Qualità dell'aria: integrazione del machine learning con il modello fisico per l'analisi di un grande dataset distribuito

In collaborazione con SSPT-MET-INAT, Mario Adani, Gino Briganti, Mihaela Mircea

Cagliari, 03 Ottobre 2018

Angelo Mariano, DTE-ICT-IGEST
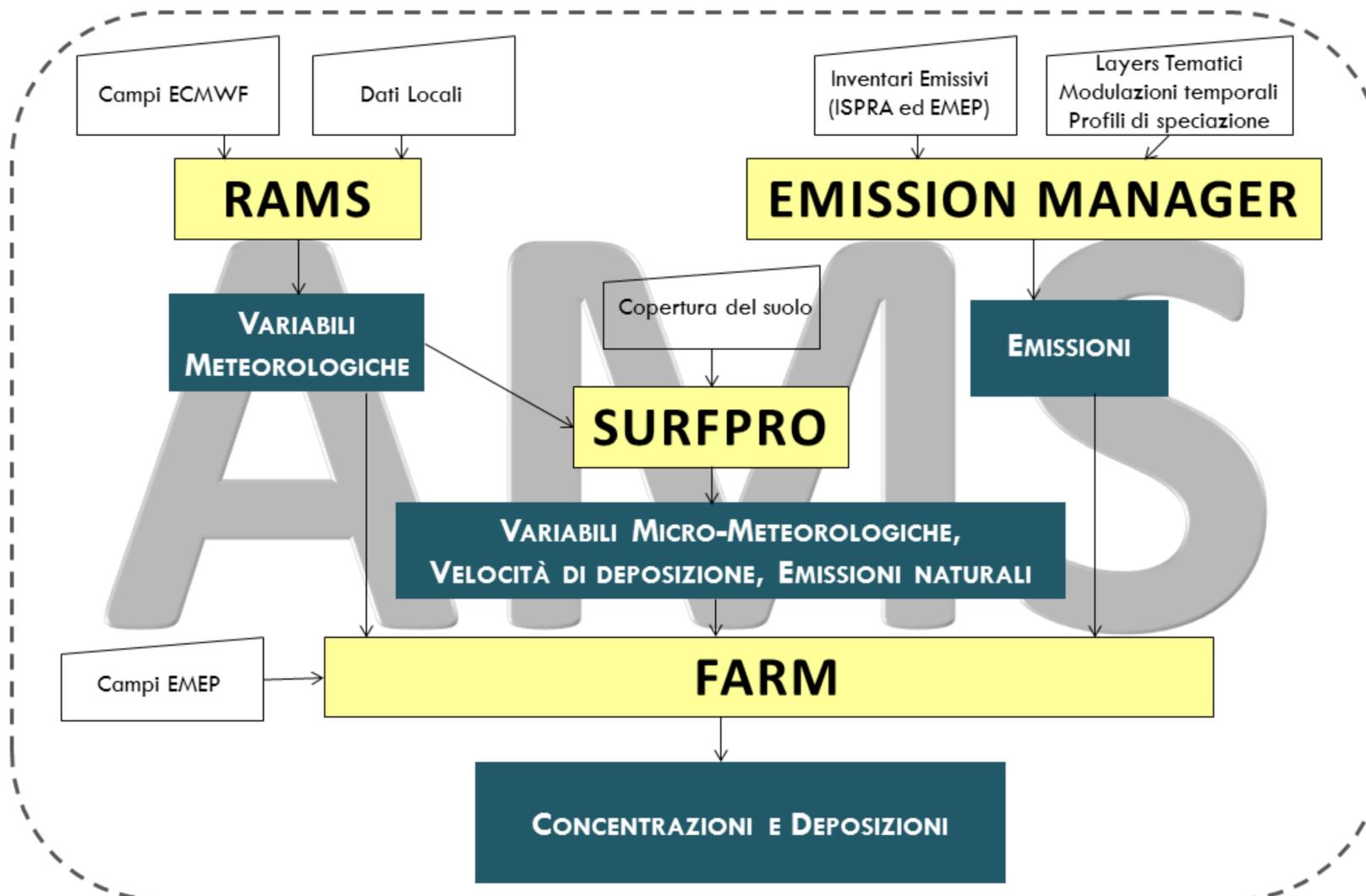
Concentrazione NO2
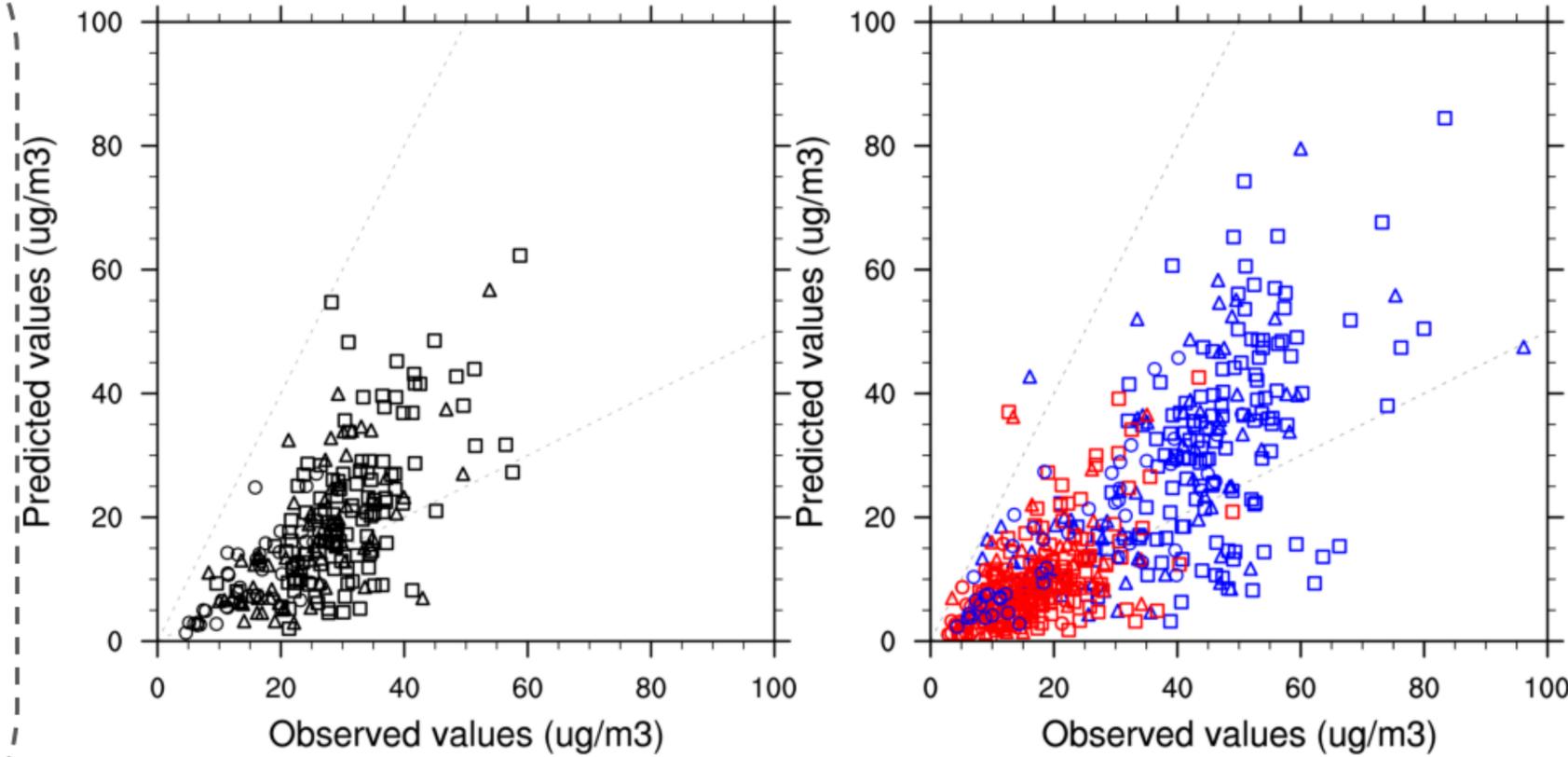
	Corr	*SR*	*FAC2*	*BIAS*	*FB*	*RMSE*	*NMSE*
Rurali **(46 Staz.)**	0.42 ± 0.18	0.93 ± 0.39	0.45 ± 0.14	-5.2 ± 4.9	-0.45 ± 0.36	13.8 ± 5.3	1.44 ± 0.82
Suburbane **(66 Staz.)**	0.52 ± 0.11	0.80 ± 0.33	0.46 ± 0.19	-8.7 ± 8.5	-0.51 ± 0.46	19.9 ± 6.6	1.59 ± 1.71
Urbane **(122 Staz.)**	0.55 ± 0.12	0.84 ± 0.30	0.49 ± 0.20	-10.1 ± 9.3	-0.47 ± 0.43	22.4 ± 5.5	1.38 ± 2.15
