A Genetically Optimised Neural Network for Prediction of Maximum Hourly PM10 Concentration

I. Kapageridis[1], A.G. Triantafyllou[2]
[1] *Maptek/KRJA Systems Ltd, United Kingdom.*
[2] *Laboratory of Atmospheric Pollution and Environmental Physics, Technological Education Institute of West Macedonia, Greece.*

Abstract

Concentrations of ambient air particles have been found to be associated with a wide range of effects on human health. PM10 concentrations are usually used as a standard measure for air pollution. Increase in the level of PM10 has been associated with increases in mortality and cardio respiratory hospitalisations. Therefore, prediction of ambient levels in certain environment is of great importance, especially in urban and industrialised areas. The present work aims to develop an adaptive system based on Artificial Neural Networks (ANN) that will allow the prediction of the maximum 24-h moving average of PM10 concentration. A special ANN architecture is employed, the Time Lagged Feed forward Network (TLFN), with genetically optimised topology and learning parameters. This type of network is able to process information over time and produce time-varying nonlinear mappings from the chosen input variables to the predicted value. The network is trained and testified by hourly data collected at two air pollutant – monitoring stations in an urban and nearby industrial location in northern Greece. The initial study presented in this paper involves a small subset of the available data that were used to validate the approach and the chosen ANN architecture.
Keywords: PM10 concentration, prediction, time lagged feed forward neural network, genetic optimisation.

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 [5, 10]. Kozani is the most populated and industrialised area of West Macedonia. The town is located in the southern part of the Eordea basin and is the centre of significant industrial activity. A number of lignite power stations are operated within the basin (Figure 1). It has been shown [11, 13, 14], 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 [12]. The prediction of maximum hourly PM10 concentrations can be beneficial in the efforts to monitor and control the effects of PM10 on the health of the local population.

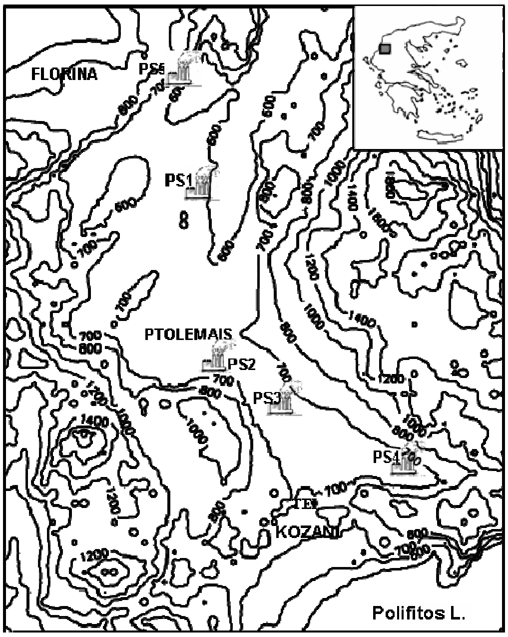

Figure 1: The topography of Eordea basin greater area, showing the location of Kozani and the pollutant point sources (PS).

1.1 Data Sources

The PM10 concentration measurements used in this study were carried out in the commercial centre in Kozani and are referred to a three month period, from 8/02 to 10/02. The PM10 measurements were made using a beta absorption (FAG FH 62 I-N) monitor having a specific sampling head for PM10 [13]. This monitor is based on the principle of β-ray absorption by particles, sampled through the instrument and collected on a fibreglass filter tape. Dust concentration calculations are performed automatically and displayed by the instrument. Automatic zero was provided for each measurement cycle by obtaining the beta absorption on the same portion of the filter tape before and after the sample was deposited. A sequenced calibration foil of known absorption was used for span check. The PM10 data were collected with a 1-hr time resolution. The meteorological data used have been taken from the meteorological station located in the same place.

1.2 Data Analysis

The PM10 concentrations were at relatively high levels in all studies in the area of Kozani [11, 12, 13, and 14]. Each yearly average was found to be higher than the US Environmental Protection Agency (EPA) limit value. There was some variation in the highest mean seasonal concentrations between warm and cold periods. During a 24-hr period, the mean diurnal variation showed relatively high concentrations, which exceeded the value of 50μg/m^3 during the whole day. A complex system of PM sources and meteorological conditions modulates the levels of particulate pollution. A synoptic climatology approach showed that the highest concentrations were associated with stagnant conditions, when the accumulation of pollutants due to urban activities and power plants results in very high concentrations. The wind speed was also associated with high concentration episodes [12].

