A Low Cost Forecasting Framework for Air Pollution

Ilias Bougoudis[1], Konstantinos Demertzis[2] and Lazaros Iliadis[2]

[1]Institute of Environmental Physics, University of Bremen, Germany
[2]School of Engineering, Department of Civil Engineering, Democritus University, Greece

RTG 2224

Motivation

- Atmospheric pollution is one of the most serious problems of the modern way of life
- 60,000 deaths each year in Europe, 3,000,000 worldwide due to long-term air pollution exposure
- To study it we used hourly in-situ measurements from a station located at the city centre of Athens, Greece
- The table below summarizes the statistical analysis performed on 89364 records, from 2000 to 2012

Table 1: Statistical Analysis of the Whole Dataset

2000 – 2012 89364	CO (mg/m³)	NO (μg/m³)	NO_2 (μg/m³)	O_3 (μg/m³)	SO_2 (μg/m³)
Max	21.4	908	377	253	259
Min	0.1	1	1	1	2
Mode	0.8	8	60	3	2
# Mode	4941	2358	1508	6649	10451
Average	1.83	59.31	63.57	32.94	9.64
Std	1.48	90.06	27.11	28.67	9.39

Forecasting and Clustering

- Our main goal is to create a **cheap** and **fast** system to predict and monitor air pollution
- Using **multi layer feed-forward neural networks** to **forecast** pollutants with temporal and meteorological inputs
- The network predicts accurately pollutant values when other pollutants are used as inputs, it does not when they are excluded

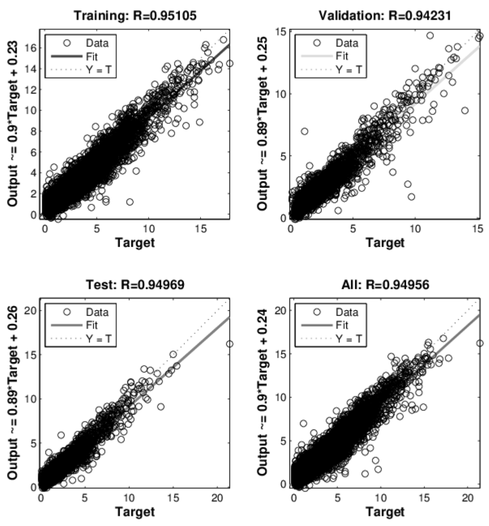

Figure 1: CO prediction **with** other pollutants as inputs

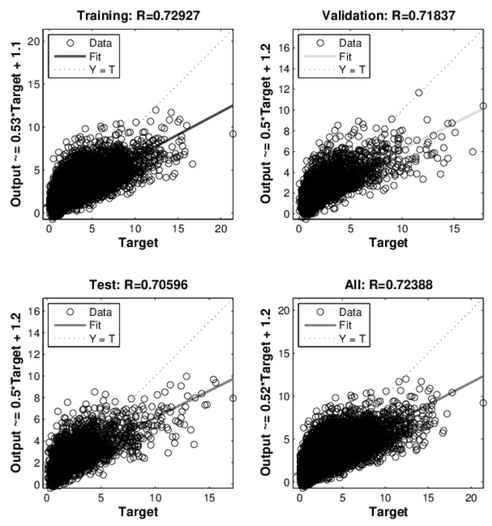

Figure 2: CO prediction **without** other pollutants as inputs

- As only **extreme** pollutant cases are important, we used **unsupervised learning** to **cluster** our records
- A novel **fuzzy semi – supervised** approach was selected
  - Uses pre-assigned classes of some records to cluster the whole dataset
  - Includes fuzzy logic

The Formula Forecast Framework

- Two extreme clusters: One with high CO, NO, NO_2, SO_2 values and one with high O_3 values
- 1st extreme cluster: Winter, early morning, low airtemp
- 2nd extreme cluster: Summer, mid-afternoon, high airtemp
- The statistical analysis of this **extreme dataset** (clusters 1 & 2) is shown on table 2 (49638 records in total)

Table 2: Statistical Analysis of the Extreme Dataset (inside the parentheses are the values for extreme cluster 2, outside for extreme cluster 1)

2000 – 2012 21635 (28003)	CO (mg/m³)	NO (μg/m³)	NO_2 (μg/m³)	O_3 (μg/m³)	SO_2 (μg/m³)
Max	21.4 (3.1)	908 (33)	377 (153)	139 (253)	259 (35)
Min	0.1 (0.1)	1 (1)	4 (3)	1 (11)	2 (2)
Mode	2.7 (0.7)	52 (7)	76 (34)	3 (49)	9 (2)
# Mode	756 (3357)	159 (2109)	444 (787)	4908 (514)	979 (5942)
Average	3.62 (0.86)	160.5 (9.7)	88.5 (42.4)	9.8 (63.1)	19.1 (6)
Std	1.91 (0.38)	133.8 (5.3)	28.8 (15.3)	12.4 (25.7)	13.5 (4.4)

- Using only **Year**, **Month**, **Day**, **Hour**, **AirTemp** and **Cluster_Id** (1 or 2) as inputs in the extreme dataset (**Formula**)

Comparative Results – Future Work

- We trained our neural networks (one for each pollutant) with data from 2000 - 2012
- The tables below show the comparison between the formula and the initial forecast (with pollutants as inputs)

Table 3: Forecast Comparison for 2000 - 2012 (inside the parentheses is the number of records in each case)

2000 – 2012	CO (mg/m³)		NO (μg/m³)		NO_2 (μg/m³)		O_3 (μg/m³)		SO_2 (μg/m³)	
	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE
All (89634)	0.94	0.5	0.95	27.63	0.87	12.93	0.93	10.23	0.79	5.52
Formula (49638)	0.9	0.81	0.88	52.4	0.91	12.9	0.94	10.7	0.87	5.37

- We also used 2013 as evaluation for the neural networks

Table 4: Forecast Comparison for 2013 (inside the parentheses is the number of records in each case)

2013	CO (mg/m³)		NO (μg/m³)		NO_2 (μg/m³)		O_3 (μg/m³)		SO_2 (μg/m³)	
	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE
All (8144)	0.93	0.37	0.96	21.11	0.89	15	0.93	10	0.28	3.18
Formula (5059)	0.88	0.54	0.91	31.43	0.7	13.5	0.83	15.48	0.35	3.34

- Future work will include:
  - Implementation of the formula in cities with different climate conditions
  - Prediction of other trace gases (e.g. PM_{10}, $PM_{2.5}$)
  - Effort to link the formula to global warming and climate change

[1]Bougoudis, I., et al. (2015, October). Fast and Low Cost Prediction of Extreme Air Pollution Values with Hybrid Unsupervised Learning. In Integrated Computer Aided Engineering (Vol. 23, No. 2, p. 115 - 127). IOP Publishing
[1]Bougoudis, I., et al. (2017, June). FuSSFFra, a fuzzy semi-supervised forecasting framework: the case of the air pollution in Athens. In Neural Computing and Applications (Vol. 28, No. 6). Springer Publishing

DFG Universität Bremen iup