Air pollutants modelling and forecast using in situ and CAMS data in Portugal.

Fundação para a Ciência e a Tecnologia

FAIR Project - FAIR-2022.01660.PTDC

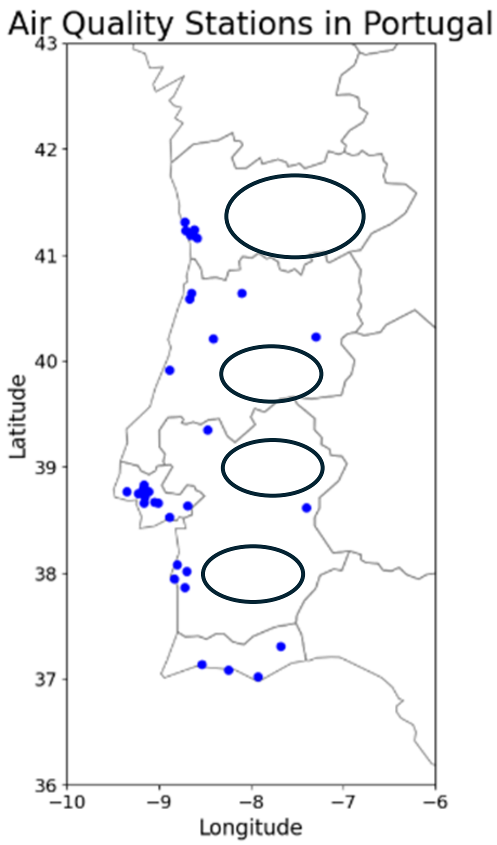
André Brito, Rita Durão, Ana Russo

2024

FAIR Project

FAIR Project goals:

- Benefit from the complementarity between **CAMS** and **QualAR** datasets.
- Characterization of space-time patterns of air pollutants and meteorological parameters for the period under study, over mainland Portugal.
- Development of an AQ forecasting tool for areas less covered by the national air quality network for each Portuguese **NUTS II** region:
  - **MLP** and **DL-LSTM** models for PM10 one-day ahead prediction.



Data

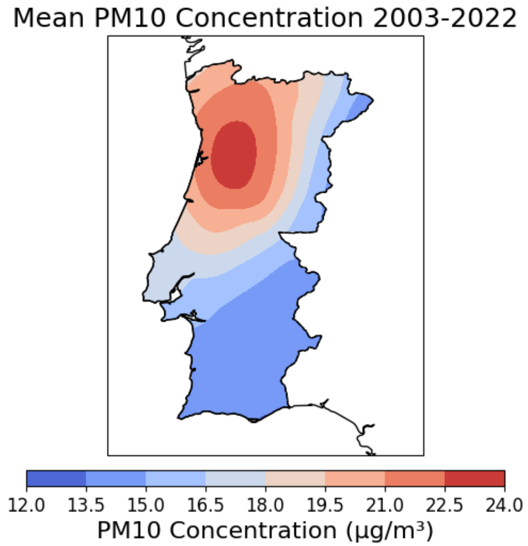
Spatial domain: Portugal
Temporal Domain: 2003-2022

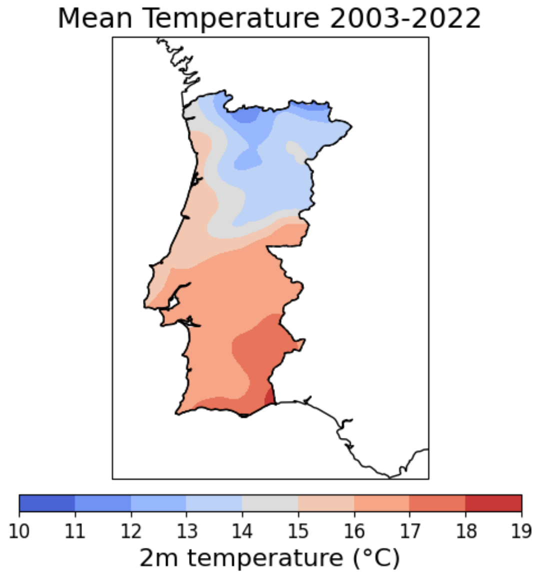
CAMS Global Reanalysis (EAC4) (0.75ºx0.75º):

