

# Prediction of PM10 Concentrations using a Modular Neural Network System and Integration with an Online Air Quality Management System

Ioannis Kapageridis<sup>1</sup>, Vasilios Evagelopoulos<sup>2</sup>, Athanasios Triantafyllou<sup>2</sup>

<sup>1</sup>Laboratory of Mining Information Technology and GIS Applications /  
Technological Educational Institute of Western Macedonia  
ikapa@airlab.teikoze.gr

<sup>2</sup>Laboratory of Air Pollution and Environmental Physics /  
Technological Educational Institute of Western Macedonia  
evagelopoulos@airlab.edu.gr atria@airlab.edu.gr

**Abstract.** The development and application of Integrated Air Quality Management Systems, that increase the awareness of the local population and lead to the enforcement of measures for the avoidance of air pollution episodes, is a subject of great interest. The development of such a system is crucial for Kozani, the most populated and industrialized area of western Macedonia in Greece, where a number of power stations are operated using lignite from nearby mines. The industrial and urban activities aggravate the air quality in the area, resulting in high PM10 concentrations. This paper presents a PM10 concentration prediction method based on a modular neural network. The neural network is integrated with an online air quality monitoring system and the results of current versus predicted air quality are available to the public through the internet.

## 1 INTRODUCTION

The level of particulate matter (PM) has been of concern in the area of Kozani in northern Greece as several studies confirmed that these particles may induce severe effects on public health [7, 3]. The town of Kozani is located in the southern part of the Eordea basin and is the centre of significant industrial activity. A number of lignite power stations operate within the basin. It has been shown [8] that under certain atmospheric conditions, pollutants emitted by these power stations reach the town of Kozani. Urban pollution sources also contribute to the problem of air quality in the town.

The air quality problem described above was the main motivation for the development of an Integrated Air Quality Management System (IAQMS) at the Laboratory of Air Pollution and Environmental Physics of the Technological Educational Institute (TEI) of Western Macedonia [9]. In 2007, a modular neural network (MNN) system for the prediction of 24h average PM10 concentration was added to the IAQMS. Prior to integration with the IAQMS, the MNN received initial training using data between 1 June 2004 and 31 May 2005 (12 months). The MNN was developed to receive several inputs from air pollution and environment monitoring stations at the center of Kozani.

The modular structure of the system allows it to better approach the three different groups of data - PM10 measurements, current and forecasted meteorological conditions. Network training is repeated regularly using updated information in order to improve its prediction capacity along time and to better approach seasonal variations of air pollution. Previous states of the system are stored for further analysis and validation. The MNN is used to predict the 24-h average PM10 concentration for the next day at the city center of Kozani.

For each day considered, eight hourly PM10 measurements are taken every three hours (PM03, PM06, PM09, PM12, PM15, PM18, PM21 and PM24). The average PM10 concentration of the last 24 hours is also calculated (AVGPM). A number of meteorological parameters for the last 24 hours are taken into account including the minimum and maximum 1-h average relative humidity (MINRH, MAXRH), the maximum temperature (MAXTMP), the temperature range (TMPRNG), the average wind direction and speed (AVGWD, AVGWS). Forecasted meteorological parameters for the following 24 hours are also considered, consisting of the minimum and maximum 1-h average of relative humidity (MINRHF, MAXRHF), the forecasted maximum temperature (MAXTMPF), the difference between the maximum and minimum temperature (TMPRNGF), and the average wind speed (AVGWSF). All these measured parameters and forecasts are used as inputs to the MNN system. The required prediction output is the 24h average PM10 concentration for the next day obtained at 06:00h of the following day. The choice of inputs was motivated by similar studies [5, 2].