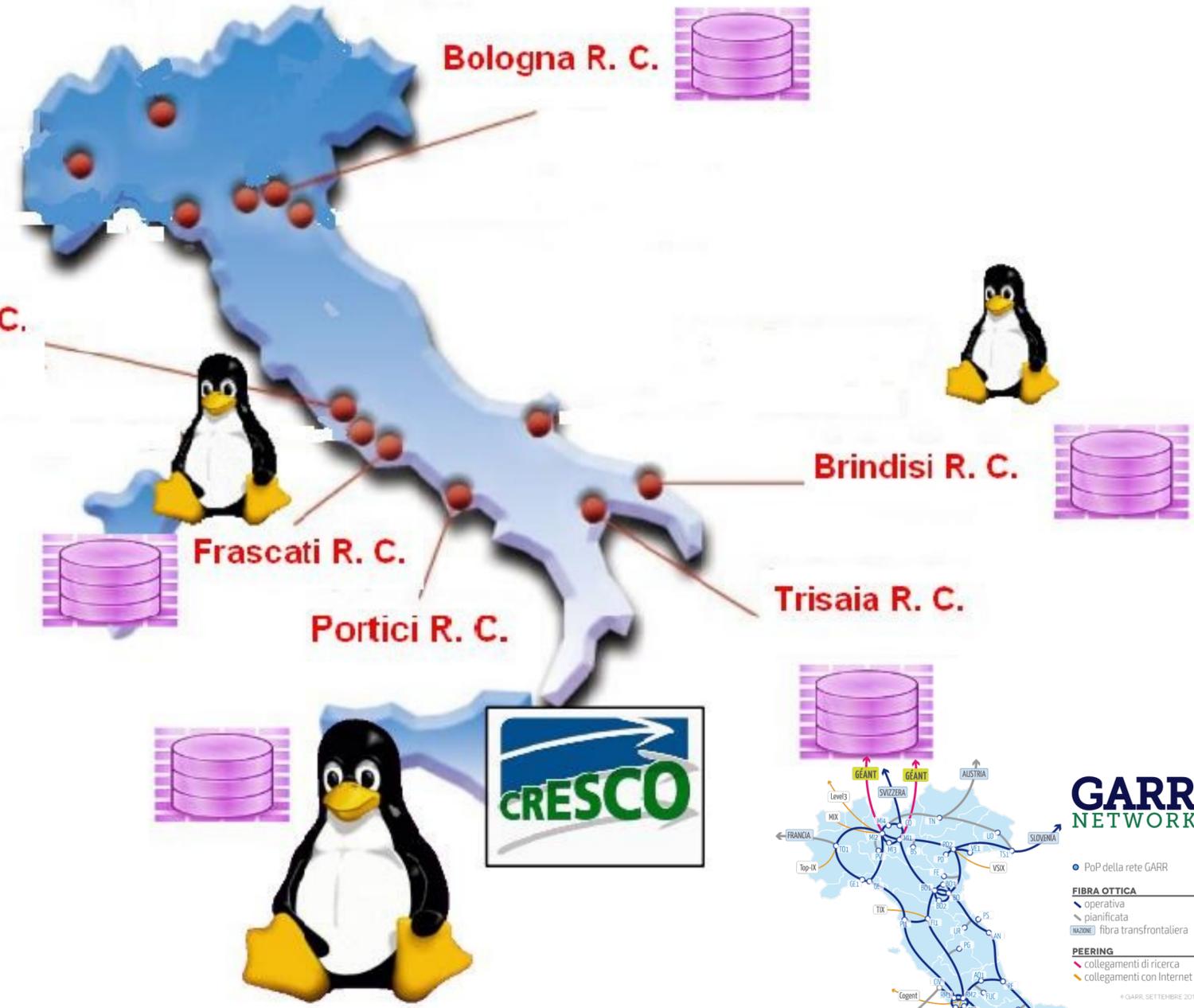
| Process	Cores	Queues	Elapsed time	Disk space	Total disk space
Boundary conditions	1	cresco4_h6	~ 20'/year	1.6 GB/year	29 GB
Meteorology	-	Local	-	200 GB/year	3.7 TB
SURFPRO	1	cresco4_h144 cresco3_h144	~ 45'/year ~ 1h20'/year	115 GB/year	2 TB
Emission Manager	12	cresco4_h6 cresco3_h144	~ 2h/year ~ 4h/year	400 GB/year	7 TB
FARM	16	cresco4minni_h24	~ 1÷1.5 day/year depending on Grid crowding and other I/O simultaneous processes.	93 GB/year	3.5 TB
Post-processing	1	cresco4_h144 cresco3_h144 cresco4_h6	~ 2÷3 day/year	38 GB/y	1.4 TB
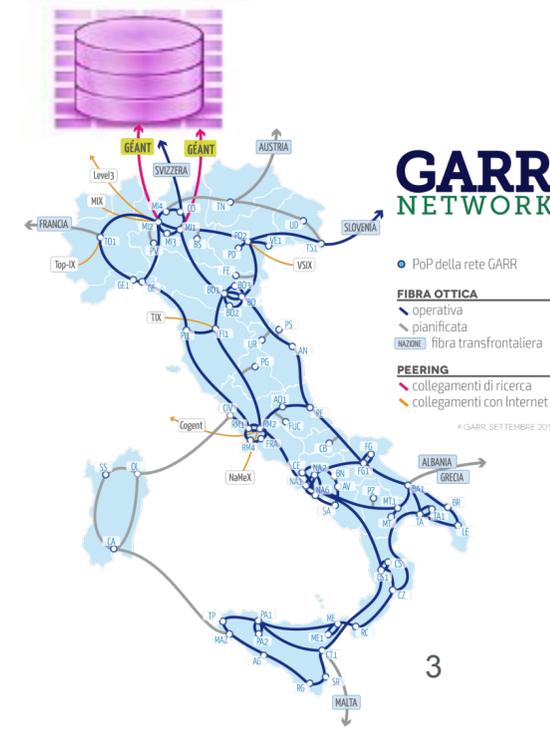
Model