- PM10
- PM2.5
- CO
- NO_2
- O_3

ERA-5 hourly data on single levels from 1940 to present (0.25ºx0.25º):

- 2m Temperature
- 10m u and v- components of wind
- Surface Pressure
- Surface Solar Radiation Downwards
- Total Precipitation
- Boundary Layer Height
- Relative Humidity

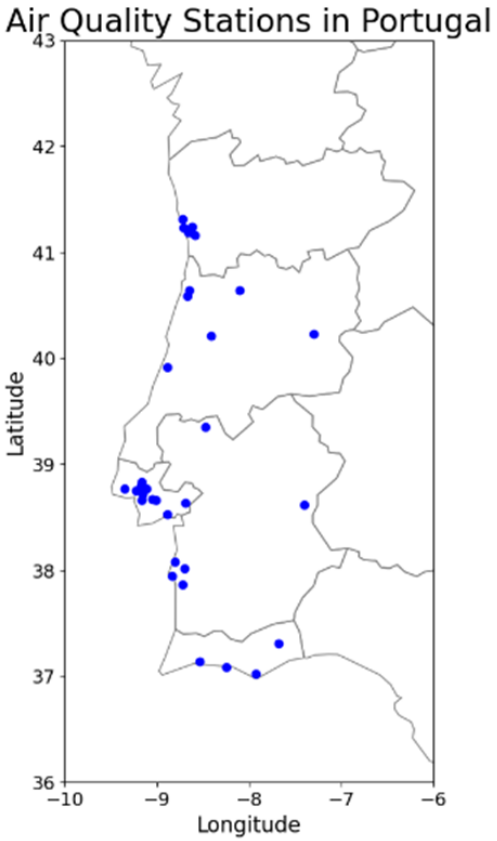




Data

Spatial domain: Continental Portugal
Temporal Domain: 2003-2022

Station name	Years active
Perafita	2003-2022
Matosinhos	2003-2022
Senhora da Hora	2003-2022
Vila do Conde	2003-2022
Custóias	2003-2016
Vermoim	2003-2018
Campanhã	2010-2019
Coimbra	2003-2022
Fundão	2003-2022
Aveiro	2003-2022
Ílhavo	2003-2022
Chamusca	2003-2022
Ervedeira	2003-2022
Fornelo	2005-2022
Lavradio	2003-2022
Olivais	2003-2022
Entrecampos	2003-2022
Av. Liberdade	2003-2022
Laranjeiro	2003-2022
Loures	2003-2022
Quebedo	2003-2022
Mem Martins	2003-2022
Reboleira	2003-2022
Escavadeira	2003-2022
Odivelas	2003-2022
Fernando Pó	2008-2022
Terena	2005-2022
Sonega	2014-2022
Santiago do Cacém	2011-2017, 2021-2022
Monte Velho	2005-2017, 2021-2022
Monte Chãos	2014-2022
Joaquim Magalhães	2004-2007, 2010-2022
Malpique	2004-2007, 2010-2022
David Neto	2004-2007, 2010-2022
Cerro	2004-2007, 2010-2022



PM10 in situ data from the QualAr network
Portuguese Environment Agency

Methods

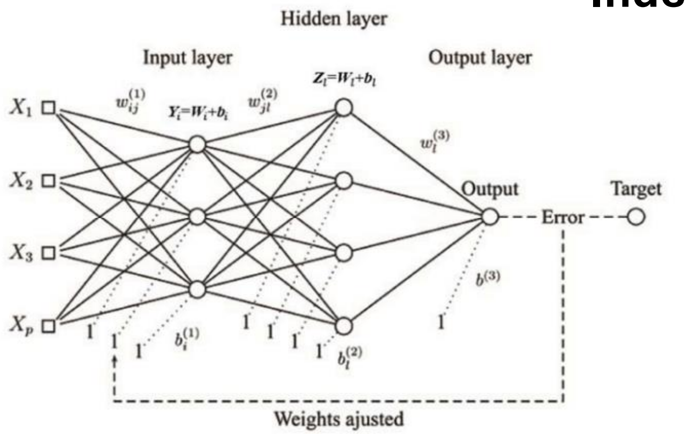
Predictors
PM10 at 00:00
Max PM10
PM10
PM2.5
CO
NO_2
O_3
Temperature
Relative Humidity
Pressure
Wind Speed
Boundary Layer Height
Precipitation
Radiation
CWT
Stations
ToySin
ToyCos

