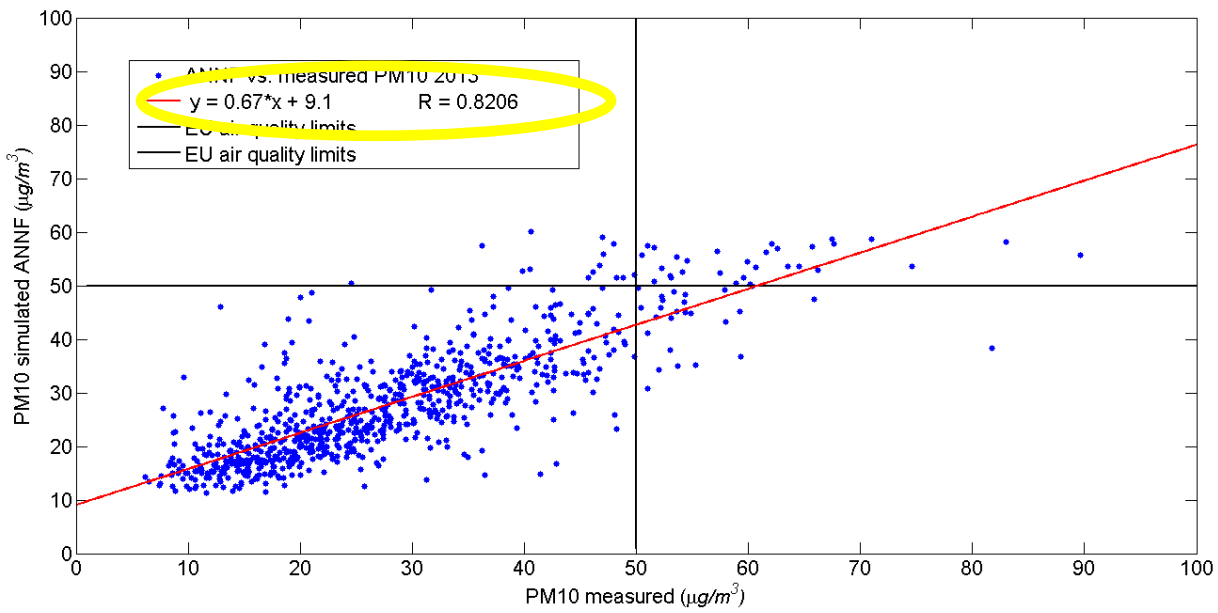
Multiple linear regression model (MLRM) vs observations