50 LSTM cells

1 Dense layer

20% dropout

1/2 hour of training on a 2016 Macbook

Legend: Layer — Pointwize op — Copy

Data
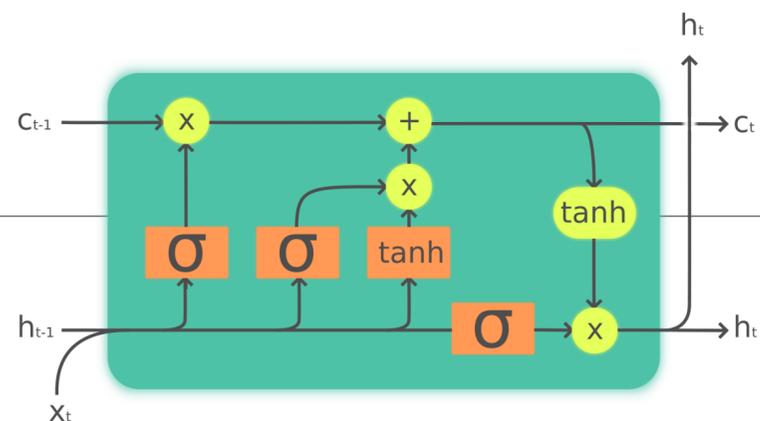
- About 400 non-zero features (anthropogenic and natural emissions, meteorological features, soil deposition and different height concentrations)