Stepwise Forward Regression

North	Center	Lisbon	Alentejo	Algarve
PM10 at 00:00	PM10 at 00:00	PM10 at 00:00	PM10 at 00:00	PM10 at 00:00
Stations	CO	BLH	Mean PM10	CO
CO	Max PM10	Radiation	Stations	NO_2
Temperature	Temperature	NO_2	Precipitation	O_3
Wind	Wind	Precipitation	Wind	Stations

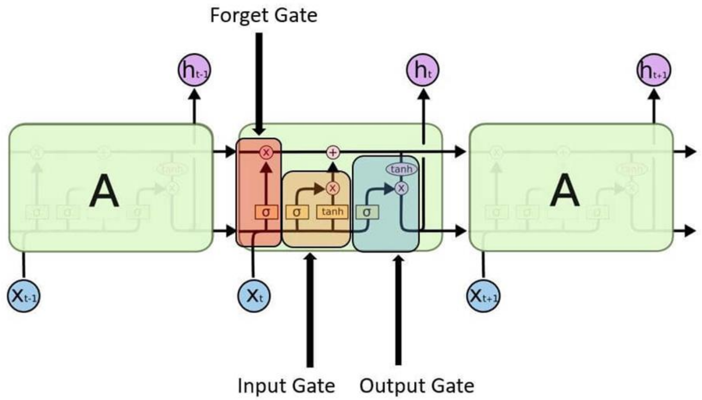
Results

Training and cross-validation: 2003-2021
Independent testing: 2022



Epochs	100
Activation function	ReLU
Number of hidden layers	2
Number of neurons in each hidden layer	32
Optimizer	Adam
Loss function	MSE