1.3 Data Processing

As mentioned above, we consider measurements of hourly averages of PM10 concentrations obtained from the commercial center of an urban and nearby industrial location. These measurements were obtained in a three-month period.
For each of the 90 days considered, four hourly PM10 measurements are taken at 0:00, 06:00, 12:00 and 18:00 hours (PM00, PM06, PM12 and PM18). The maximum PM10 concentration between the 19:00 hour of the previous day and 18:00 hour of the current day is also calculated (24MAXPM). A number of environmental parameters are taken into account including:

- the minimum 1-h average relative humidity between 19:00 hour of the previous day and the 18:00 hour of the current day (24MINRH),
- the maximum temperature over the same time period (24MAXTM)
- the difference between the maximum and minimum temperature over the same time period (24DIFTM)
- the average wind speed over the same time period (24AVGWS)

Forecasted meteorological variables are also taken into account:

- minimum 1-h average of relative humidity for the next day (NDMINRH),
- maximum temperature for the next day (NDMAXTM),
- difference between the maximum and minimum temperature for the next day (NDDIFTM),
- and the average wind speed for the next day (NDAVGWS).

All these measured parameters and forecasts are used as inputs to the neural network based prediction system presented in this study. The required prediction output from the network is the maximum of the 24-h moving average of PM10 concentration for the next day that will be obtained at 18:00h of the current day. The choice of inputs was motivated by similar studies [1, 2, 6, 8, and 15].

2 Time Lagged Feedforward Neural Networks

Standard feed forward networks based on supervised learning have a long-term memory built during learning and stored in the synaptic weights of the network. These structures however are not suitable for problems with a temporal dimension such as the prediction of PM10 concentrations. Such problems require some form of a short-term memory to make the network dynamic. One simple way of building short-term memory into the structure of a neural network is through the use of *time delays*, which can be implemented at the synaptic level inside the network or at the input layer of the network [4].

Time Lagged Feedforward Networks (TLFN) offer a powerful architecture for the prediction of patterns that evolve over time, with the response at a particular instant of time depending not only on the present value of the input but also on its past values. There are two types of TLFN networks: *focused* and *distributed*. Focused TLFN networks are limited to maps that are shift invariant, i.e. they are only suitable for use in stationary (time-invariant) environments. The problem of PM10 prediction is a time variant one as all input variables change with time. The distributed TLFN architecture overcomes this limitation by distributing the implicit influence of time throughout the network.

2.1 Distributed TLFN Network

Distributed TLFN networks rely on the use of a spatiotemporal model of a neuron, namely, a *multiple-input* or *distributed neuronal filter*. This model uses finite-duration impulse response (FIR) filters as synaptic filters. As such, the multiple inputs neuronal filter provides a powerful functional block for spatiotemporal signal processing built around a single neuron. Each memory neuron in the distributed TLFN is effectively processing information over time by working with the projections of the neuron activations of the previous layer on its local linear memory space [9]. The size of each memory space (i.e., the number of bases) is determined by the number of memory taps.

2.2 Distributed Neuronal Filter

The distributed neuronal filter builds on the processing power of the finite-duration impulse response (FIR) filter of order p, shown in Figure 2. The FIR filter is one the basic building blocks in digital signal processing [3, 7].

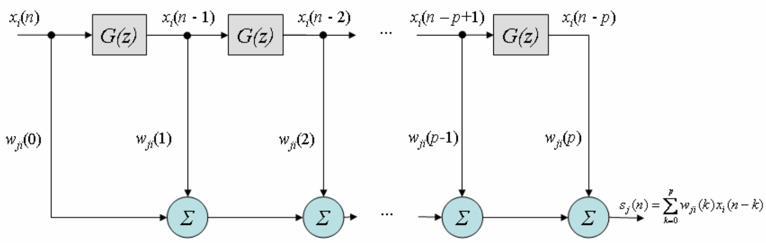

Figure 2: Finite-duration impulse response (FIR) filter.

The processing power of the FIR is extended by the use of multiple inputs, m_0 in number, as depicted in Figure 3. The spatiotemporal model of the distributed neuronal filter is also referred to as a *multiple input neuronal filter* [4].

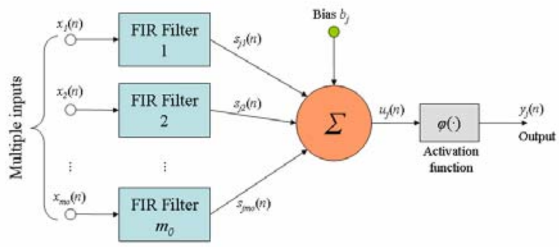

Figure 3: Distributed (multiple-input) neuronal filter.