- 365 days, 24 hours a day

- 2 lag hours

- Only 1 site (non-urban)
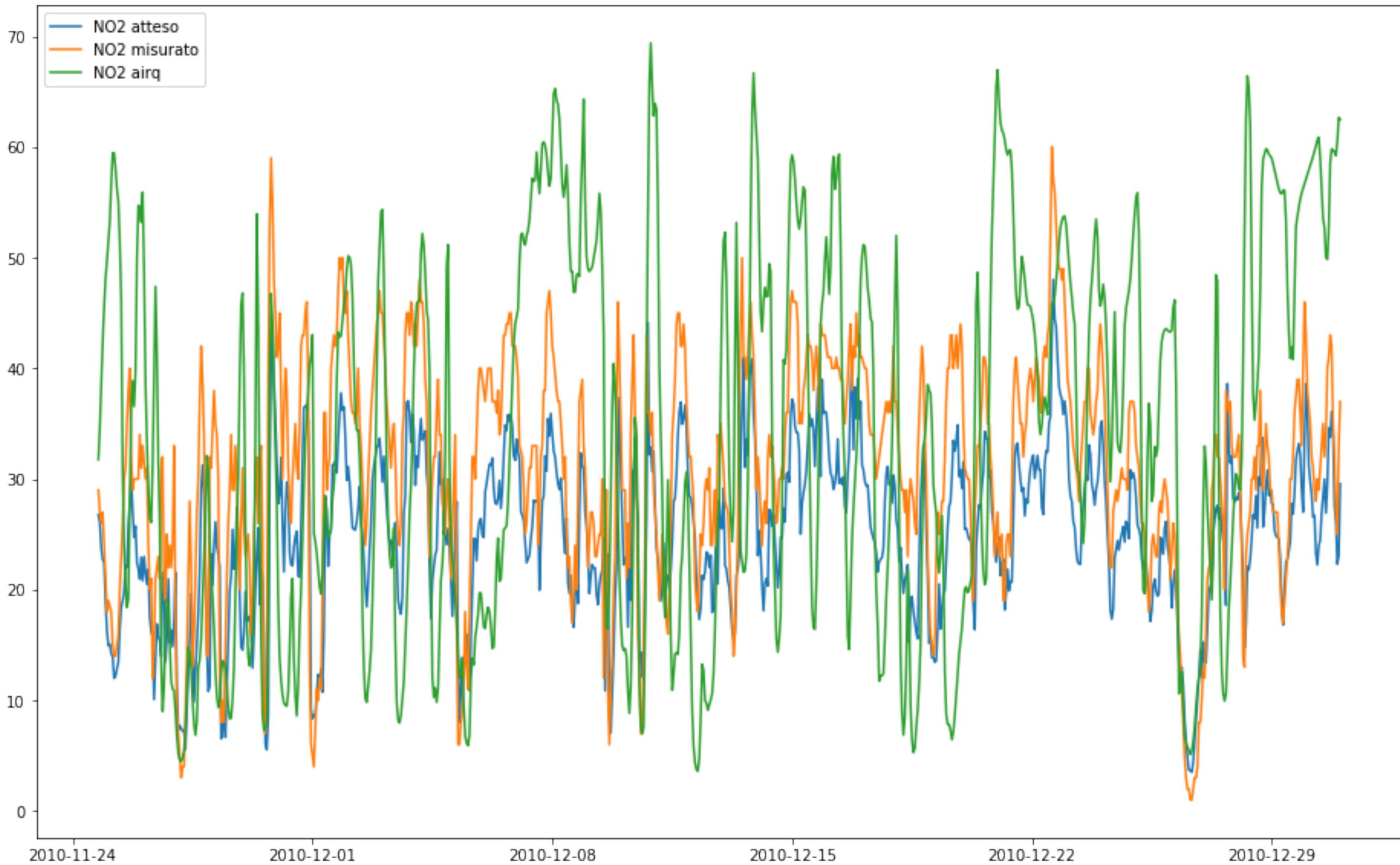
FARM:

RSE = 1.759

ML:

RSE = 0.765
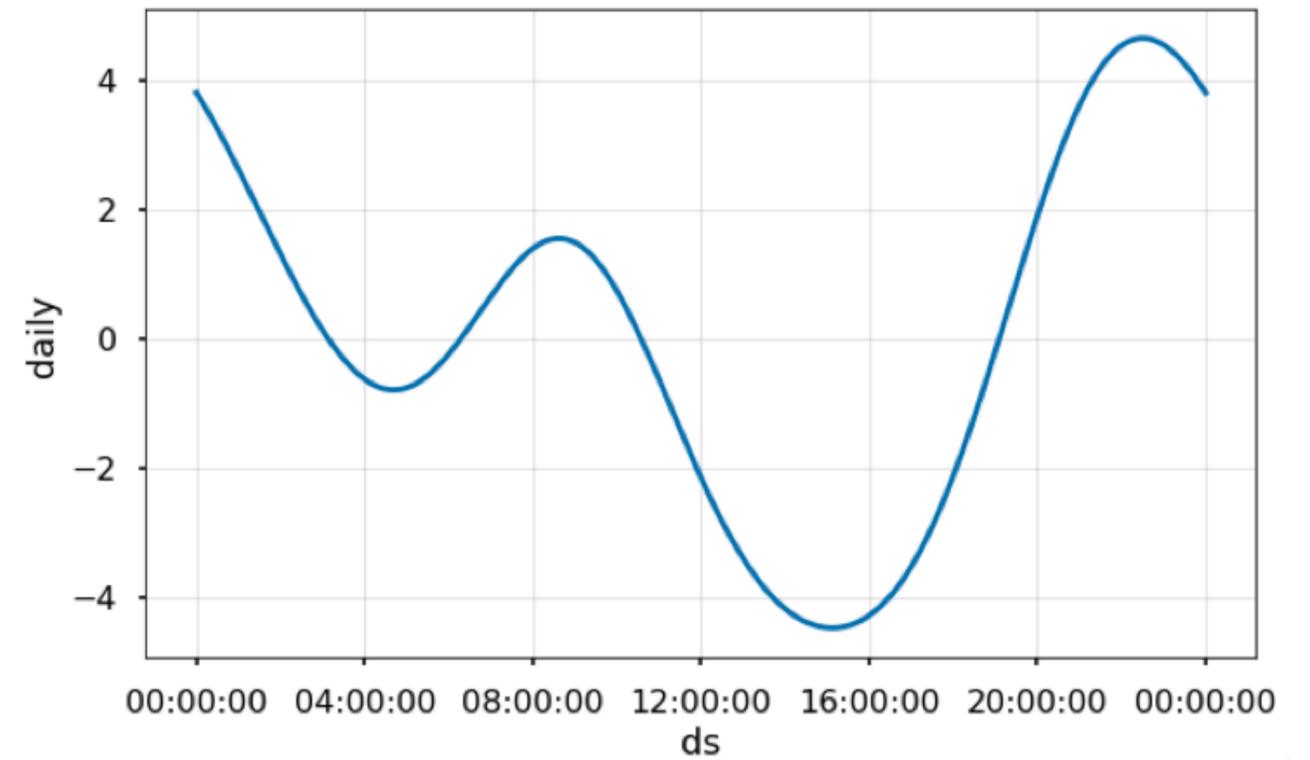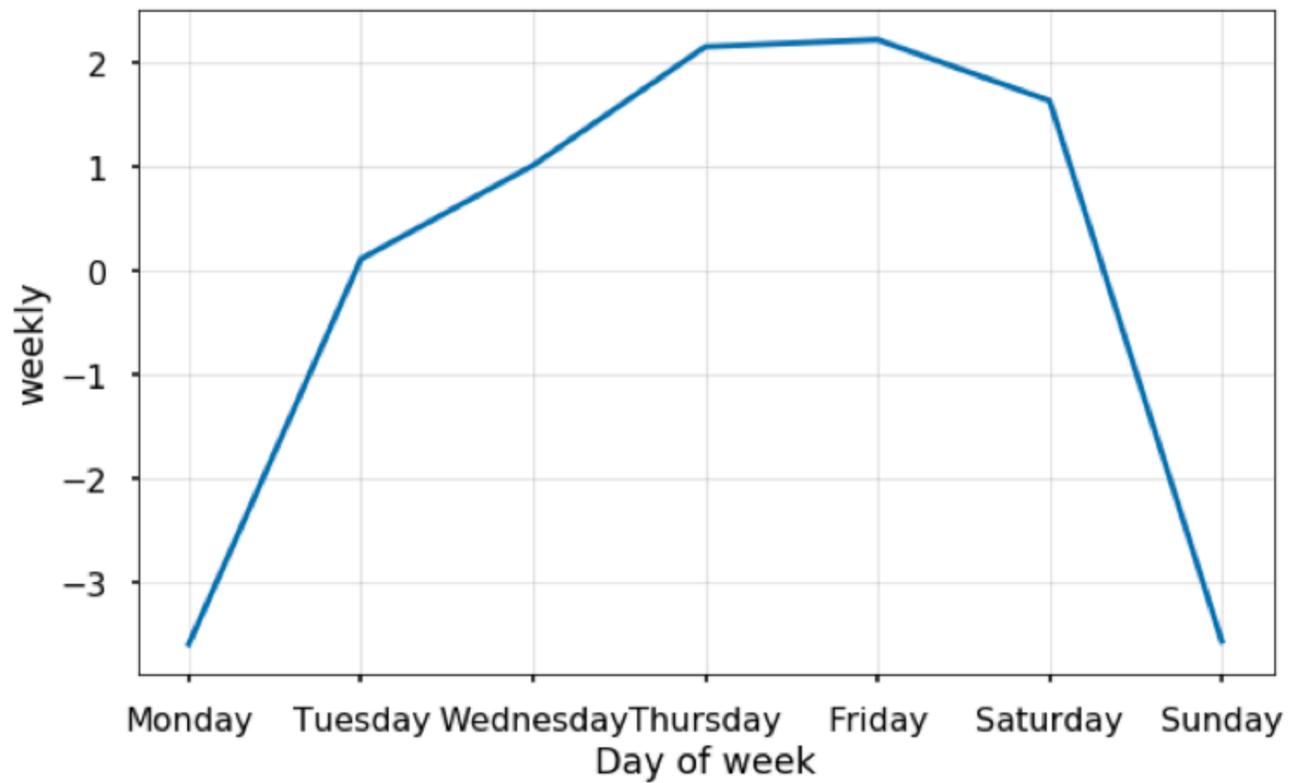