## **2 MODULAR NEURAL NETWORK STRUCTURE**

The MNN system is based on Radial Basis Function (RBF) networks. The networks are arranged in two levels (Figure 1). The first level consists of three networks, each receiving different inputs and each producing a forecast of next day 24h average PM10 concentration. The first receives the eight PM10 hourly measurements, the second network receives meteorological inputs (observed values) regarding the last 24 hours, and the third network receives meteorological inputs (forecasts) regarding the following 24 hours. The outputs from the three networks of the first level are directed as inputs to the single RBF network of the second level. This network combines the outputs from the three previous networks into one forecast of next day's average PM10 concentration.

The philosophy of this structure is that by decomposing the prediction problem into three separate smaller problems, a better forecast may be obtained. The separate RBF networks can build a model between inputs and output easier and faster, requiring a smaller number of training patterns, as only a subset of the total inputs are used in each of them leading to a smaller number of network parameters to fix through training. The overall system is less dependant on each of the three different groups of inputs. It is also easier to examine the sensitivity of the system's output to each of the inputs. In the future, this modular structure will allow prediction even when one of the RBF networks of

the first level is unable to produce an output due to unavailability of one or more of the inputs for various reasons (lack of meteorological forecasts, measuring equipment breakdown, etc.)

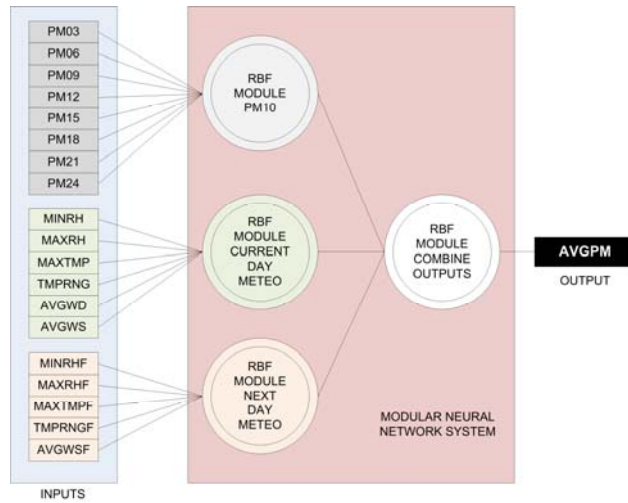


Figure 1: Structure of the modular neural network system.

### 3 MODULAR NEURAL NETWORK DEVELOPMENT

As discussed before, the MNN was developed using data spanning over 12 months. This time period provided a total of 2603 patterns (approximately equal to 12 months x 30 days x 8 time periods) separated by 3 hours, i.e. complete sets of all 19 inputs and 1 output. These were split equally in two subsets, one used for training and one used for validation. The generation, training and testing of the modular neural network during development was achieved using a commercial package (NeuroSolutions by NeuroDimension Inc.) Training of the network included fixing the number of radial basis functions, function centres, function widths, and output weights, for each of the sub-networks. Figure 2 shows the performance of the network on the validation set.