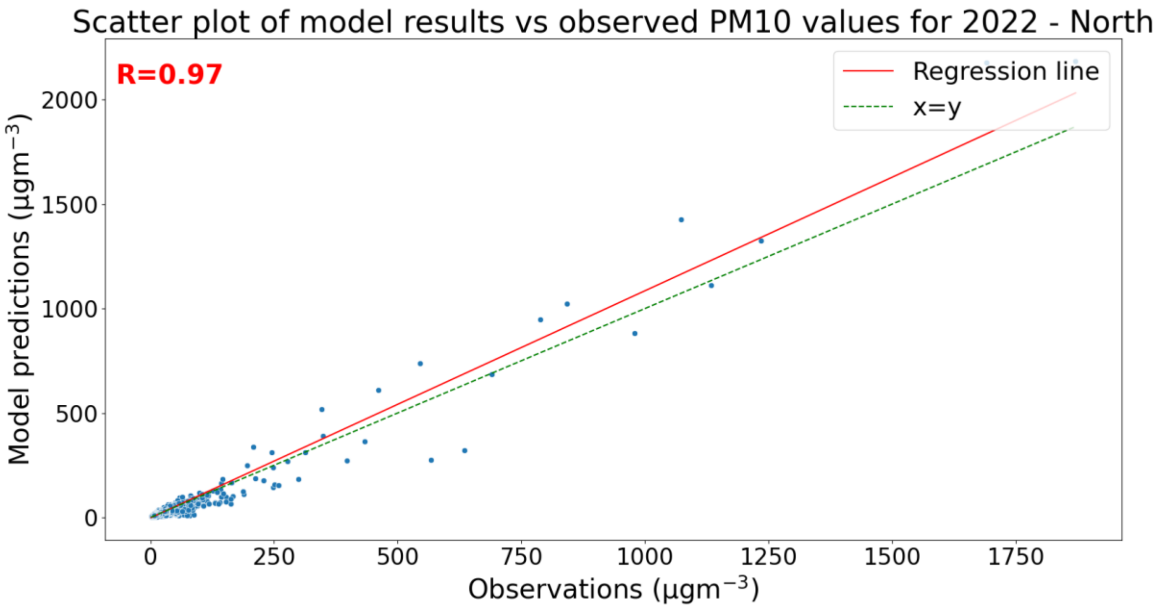
MLP



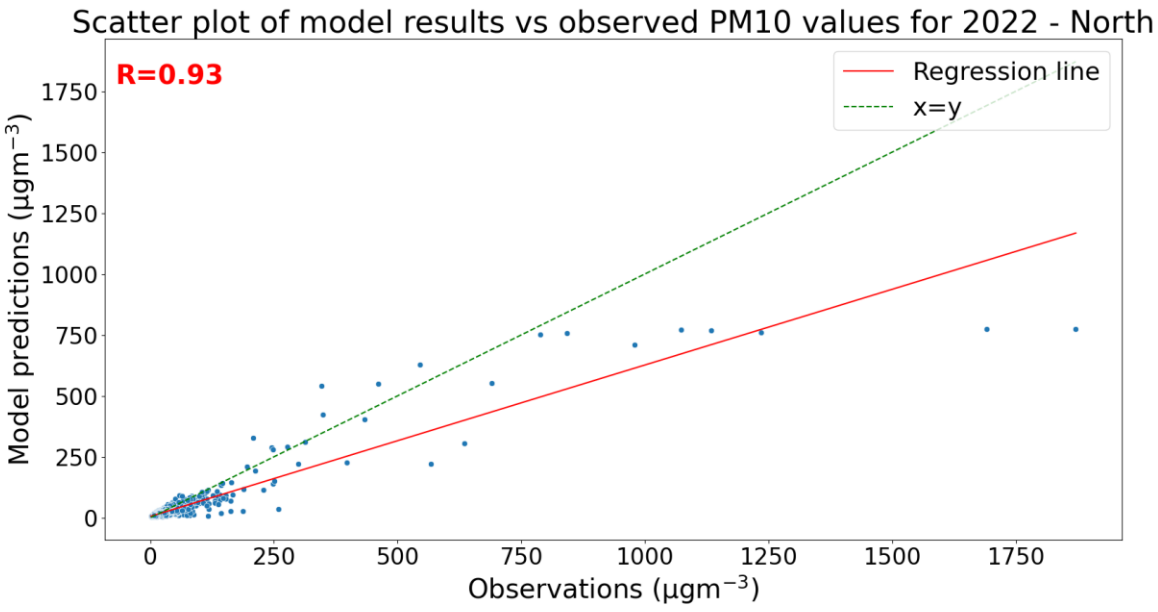
Epochs	100
Activation Function (LSTM layer)	Tanh(cells) / sigmoid (gates)
Number of LSTM hidden layers	2
Number of neurons in each hidden layer	64
Optimizer	Adam
Loss function	MSE