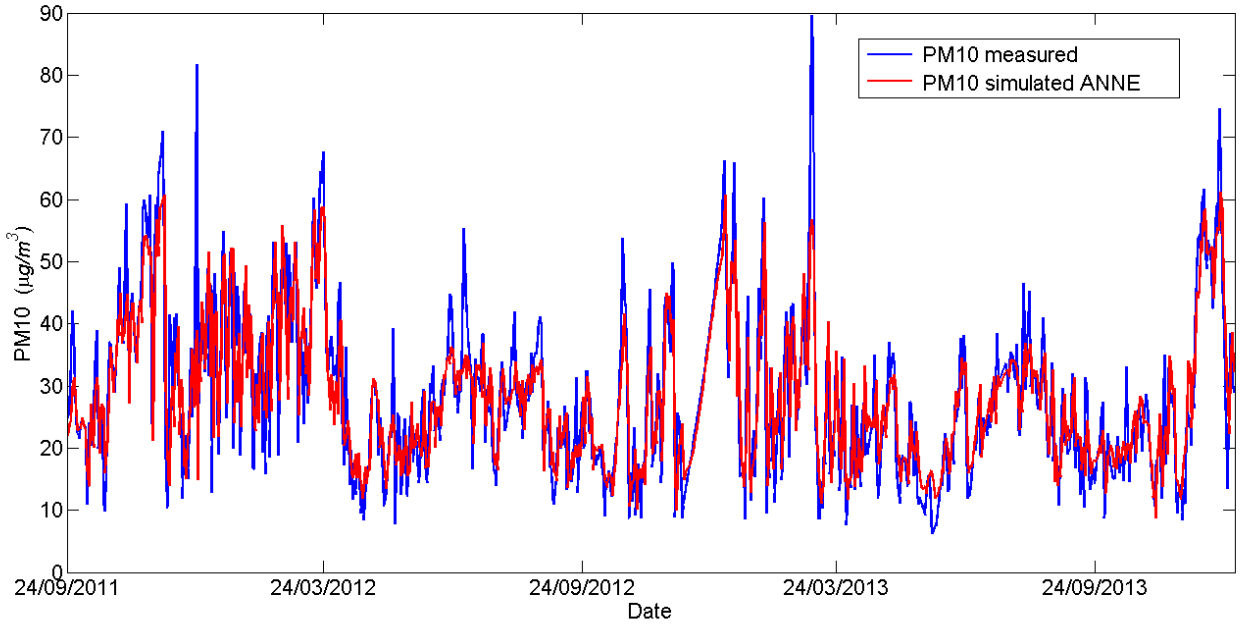


Multi-layer Neural Network feed-forward (ANNF) vs observations

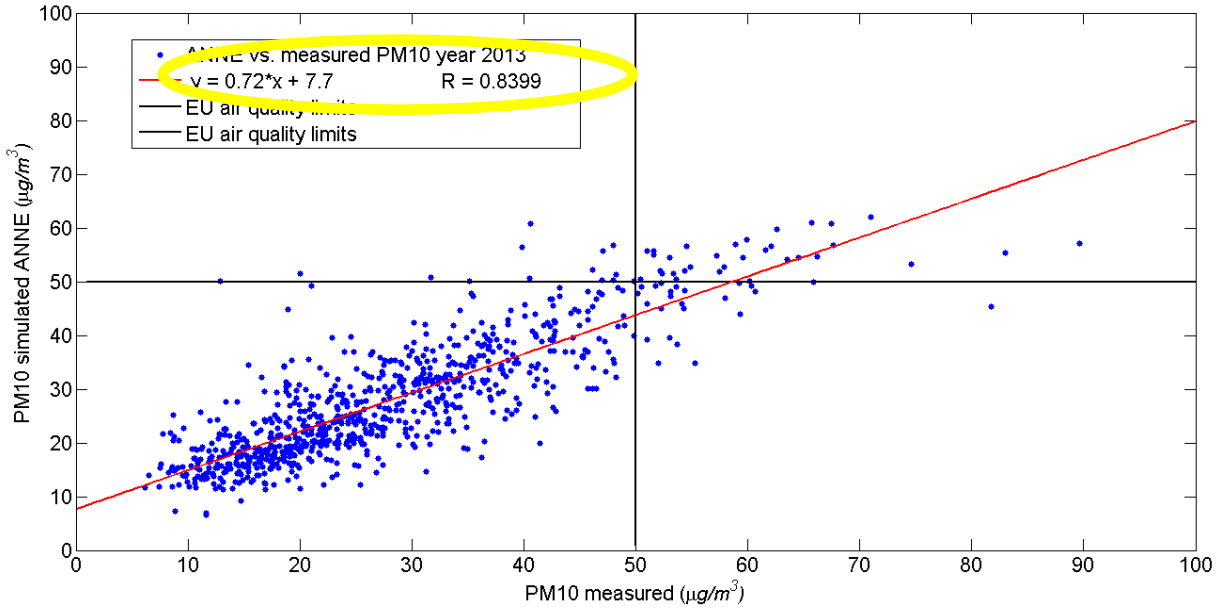




Multi-layer Neural Network recurrent (ANNE) vs observations

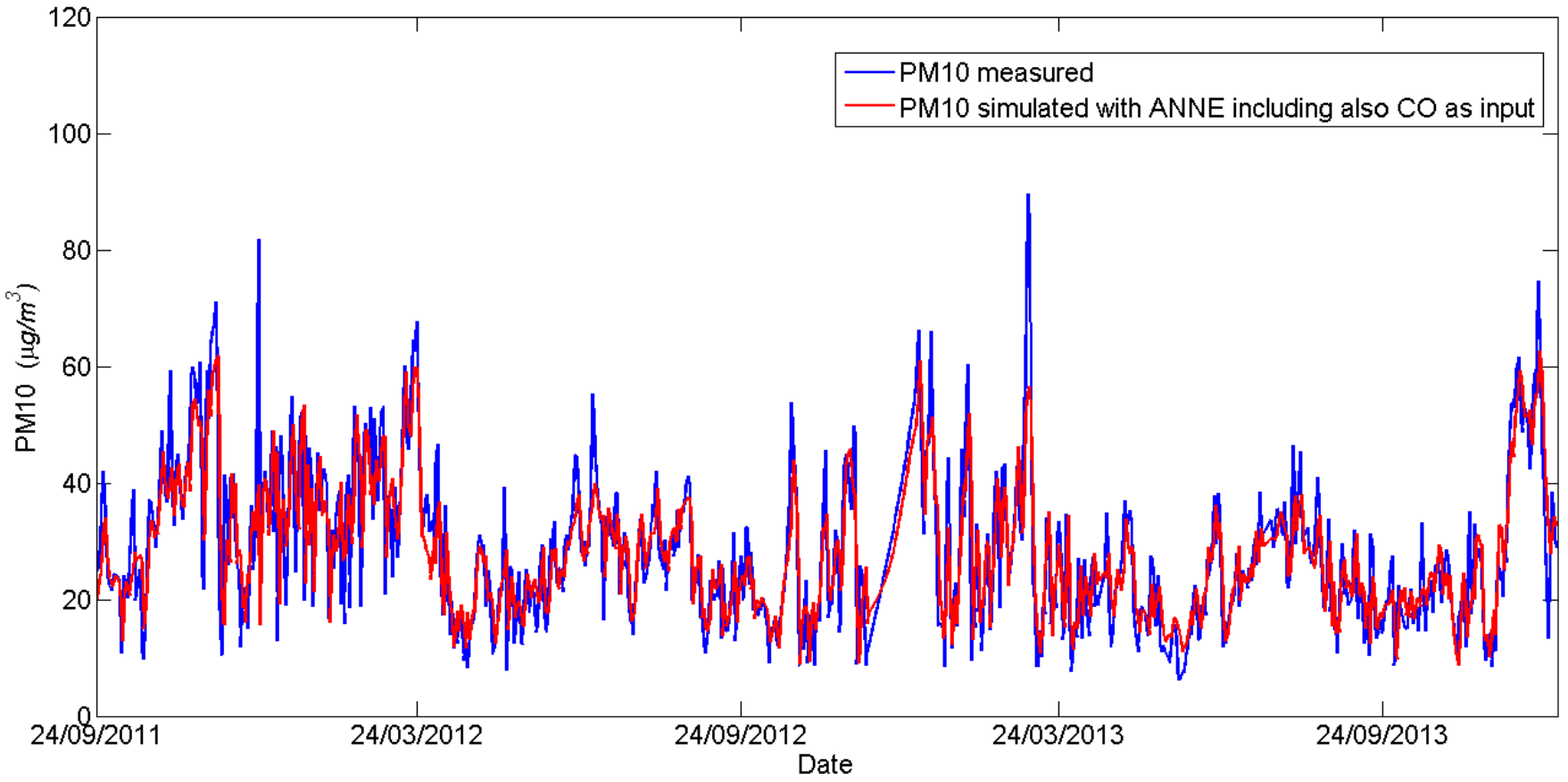


The neural network has a great ability to foresee if the next day the PM10 will be above or below the EU limit: in 95% of the days considered in this study the prediction is correct.



Role of CO as proxy in the PM simulation

Modello ibrido meteo + proxy di emissioni antropiche, incendi ecc.