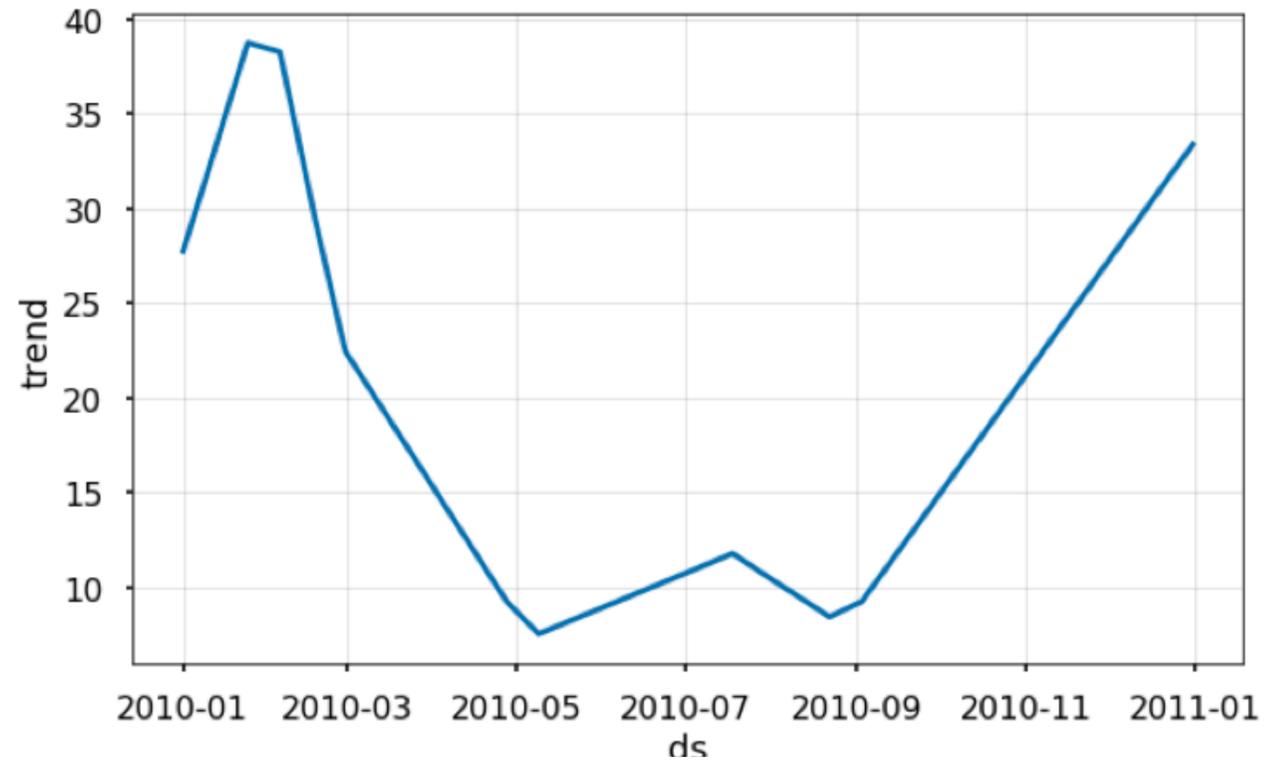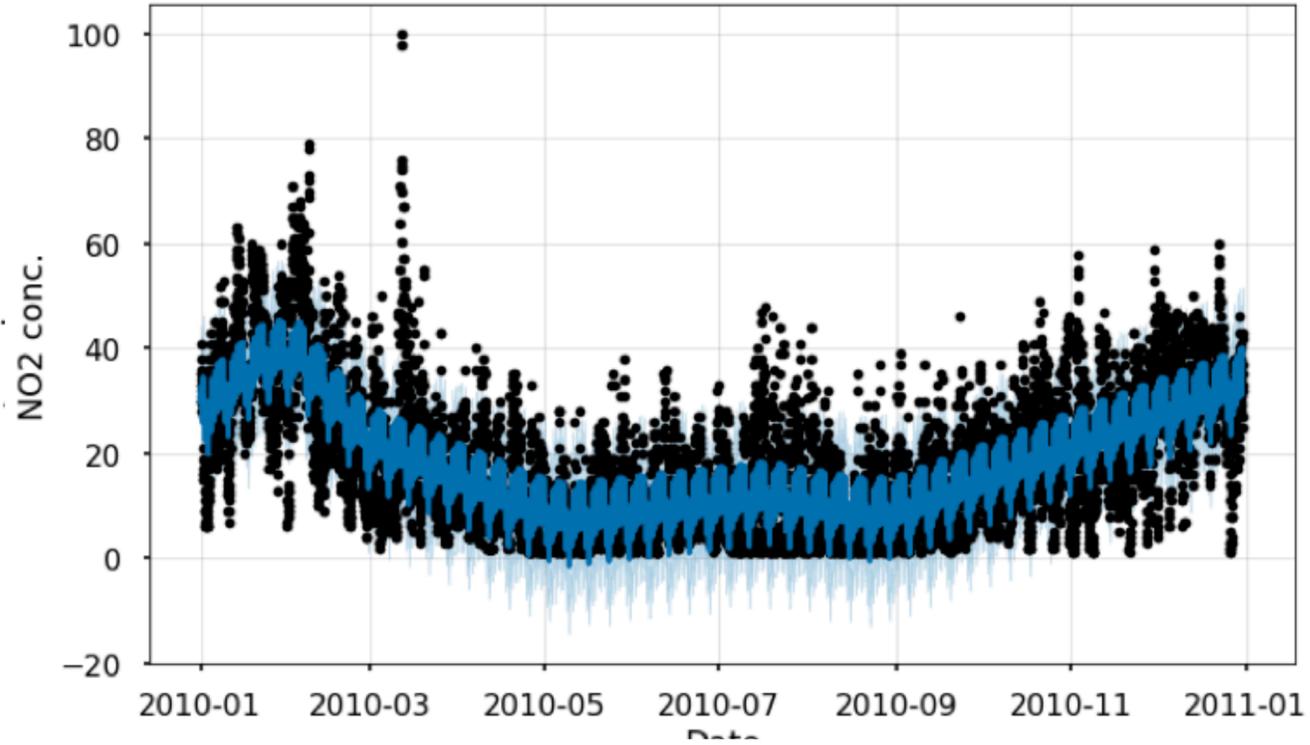
1st urban site