DL-LSTM

Results – Scatter Plots

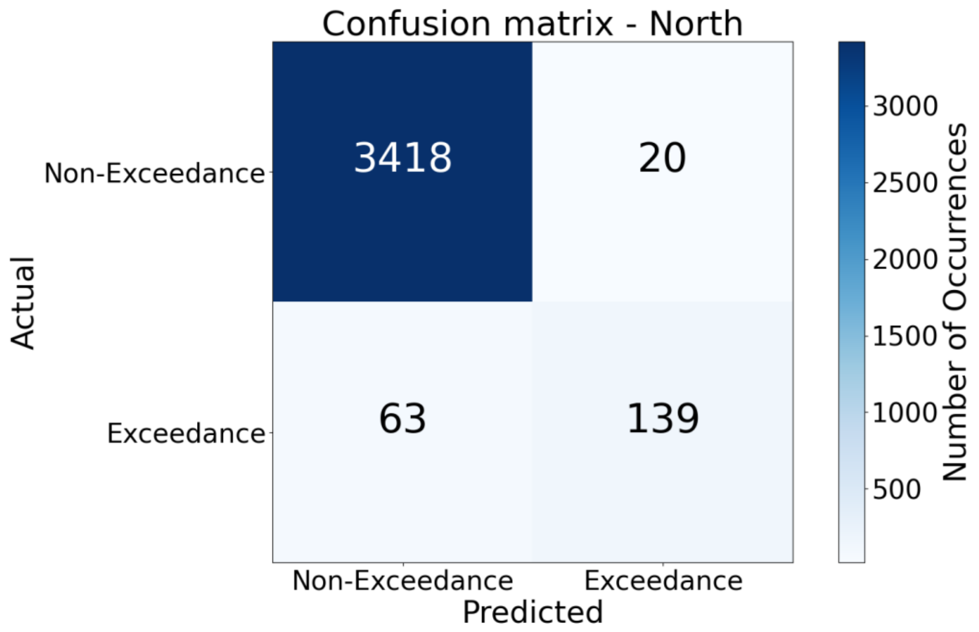


MLP

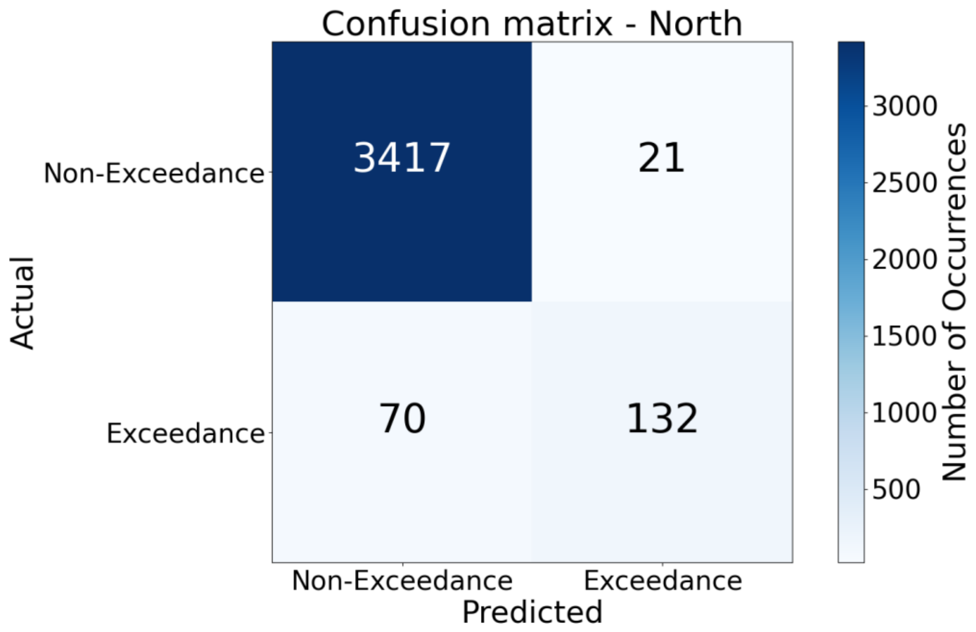


DL-LSTM

Results – Confusion Matrixes

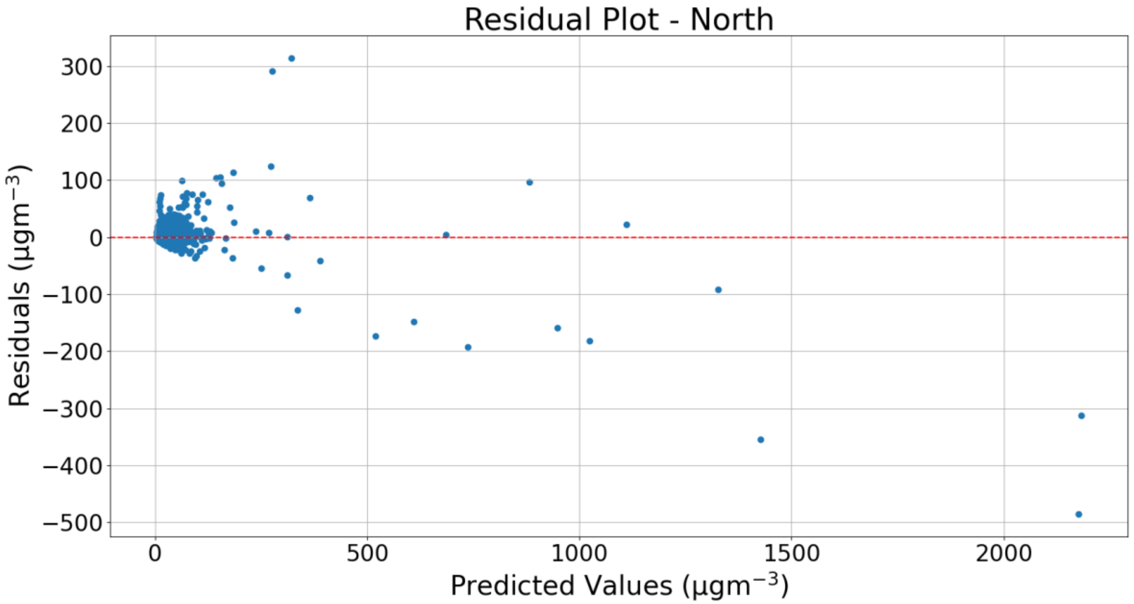


MLP

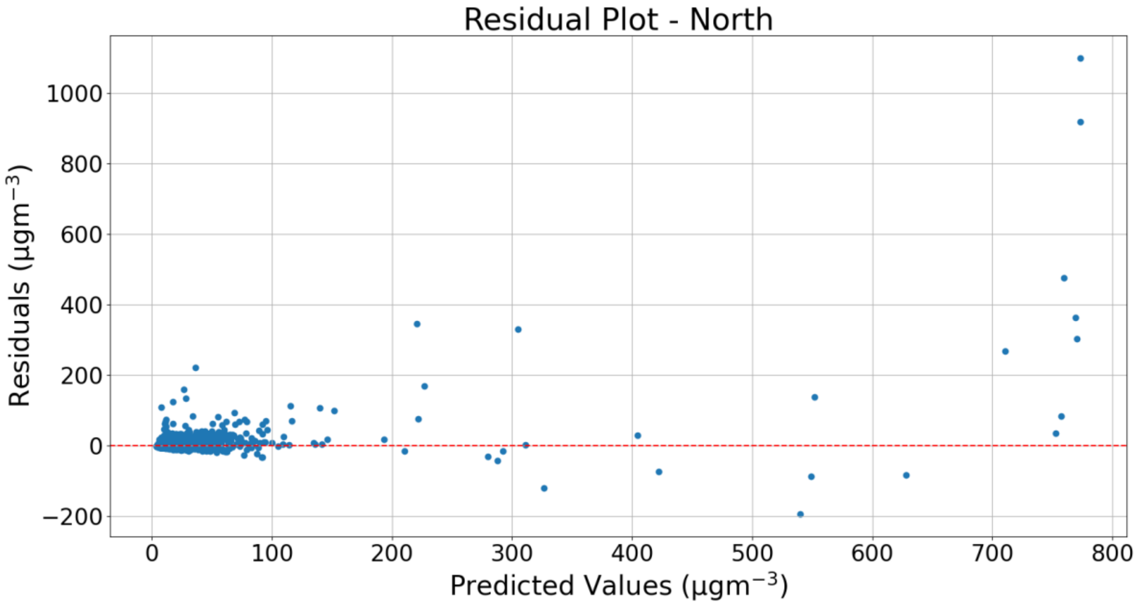


DL-LSTM

Results – Residual Plots



MLP



DL-LSTM

Results – Model Comparison

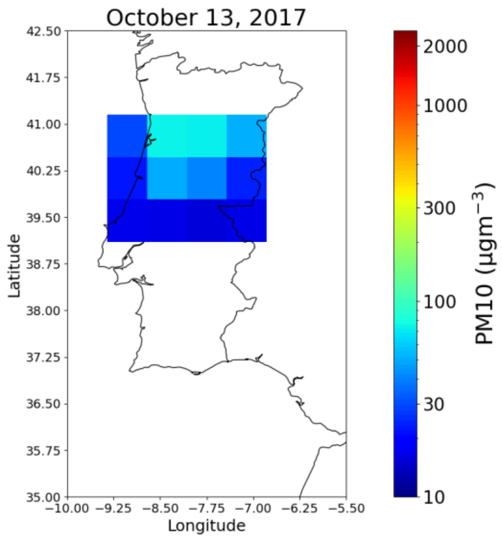
NORTH – MLP	