Summary of the performances of the three models in forecasting PM10

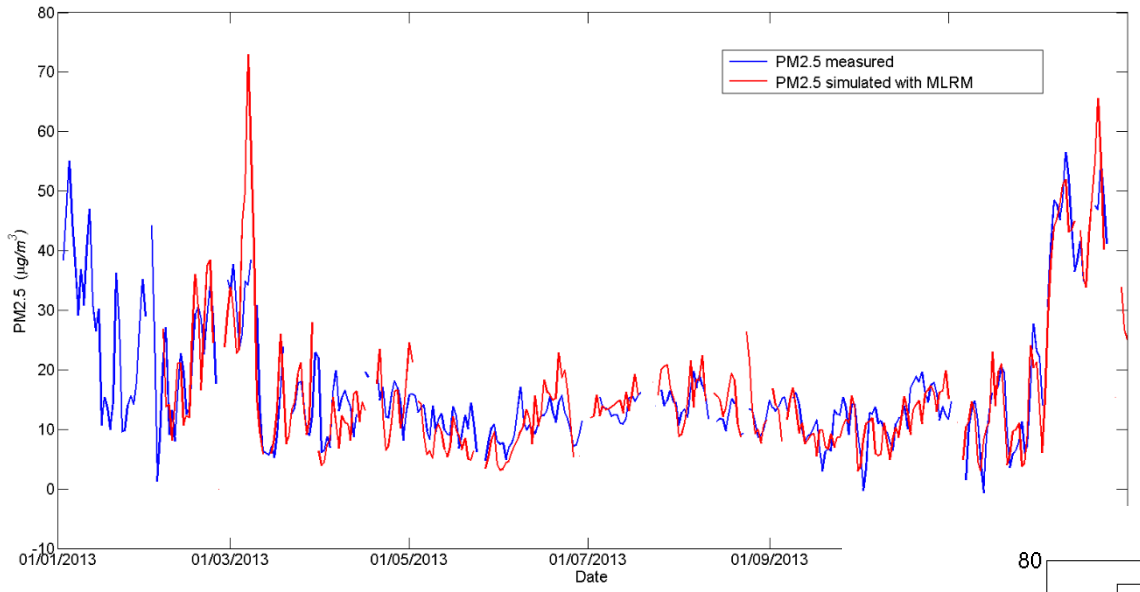
Δt	Model and inputs	R	FB	NMSE	FA2
1	ANNE; Meteo	0.8399	0.0002	0.0652	0.9723
1	ANNF; Meteo	0.8206	0.0001	0.0723	0.9614
1	Regression, meteo	0.5659	0.0100	0.1619	0.8968
2	ANNE; Meteo	0.7770	0.0025	0.0881	0.9577
2	ANNF; Meteo	0.7270	-0.0007	0.1044	0.9420
2	Regression, meteo	0.5611	0.0067	0.1638	0.8955
3	ANNE; Meteo	0.7032	0.0015	0.1123	0.9395
3	ANNF; Meteo	0.6783	0.0013	0.1200	0.9299
3	Regression, meteo	0.5193	0.0077	0.1725	0.8805
1	ANNE; Meteo, CO	0.8478	0.0023	0.0624	0.9735
1	ANNF; Meteo, CO	0.8271	0.0008	0.0700	0.9638
1	Regression, meteo, CO	0.7176	0.0103	0.1153	0.9374
2	ANNE; Meteo, CO	0.7556	0.0017	0.0953	0.9529
2	ANNF; Meteo, CO	0.7354	0.0016	0.1020	0.9444
2	Regression, meteo, CO	0.6222	0.0085	0.1464	0.9082
3	ANNE; Meteo, CO	0.7297	0.0015	0.1039	0.9407

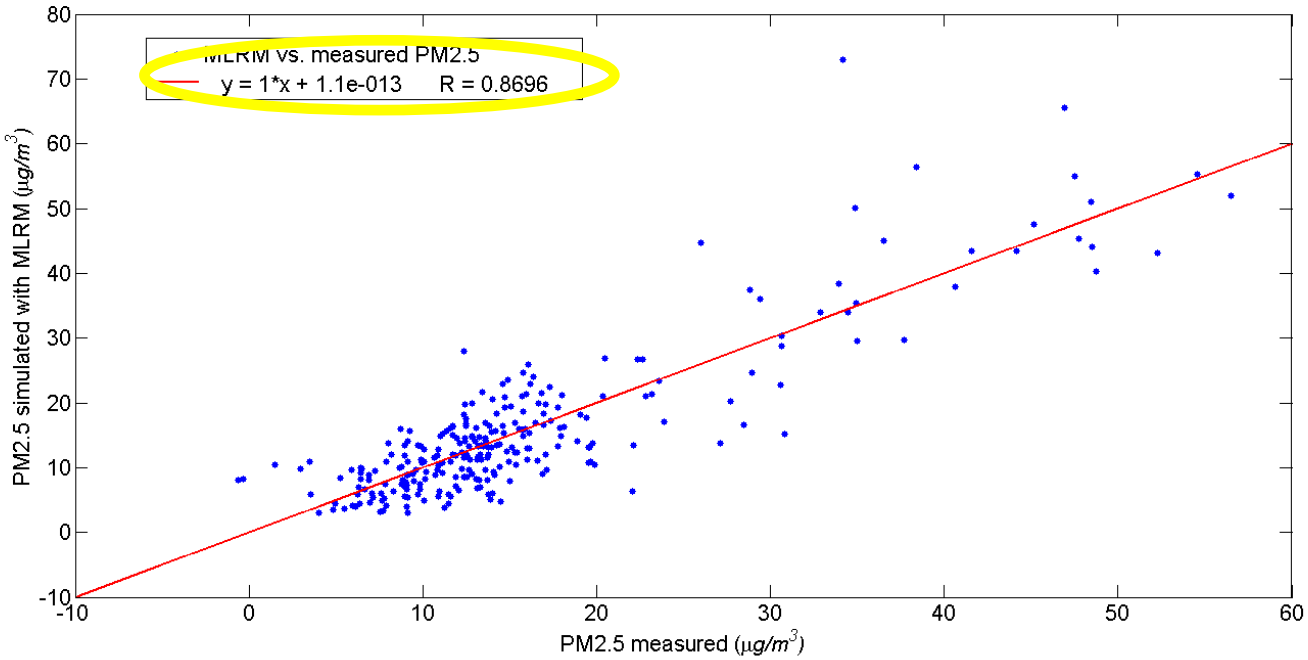
Δt is the forecast time: 1, 2 and 3 days ahead