1st urban site

2nd urban site

1st urban site

2nd urban site

non-urban site
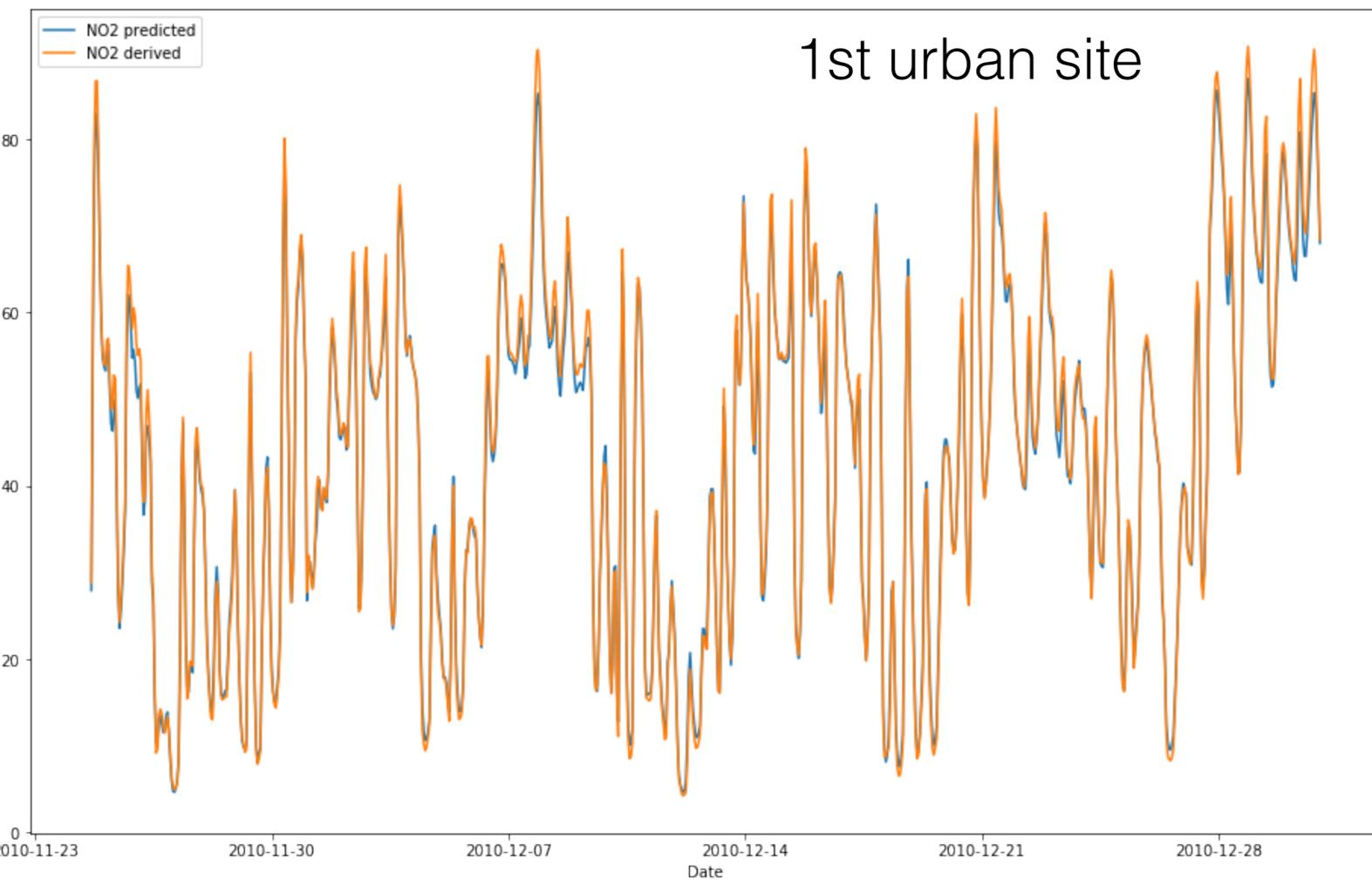