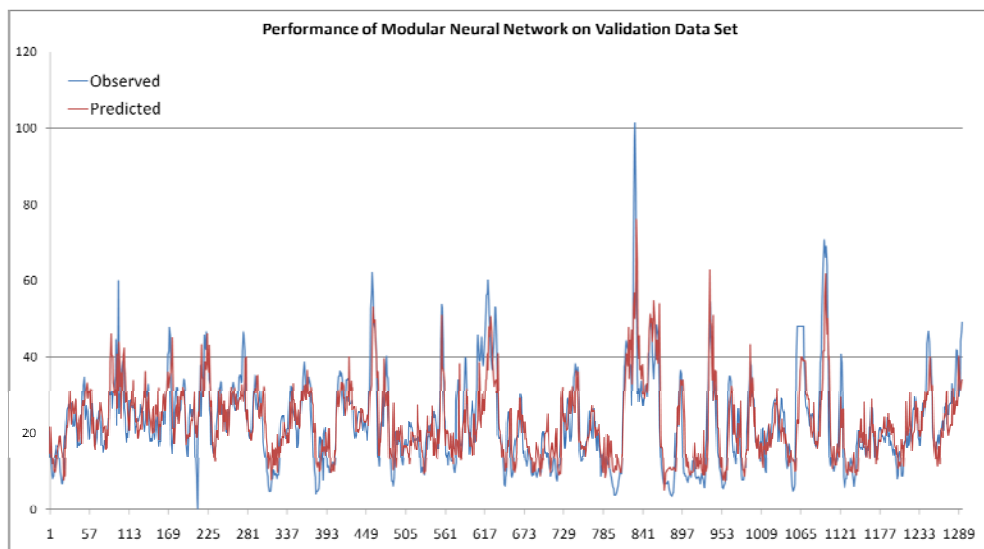


Figure 2: Target values and network outputs from the validation data set.

The network performed particularly well on the validation set (coefficient of correlation,  $\rho = 0.78$ ). The developed network was examined as to the derived complexity of the RBF sub-networks (number of RBFs, function center locations and widths) and their sensitivity to each of their inputs. Figure 3 presents in graphical form the variation of each sub-network's output to pairs of inputs. All outputs shown are normalised in a scale between 0 and 1.

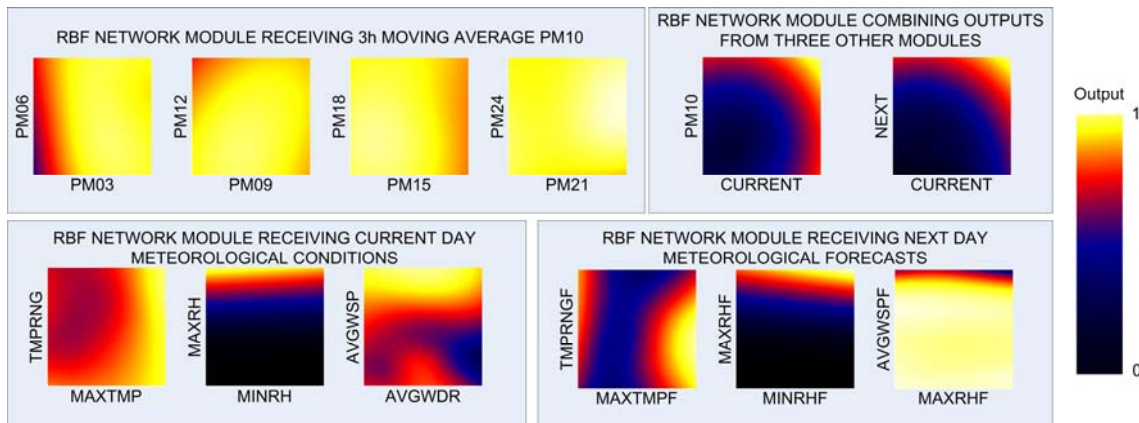


Figure 3: Sensitivity of RBF networks' output to each of the inputs after development.

One of the conclusions drawn here is that the MNN output is not affected by the minimum relative humidity inputs (MINRH and MINRHF). Relative humidity is one of the meteorological parameter that influences PM10 concentration in ambient air, as it was indicated by many studies [1, 4]. Precipitation and relative humidity largely removes pollutants from the atmosphere [6]. However, minimum relative humidity has a weak inverse correlation with PM10 [10] and this is demonstrated by the low sensitivity of the MNN output to both MINRH (measured minimum relative humidity of the past 24-h) and MINRHF (forecasted minimum relative humidity for the following 24-h). This weak inverse correlation is confirmed by the training and validation set, where the 24-h average PM10 concentration has a -0.108 correlation to the minimum relative humidity of the past 24 hours and a -0.099 correlation to the forecasted minimum relative humidity for the next 24 hours.