- **R is the correlation coefficients.**
- **FB is the fractional bias** (-2-2, > 0 overestimation,< 0 underestimation**).**
- **NMSE is the normalized mean squared errors** (close to 0 the best).
- **FA2 is the factor of two** (0-1, close to 1 the best**)**

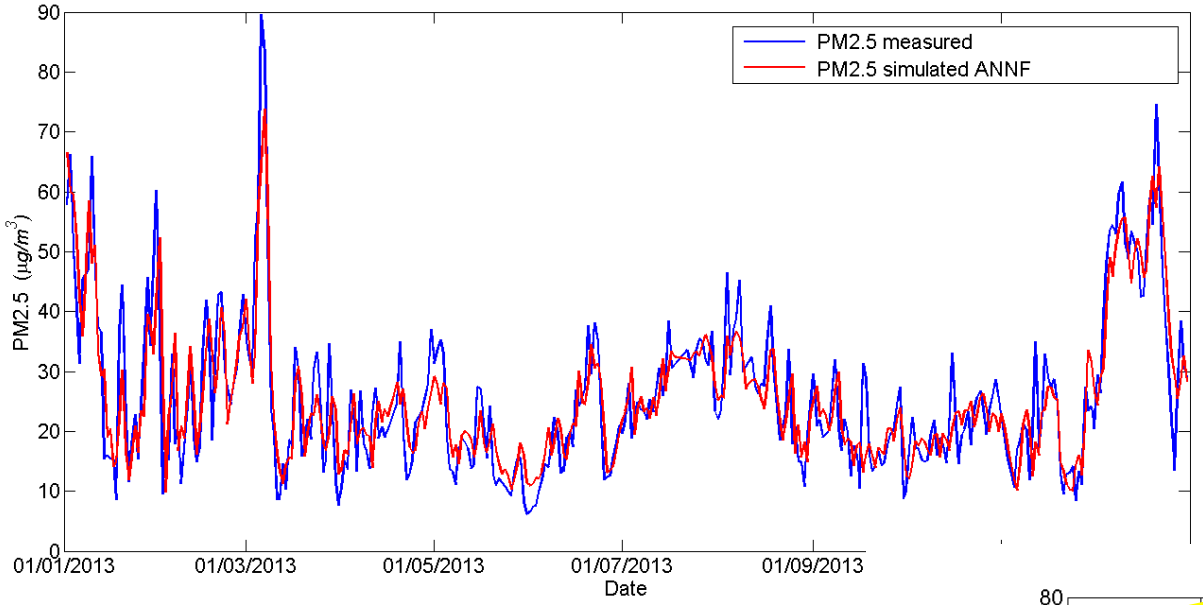
- ANNE (il modello neurale ricorrente) è quello più accurato nelle previsioni a 1,2 e 3 giorni
- La CO in input ha poco peso per i modelli neurali, mentre migliora molto le performance del modello regressivo