R	0.97
RMSE	17.73
MAE	5.36
MAPE	31.80%
ACCURACY	0.98
PRECISION	0.86
RECALL	0.68
F1-SCORE	0.76

MLP

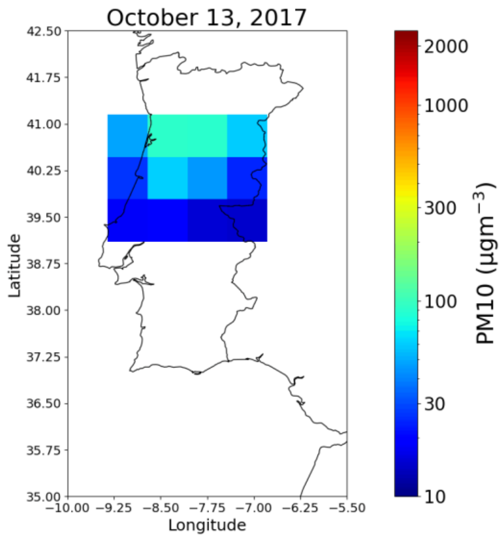
NORTH – LSTM	
R	0.93
RMSE	30.39
MAE	5.54
MAPE	30.57%
ACCURACY	0.97
PRECISION	0.85
RECALL	0.64
F1-SCORE	0.73

DL-LSTM

Case study – October 2017 (MLP)



MLP Predictions

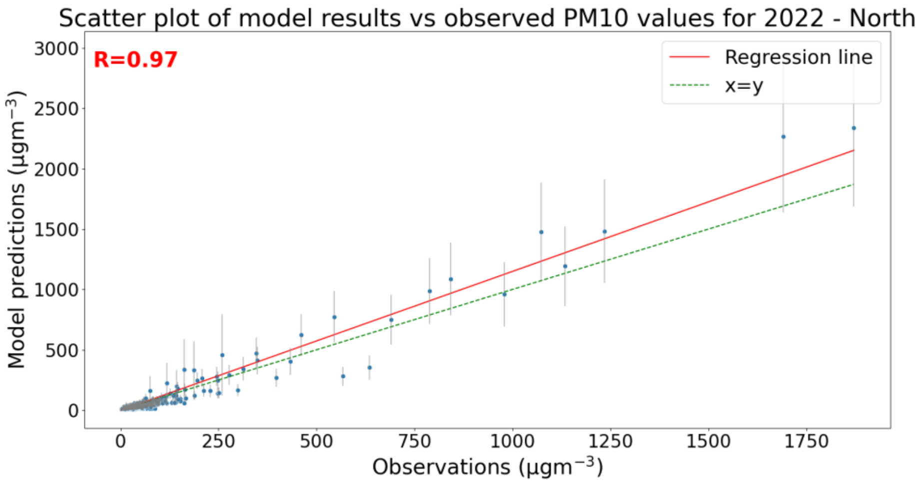


Observations

Conclusions and next steps

Conclusions

- Both models present **high** effectiveness in predicting PM10 levels one-day ahead.
- The **MLP** model exhibits lower error rates.
- Tendency to underestimate higher values.



Next steps

- Development of **Bayesian Neural Networks** to introduce uncertainty in the predictions.
- Creation of an operational tool to help authorities during AQ emergencies.

Thank you!