MLP based models to predict PM_{10} and O_3 concentrations, in Sines industrial area

R. Durão[1], M. J. Pereira[1] and M. mendes [2]

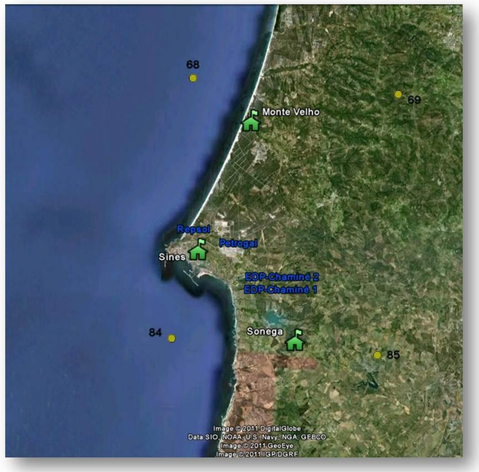
[1]CERENA, [2]IM

IST/UTL - Instituto Superior Técnico, Lisboa, Portugal; Instituto de Meteorologia

1. OBJECTIVES

The aim of this study is to develop prediction models of pollutant concentration categories (e.g. low and high concentrations) in Sines industrial area.

2. DATA



Continuous industrial emissions of Sulfur Dioxide (SO_2), Nitrous oxides (NOx), Total Suspended Particles(TSP)

Meteorological (ECMWF): Temperature, Relative Humidity, Wind (U, V components) Precipitation, Sunshine duration, Boundary layer height , Weather types

Air Quality concentrations: Ozone (O_3), Particulate matter (PM_{10}), Nitrous oxide (NO_2), Nitrous oxides (NOx) and Sulfur Dioxide (SO_2)

Time of Year - sinusoidal and cosinusoidal variables of the "day of the year"

- The data set consited of hourly mean values corresponding to the 4-year period 2006-2009.

All data values were standardized to zero mean and unity deviation.

3. METHODOLOGY

Input Data → C&RT → MLP (Training, Validation Generalization) → Prediction of categories

I. Classification and Regression trees (C&RT) applied to select the best predictor variables in each station.

C&RT is based on binary recursive partitioning. Each predictor is examined and the data set is split into 2 groups based on the predictor value that maximizes the dissimilarity between groups. The tree grows by exhaustively searching the predictors at each branch for the best split.

II. Multi-layer perceptron (MLP) network developed for each station through BFGS quasi-Newton algorithm.

The input layer of each NN is consisted by the best predictors, whereas the output layer is the pollutant concentrations (High-Low class) for the next day.

MLP performance values are obtained using cross entropy error functions. This error functions are exclusively for classification problems and ensure that the network outputs are true class membership probabilities, improving NN classification performances.

4. C&RT BEST PREDICTORS

Input Variables	Stat.1_O3	Stat.2_O3	Stat.3_O3	Stat.1_PM10	Stat.2_PM10
Wtype	x	x	x	x	x
BLH_Bin				x	
Prec_Bin					
Temperature (º)		x		x	x
Relative Humidy (%)	x	x	x		
U wind component (m/s)	x	x		x	x
V wind component 8m/s)	x				
Sunshine Duration. (s)	x	x	x	x	x
Temperature_84level (º)					x
U velocity_84level (m/s)					x
V velocity_84level (m/s)					
Vertical Veloc_84level(Pa/s)		x	x	x	x
TOY_Cos	x				
TOY_Sen		x	x		x
NOx_emission(mass flow)					
SO2_emission(mass flow)				x	
PTS_emission(mass flow)		x			x
NOx_air quality		x	x		
NO2_air quality					
SO2_air quality					
PM10_air quality					
O3_air quality					

Acknowledgments: The authors would like to acknowledge the GISA Project partners; FCT: SFRH / BD / 64348 / 2009

5. THRESHOLDS

Percentile	O_3 (µg/m³)_Station 1	PM_{10} (µg/m³)_Station 1
98	118	73
95	105	56
70	77	31
50	64	23
20	35	13
10	20	9

High Frequency of Low concentrations → Decreasing of Exceedences → Limit Value = Percentiles Data

- Health effects due to long-term exposures to low concentrations (*WHO, 2000*)

6. MLP RESULTS

Station/Pollutant	Percentile	MLP	Performance
Station 1_O3	70 = 77 µg/m3	5-45-2	87%
Station2_O3	60 = 68 µg/m3	18-39-2	64%
Station3_O3	50 = 55 µg/m3	16-5-2	74%
Station 1_PM10	50 = 25 µg/m3	52-25-2	69%
Station2_PM10	80 = 38 µg/m3	37-8-2	66%

7. DISCUSSION

I. **C&RT** identified the best predictors affecting O_3 and PM_{10} concentrations. Mainly dependent on industrial emissions, the time of the year and a combination of meteorological factors (wind speed and direction, boundary layer height, temperature, sunshine duration, relative humidity, weather type).

II. **MLP** models showed performances between 0.64 and 0.87, indicating a reasonable accuracy for models development and generalization capability.

8. FINAL REMARKS

The present work addresses an unusual approach to pollutant concentration prediction methodology :

i. **Mass flow** of industrial emissions considered as input variable

ii. **C&RT** methodology to find the best concentration predictors

iii. **Regional thresholds** based on the concentration data percentiles (WHO guidelines)

iv. **MLP models** exhibit a good emission-monitoring classification prediction 24h in advance in the study area.