The neuron has m_0 primary synapses, each of which consists of a linear discrete time filter implemented in the form of an FIR filter of order p; the primary synapses account for the spatial dimension of signal processing. Each primary synapse has $(p + 1)$ secondary synapses that are connected to its respective input and the memory taps of its FIR filter, thereby accounting for the temporal dimension of signal processing (Figure 4).

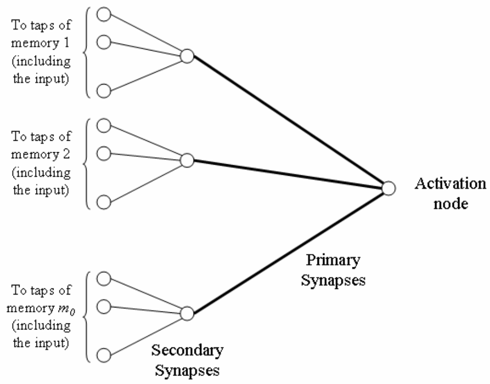

Figure 4: Synaptic structure of a distributed neuronal filter.

The spatiotemporal processing performed by the neuronal filter in Figure 5 can be expressed mathematically in terms of its output, $y_j(n)$, as

$$y_j(n) = \varphi\left(\sum_{i=1}^{m_0}\sum_{l=0}^{p} w_{ji}(l)x_i(n-1) + b_j\right)$$

Where $w_{ji}(l)$ is the weight of the lth secondary synapse belonging to the ith primary synapse, $x_i(n)$ is the input to the ith primary synapse at time n, and b_j is the bias applied to the neuron. The overall structure of the distributed TLFN network is shown in Figure 4.

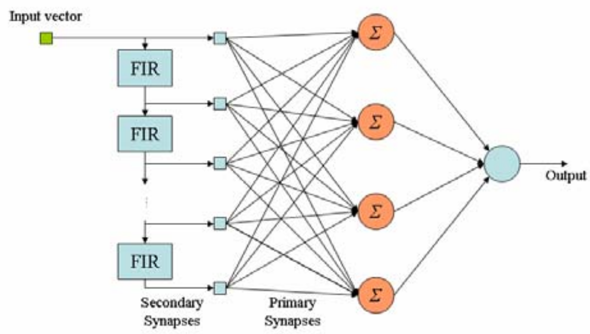

Figure 5: Distributed TLFN network architecture.

A supervised learning algorithm is necessary to train a distributed TLFN network, in which the actual response of each neuron in the output layer is compared with a desired response at each time instant. In this study, the *temporal back propagation* algorithm is used to train the distributed TLFN networks as described by Haykin [4]. The optimum number of hidden nodes as well as the learning parameters are determined using genetic optimisation.

3 Application

The application of the distributed TLFN network described above to the PM10 data is split into three main stages: genetic optimisation, training and validation, and testing. These stages are completely separate from each other. The first two stages correspond to the development of the TLFN network based on the available data while the third stage will give a measure of the validity of the approach using part of the available data that was hidden from the development stages.

3.1 Genetic Optimisation

The first stage of the network development is the choice of the optimum number of hidden nodes and the choice of the most suitable learning parameters for the learning algorithm. In this study the following components of genetic optimisation are used:

- **Selection:** Selection is a genetic operator that chooses a chromosome from the current generation's population for inclusion in the next generation's population. Before making it into the next generation's population, selected chromosomes may undergo crossover and/or mutation (depending upon the probability of crossover and mutation) in which case the offspring chromosome(s) are actually the ones that make it into the next generation's population. In this study selection is based on the ***Roulette*** selection operator – the chance of a chromosome getting selected is proportional to its fitness (or rank).
- **Crossover:** Crossover is a genetic operator that combines (mates) two chromosomes (parents) to produce a new chromosome (offspring). The idea behind crossover is that the new chromosome may be better than both of the parents if it takes the best characteristics from each of the parents. Crossover occurs during evolution according to the Crossover Probability. This probability should usually be set fairly high (0.9 is a good first choice). In this study the ***One Point*** crossover operator is used – a crossover point is randomly selected within a chromosome then the two parent chromosomes interchange at this point to produce two new offspring.
- **Mutation:** Mutation is a genetic operator that alters one ore more gene values in a chromosome from its initial state. This can result in entirely new gene values being added to the gene pool. With these new gene values, the genetic algorithm may be able to arrive at a better solution than was previously possible. Mutation is an important part of the genetic search as it helps to prevent the population from stagnating at any local optima. Mutation occurs during evolution according to the probability defined in this cell. In this study the ***Mutation Probability*** is

set to 0.01. The mutation operator is the ***Uniform*** – it replaces the value of the chosen gene with a uniform random value selected between the user-specified upper and lower bounds for that gene.

