

# Hybrid forecasting of PM2.5 using SOFM and ELM

Krzysztof Siwek<sup>1,\*</sup>, Tomasz Grzywacz<sup>1</sup>

IETiSIP, Faculty of Electrical Engineering, Warsaw University of Technology, Warsaw, Poland

**Abstract.** The article presents a new approach to atmospheric PM2.5 dust prediction using an Extreme Learning Machine (ELM) neural network with clusterization done by the Self Organizing Feature Map (SOFM). This work is concerned with the calculation of the average level of air particulate matters PM2,5 in Warsaw's Ursynow one day ahead. The brief description of the hazards posed by air pollution is included. The work presents a short description of the SOFM and ELM networks, and their hybridized system used as a prediction tool. The analysis of the obtained results was presented and discussed.

## 1 Problem statement

We have observed growing pollution of the natural environment in recent years. Water, soil and atmospheric air are polluted. Observing the information provided by the media, and these available directly from the measurement stations, we know this is a serious and global problem.

As a result of civilization progress and industrial development, the impact of humans activities on the natural environment is growing and, unfortunately, in most cases unfavourable for it. The air pollution caused by dust influences human life and health, and can adversely affect the length and quality of human life [1,2,3].

### 1.1 Particulate matters pollutions

Particulate matters in the atmospheric are classified due to its diameter; the most important are:

- PM10 - dust with a diameter smaller than 10 $\mu$ m, which can reach the upper respiratory tract and lungs,
- PM2.5 - dust with a diameter of less than 2.5 $\mu$ m that penetrates deep into the lungs and enters the bloodstream [2].

The suspended dust consists of an undefined mixture of particles of inorganic and organic substances. It may contain a variety of toxic substances, for example polycyclic aromatic hydrocarbons - benzo(a)pyrene and benzo(e)pyrene, dioxins, furans, heavy metals and many other substances. Air pollution is absorbed by people mainly during breathing and contributes to the development of respiratory and allergic diseases, cancer, as well as heart and blood vessel diseases.

Sources of atmospheric air pollution are of two types: natural and anthropogenic. Natural sources include, among others: volcanic eruptions, wind erosion of rocks, dust blown from deserts, forest and steppe fires.

The second category includes air pollution resulting from human activities: chemical conversion of fuels

(combustion), mining, chemical industry, raw and waste landfills, and transport.

Air pollution is usually derived from two sources: combustion of poor quality of coal in old, ineffective boilers and domestic furnaces, and from communication sources (engine exhausts) and the street dust carried from under cars. The exceeding the norms can be serious and good forecasting can help to reduce air pollutions.

## 2 Database of PM

The measurement data used for experimental research were from the WIOS (Warsaw Inspectorate for Environmental Protection) measurement station located in the residential Ursynow district in Warsaw, Poland measured in the years 2011-2013.

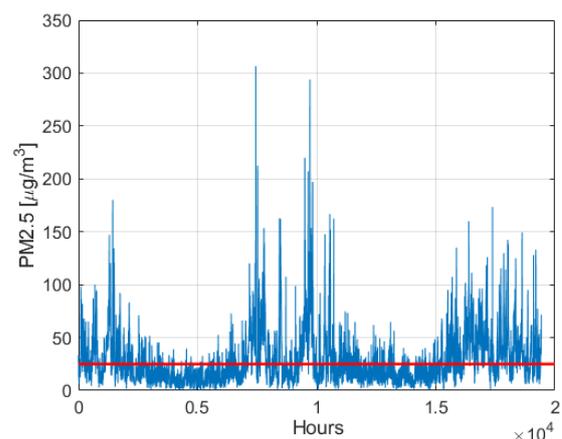


Fig. 1. PM2.5 measured in Warsaw-Ursynow in 2011-2013.

The measurements of PM2.5 are carried out every hour. Additionally the weather conditions (temperature, strength and direction of the wind, humidity, pressure) were recorded at the same time.

\* Corresponding author: [Krzysztof.Siwek@ee.pw.edu.pl](mailto:Krzysztof.Siwek@ee.pw.edu.pl)

### 3 The forecasting system

#### 3.1 The Self Organizing Feature Map (SOFM)

The Self Organizing Feature Map is unsupervised and competitive learning algorithm forming the topological maps, in which the relations of spatially correlated input data in high dimensional space are converted into spatial relationships among the neurons in one or two dimensional grid. In such way the SOFM represents the centres of input data and divides them to separate clusters [4].