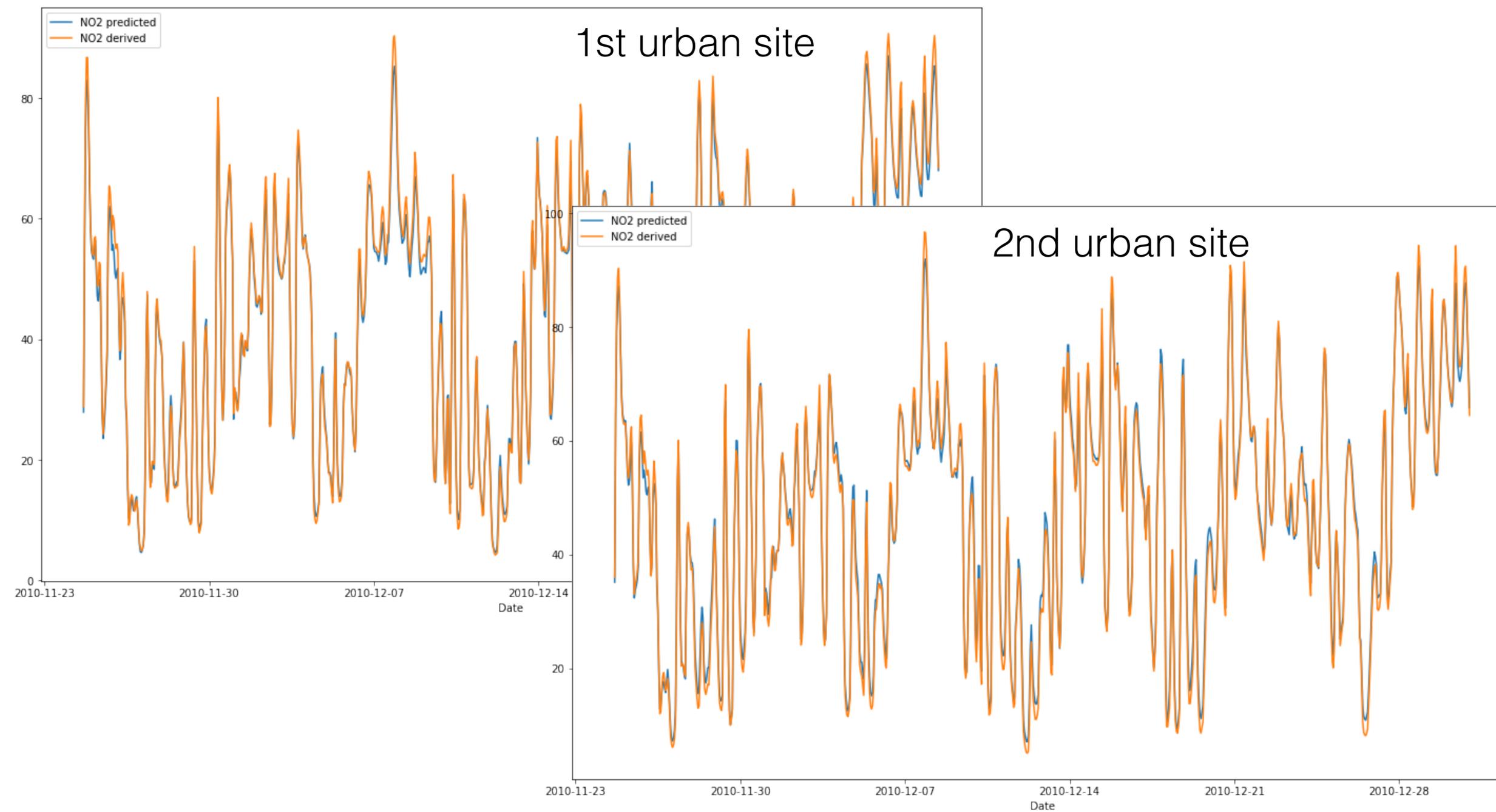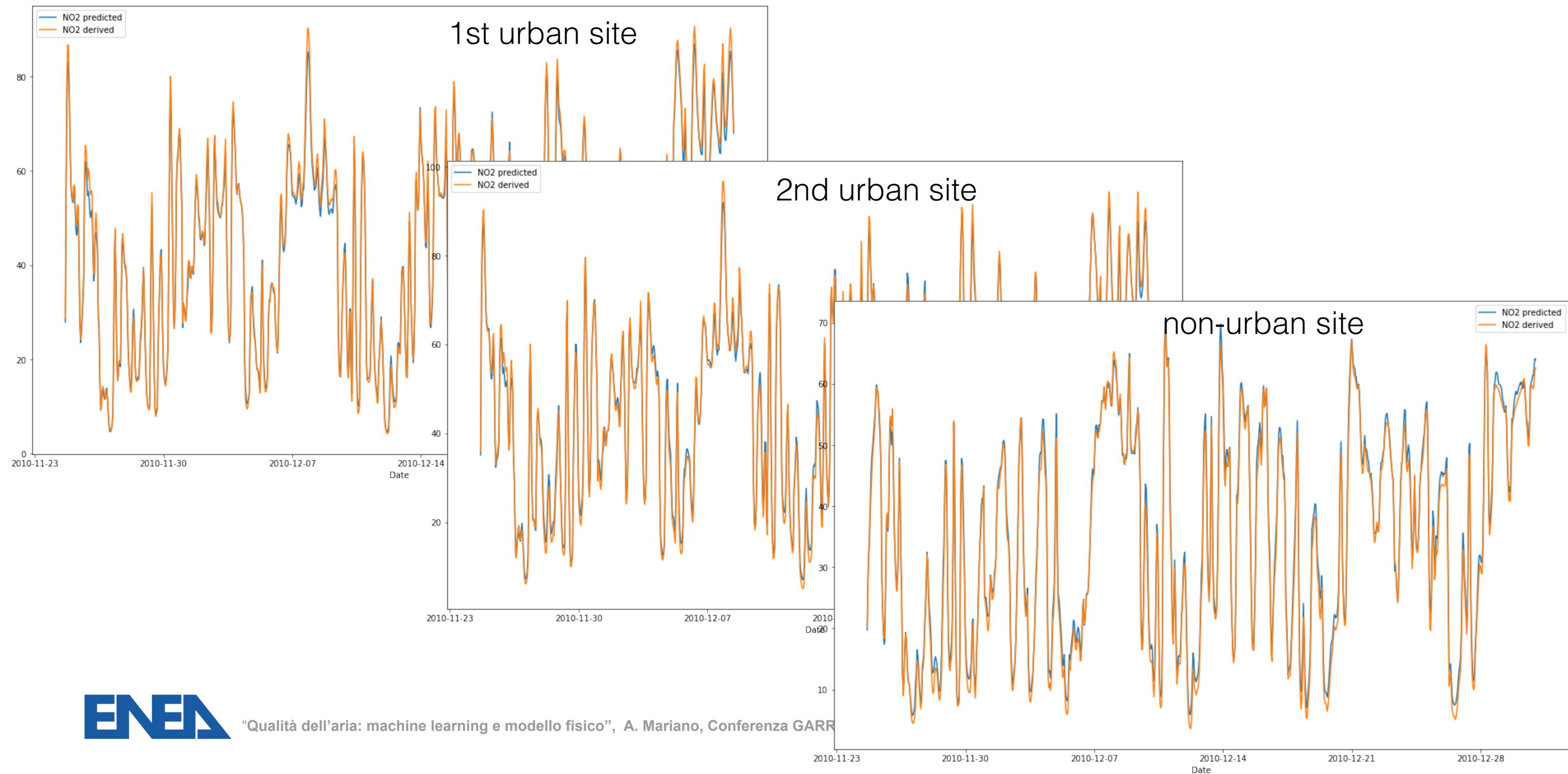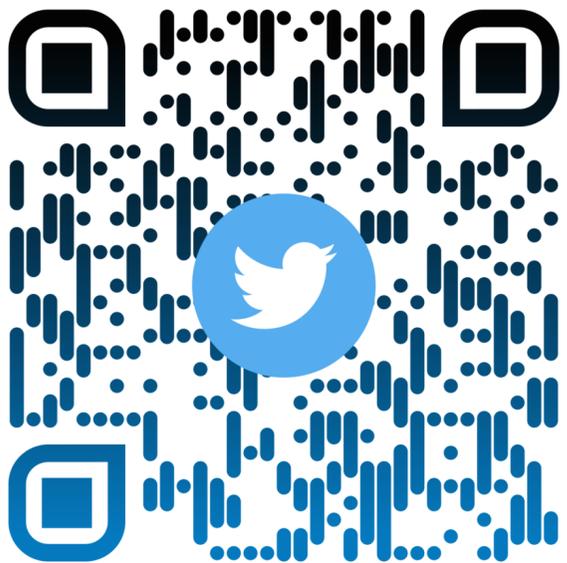
Angelo Mariano
angelo.mariano@enea.it