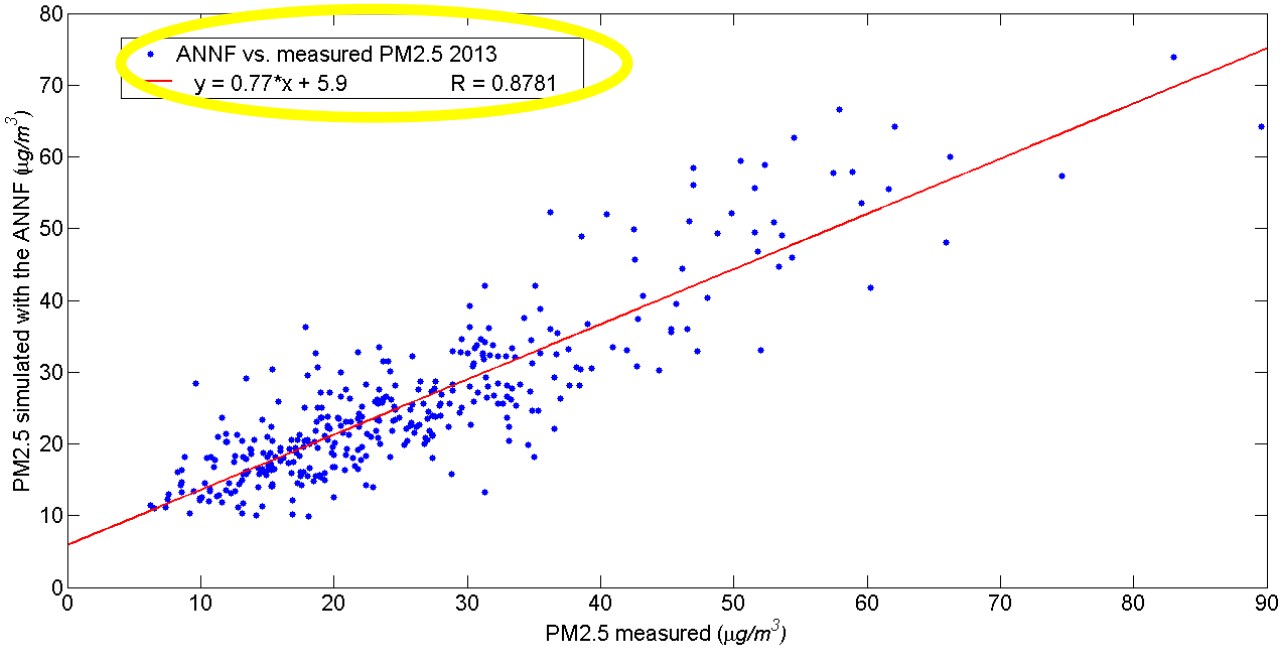
Multiple linear regression model (MLRM) vs observations