#### 3.2 The Extreme Learning Machine (ELM)

The Extreme Learning Machine is a relatively new structure of one-way neural networks, usually with one hidden layer of neurons with non-linear sigmoidal activation function [5]. The main advantage of this neural network is the extremely short learning time, based on random selection of hidden layer weights and calculation of the output layer weights using the SVD decomposition method [6].

#### 3.3 The hybrid forecasting system

The proposed two-stage hybrid system combines the ELM prediction with the SOFM clustering algorithm. In the first stage, SOFM divides the data into clusters. Each cluster contains similar input data. In the second stage, an individual ELM network for each cluster is build. In that way the individual ELM networks are trained only using similar input vectors. In the testing mode the membership to the specific cluster is computed by SOFM firstly, then ELM, associated with this cluster, calculates the final prognosis of the mean value of PM2.5 for the next day.

### 4 The results

The quality of prediction was assessed using a measure of errors made by the network:

- Average absolute percentage error MAPE

$$MAPE = \frac{1}{n} \sum_{j=1}^n \left| \frac{P(j) - \hat{P}(j)}{P(j)} \right| \cdot 100\% \quad (1)$$

- RMSE root mean square error

$$RMSE = \sqrt{\frac{1}{p} \sum_{i=1}^p |d_i - y_i|^2} \quad (2)$$

The input vectors formed from the process features, were the same in each network: 24 previous values of PM2.5 of the following hours preceding the forecast, weather forecast for the next day (temperature, wind power, humidity, pressure – each 24 values), type of the day and season code, totally 171 values. Only those features that are strongly correlated with the destination (the absolute value of Pearson's correlation coefficient above 0.5) were presented to the network input. ELM

networks has calculated the mean values of PM2.5 for the next day.

**Table 1.** Results of numerical experiments.

Number of clusters	MAPE [%]	RMSE [ $\mu\text{g}/\text{m}^3$ ]
5	28.14 ± 1.62	11.67 ± 1.22
15	24.94 ± 1.33	10.42 ± 0.88
30	<b>23.43 ± 1.29</b>	<b>10.01 ± 0.81</b>
50	25.58 ± 1.39	10.34 ± 0.94
100	26.99 ± 1.40	10.34 ± 0.98

Due to the random initial weight values in the SOFM and random weights generated during learning for the ELM network, 10 learning and testing processes were carried out for each network. Available data of pollution was divided into two randomly selected sets: teaching and testing in 70:30 proportions. After each learning process, MAPE and RMSE errors were calculated, and on their basis the average values of these errors and their standard deviations were computed. Table 1 presents the obtained results for different number of clusters (number of neurons in SOFM).

### 5 Conclusions

The paper has presented fast hybrid method for PM2.5 forecasting. The complexity of the PM2.5 prediction problem results from the dependence of the local level of atmospheric air pollution on the local weather conditions, which change very quickly and accidentally. The next step of performing a more accurate prediction will be the aggregation of the results of many forecasts calculated using various methods - neural and classical. An interesting approach is also the use of selection of features based on genetic algorithms and not only the Pearson linear correlation coefficient [1].

The numerical results presented in the paper have confirmed efficiency of the proposed method.

### References

1. K. Siwek, S. Osowski, Int. J. Appl. Math. Comput. Sci. **26** (2), 467–478 (2016)
2. R. Ruckerl, A. Schneider, S. Breitner, J. Cyrus, A. Peters, Inhalation Toxicology, **23** (10), 555–592 (2011)
3. European Environment Agency (EEA) raport, Air quality in Europe (2016)
4. T. Kohonen, *Self organization and associative memory*, 3rd ed., (Springer-Verlag, Berlin, 1989)
5. G. B.Huang, Q. Y. Zhu, C. K. Siew, Extreme Learning Machine: Theory and applications, 1-10 (2016)
6. G. B. Huang, Q. Y. Zhu, C. K. Siew, Extreme learning machine: a new learning scheme of feedforward neural networks, IEEE International Joint Conference on Neural Networks, Budapest, Hungary (2004)