#### 4 APPLICATION IN IAQMS

The developed system was converted to computer code (a DLL application extension) that was integrated with the IAQMS in August 2007. Since then, it has been used to forecast the 24h average PM10 concentration in Kozani. The daily forecasts become available through the Laboratory's website (<http://www.airlab.edu.gr>). Figure 4 shows how the network's forecasts compare with the observed concentrations. The coefficient of correlation between observed and predicted values is lower (0.52) compared to the validation set used during training of the network. The validation data values (Figure 2) were possibly easier to approach by the MNN as they were more uniform than the ap-

plication data in Figure 4. The mean absolute error was 13.7% on the validation set and 26.3% during application.

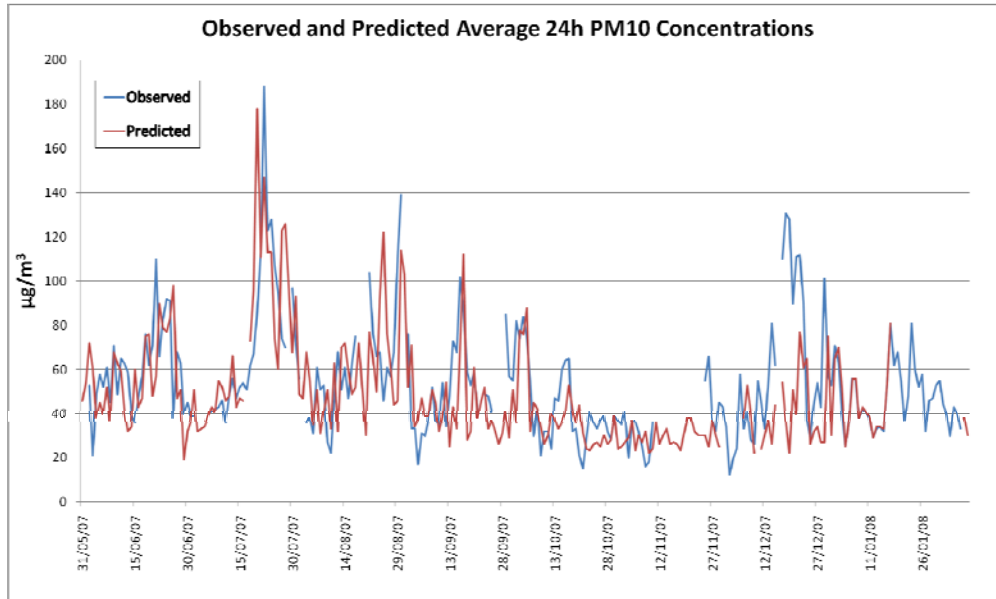


Figure 4: Line chart showing observed and predicted 24h average PM10 concentrations.

The predicted values were converted to air quality index values using the index limits for PM10. Figure 5 shows the histograms of the observed and predicted index values and a cross tabulation matrix for better comparison. These should be examined together with Figure 4 as there are specific time periods where either the network did not produce a forecast such as the low PM10 period at the beginning of December 2007, or there were no observed values (e.g. November 2007), both due to various technical problems.