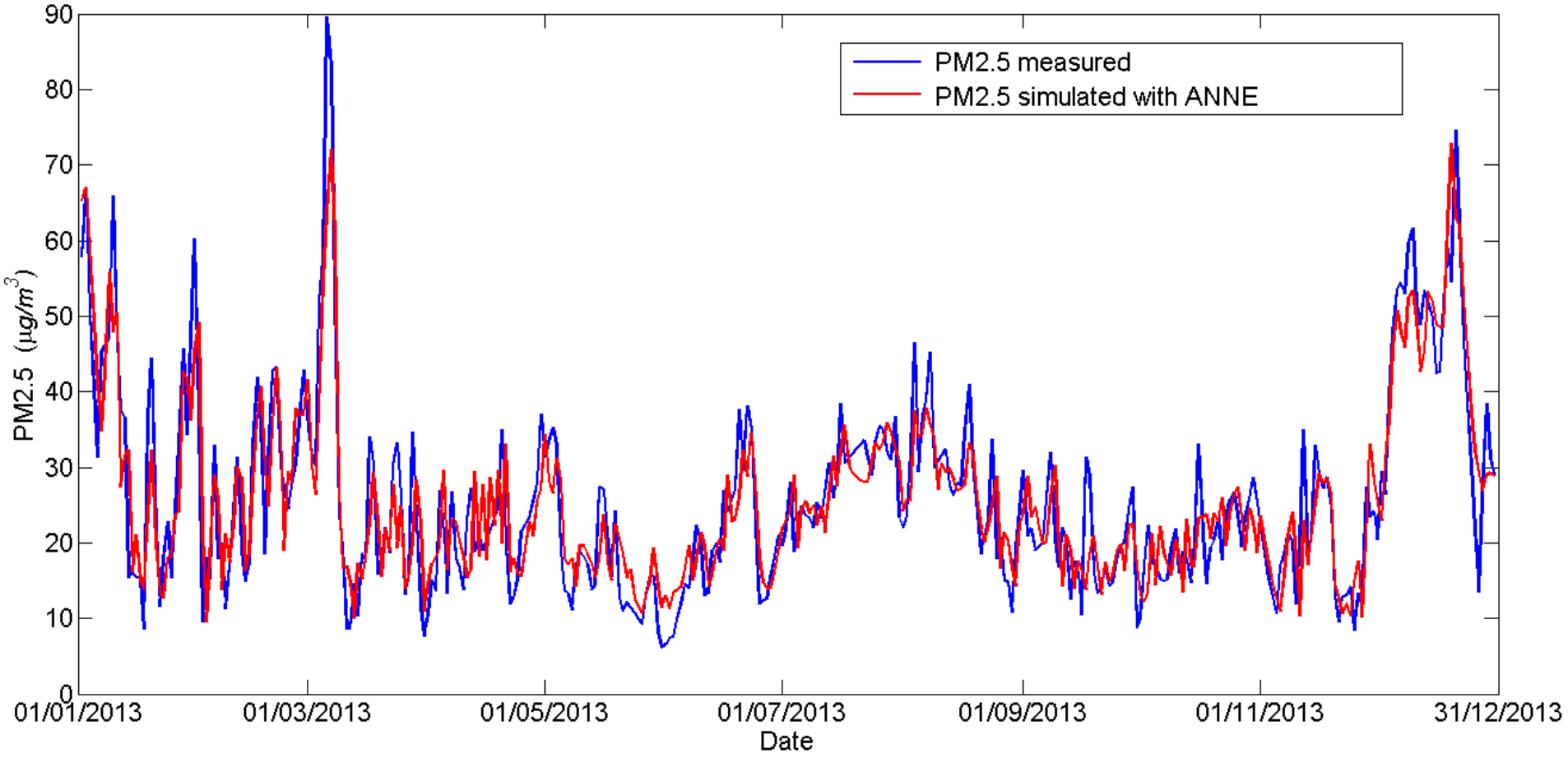


Multi-layer Neural Network feed-forward (ANNF) vs observations





Multi-layer Neural Network recurrent (ANNE) vs observations



Summary of the performances of the three models in forecasting PM2.5

Δt	Parameters	R	FB	NMSE	FA2
1	Neural network (ANNE)	0.8855	0.0019	0.0559	0.9863
1	Regression,	0.8696	0.0532	0.1291	0.9222
2	Neural network (ANNE)	0.8440	0.0029	0.0740	0.9780
2	Regression	0.7891	-0.0464	0.2034	0.8604
3	Neural network (ANNE)	0.8308	0.0007	0.0782	0.9669
3	Regression	0.7203	-0.0430	0.2505	0.8473

Δt is the forecast time: 1, 2 and 3 days ahead

All the forecast proxy: Meteo, PM10 and CO

R is the correlation coefficients, FB is the fractional bias, NMSE is the normalized mean squared errors and FA2 is the factor of two

Conclusioni

- Sviluppato e testato modello Neurale ricorrente per il forecast di PM10 e PM2.5
- Usato per lo studio di dati osservati dall'ARTA in 3 anni nel sito di Pescara
- Confronto con modello regressivo e modello Neurale non ricorrente
- Buone capacità di forecast da 1 a 3 giorni

Torre nel mare Adriatico

a 5 km dalla costa di Pescara struttura di 4 piani per circa 1000mq e area di mare circostante protetta per 900 ettari

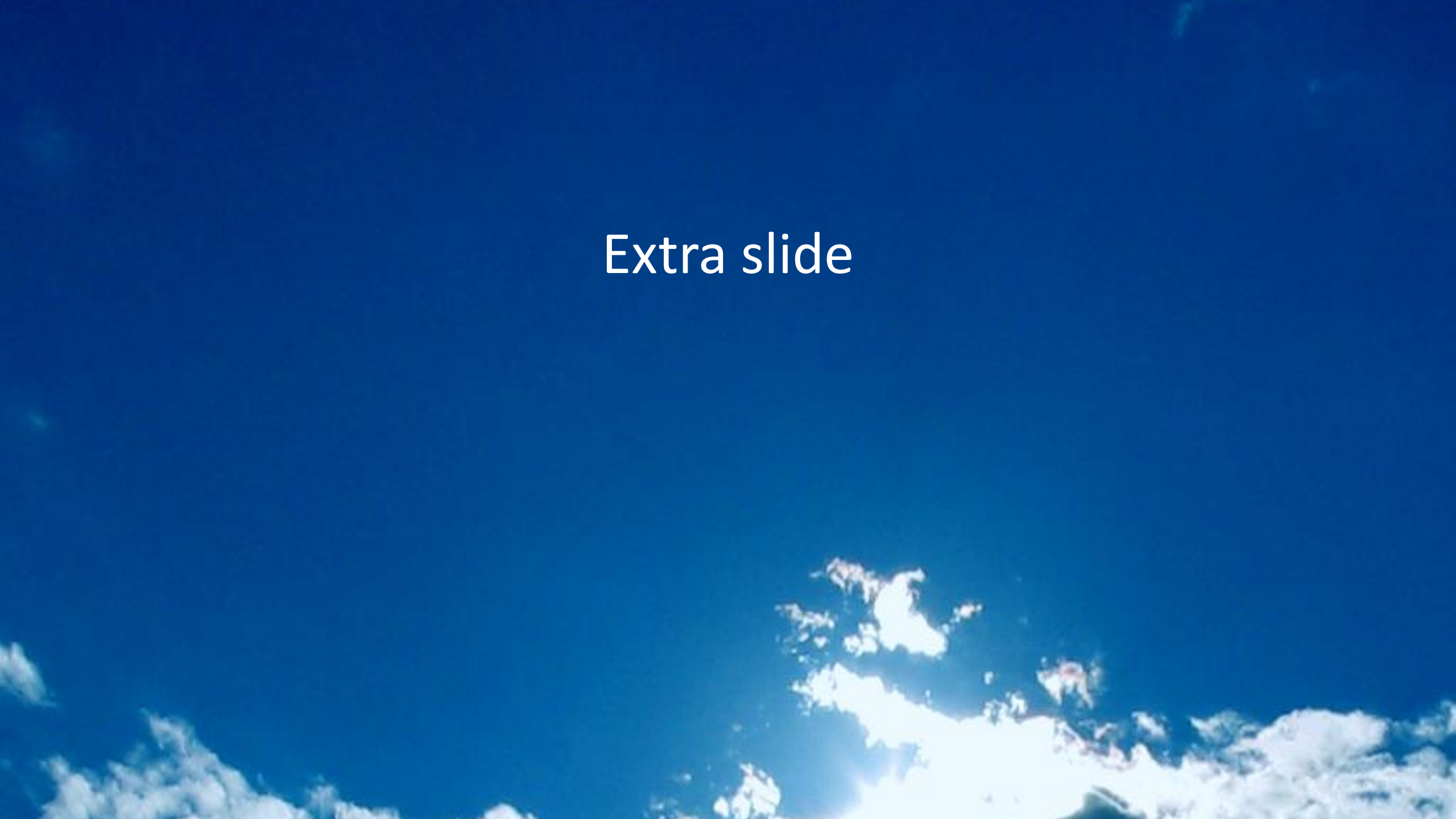
Stazione meteo ed analizzatore di Ozono

Pescara

Gran Sasso Stazione Portella
a 2400 m. slm (Meteo, O3, OPC)

Collaborations are very welcome:
Email: piero.dicarlo@aquila.infn.it

Extra slide



Statistical parameters

- NMSE is found calculating the mean of the square of the difference between the pairs of modelled and measured values, finally this value is normalized by dividing by their product. This parameter gives emphasis to the whole error of the data set. The best NMSE value that can be achieved is 0.
- FB is calculated subtracting from the mean of the measured data, the mean of simulated data, then dividing by their mean. This parameter vary between -2 and 2. A positive value indicate that the model underestimates the measured data, while a negative value indicates an overestimation.
- FA2 is the percentage of ratio between measured and simulated data that lies in the range 0.5 and 2. FA2 ranges from 0, which indicates that no ratio is in that range, to 1, which indicates that all ratio lies in this range.
- Model SM200 (Enviro Technology)
- SWAM 5° FAI