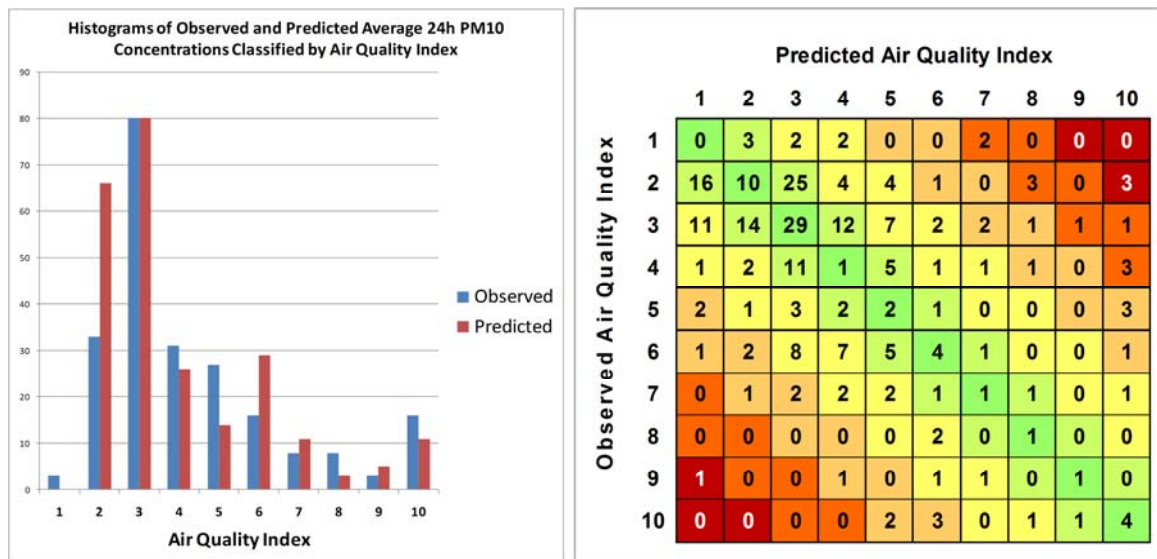


Figure 5: Histograms and cross tabulation matrix of observed and predicted 24h average PM10 concentrations converted to respective air quality index values.

## 5 CONCLUSIONS

This paper presented a modular neural network system based on RBF networks for the prediction of 24h average PM10 concentrations. The development of the system and its integration with an Air Quality Management System was explained. The predicting capacity of the presented system has been demonstrated using observed and predicted concentrations collected since it became operational as part of the IAQMS. Further improvement of the system will include the development of more representative training datasets and developing the capability to generate a forecast even when one the sub-networks cannot due to lack of input data.

## 6 REFERENCES

- [1] P. Alpert, Y.J. Kaufman, Y. Shay-El, D. Tanré, A.S. Da Silva, S. Schubert and J.H. Joseph. Quantification of dust-forced heating of the lower troposphere. *Nature*, 395, 367–370, 1998.
- [2] I.K. Kapageridis, and A.G. Triantafyllou. A Genetically Optimised Neural Network for Prediction of Maximum Hourly PM10 Concentration. In: 12th International Conference on Modelling, Monitoring and Management of Air Pollution, Wessex Institute of Technology, Rhodes, pp. 161-170, 2004.
- [3] J.E. Kelsall, J.M. Samet, S.L. Zeger, and J. Xu. Air pollution and mortality in Philadelphia, 1974-1988. *Am J Epidemiol*, 146(9):750-762, 1998.
- [4] C. Monn. Exposure assessment of air pollutants: A review on spatial heterogeneity and indoor/outdoor/ personal exposure to suspended particulate matter, nitrogen dioxide and ozone. *Atmos. Environ.*, 35, 1-32, 2001.
- [5] P. Perez, and J. Reyes. Prediction of maximum of 24-h average of PM10 concentrations 30h in advance in Santiago, Chile. *Atmospheric Environment* 36: 4555-4561, Elsevier, 2002.
- [6] J. Seinfeld, and N. Spyros. *Atmospheric chemistry and physics from air pollution to climate change*. New York, John Wiley & Sons, Inc, 1998.
- [7] J. Schwartz, D.W. Dockey, and L.M. Neas. Is daily mortality associated specifically with fine particles? *Journal of Air Waste Manage Association*; 46: 927-939, 1996.
- [8] A.G. Triantafyllou. PM10 pollution episodes as a function of synoptic climatology in a mountainous industrial area. *Environmental Pollution* 112: 491-500, Elsevier, 2001.
- [9] A.G. Triantafyllou, V. Evagelopoulos, and S. Zoras. Design of a web-based information system for ambient environmental data. *Journal of Environmental Management* 80:230–236, Elsevier, 2006.
- [10] Multiple Pollutants and Risk of Cardiac and Respiratory Emergency Department Visits in Atlanta - Air Quality Data. Annual Progress Report to Emory University, under USEPA contract , 2002.