The genetic optimisation resulted to 10 hidden units, a learning step of 0.55 and a momentum of 0.7. A total of 100 generations and 50 different chromosomes were developed during the genetic optimisation. The best architecture of the TLFN was passed to the next development stage, training and validation.

3.2 Training and Validation

Due to the relatively small number of available samples, training and validation required a very little time to complete. Of the 90 available samples, 50 were used for training, 27 for validation and 13 for testing. All input variables were normalised. Training and validation are processes that operate concurrently. The mean square error (MSE) produced after each training step based on the entire validation set (batch method) is used for adjusting the TLFN's parameters using the temporal back propagation algorithm. The training procedure is stopped after a maximum of 1000 epochs (complete propagations of the entire training set) or once the MSE on the validation set shows no signs of improvement. The learning curve (Figure 6) shows an ideal progression of the training process. The network's generalisation improved with time and the MSE was decreasing on the training and validation set, meaning that over-fitting the training set was avoided. The MSE of the fully trained TLFN was 0.081 on the training set, 0.09 on the validation set (18% and 24% respectively).

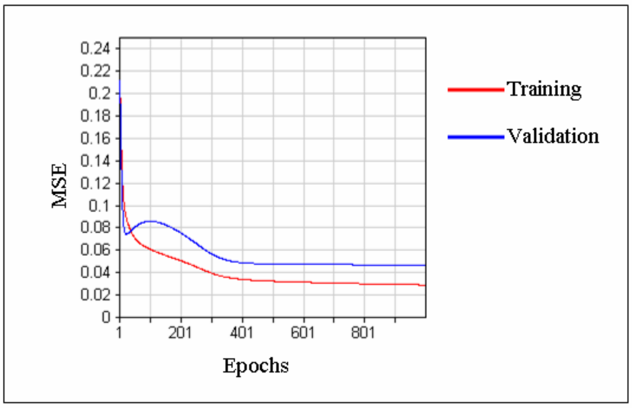

Figure 6: Learning curve of the optimised TLFN trained with PM10 data.

3.3 Testing

Testing the fully developed TLFN network was based on the 13 values hidden from the development process. The network was used to predict the maximum 24h PM10 value for the next day for 13 days that were not part of its development. This way the true predictive capabilities could be determined. The network produced a MSE of 0.082 (24.9%). The network was able to be unbiased in the predictions and was able to maintain the same levels of accuracy experienced on the training and validation set (Figure 7).

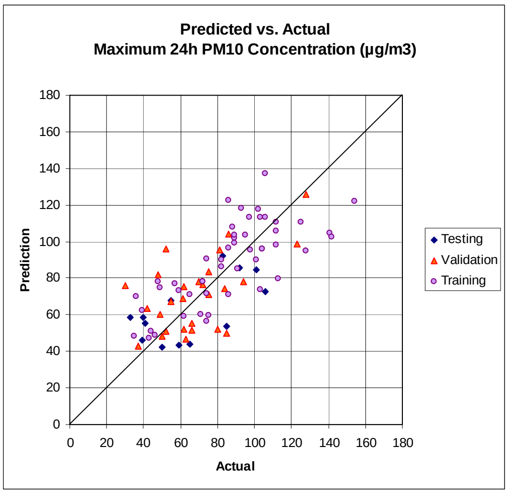

Figure 7: Scatter diagram of TLFN's predictions on the three parts of the dataset.

4 Discussion and Conclusions

In this paper a new approach for predicting the maximum 24h PM10 concentration based on a temporal model of neural network has been presented. The structure and operation of the Time Lagged Feedforward Network was analysed as well as its advantages over other neural network architectures. Genetic optimisation was used to derive the best possible TLFN network for the available data. The predictive power of the optimised TLFN network has been tested on a small dataset consisting of PM10 measurements from an urban area and nearby power plant. Testing was based on part of the data that were hidden from the network's development stages. The results showed the potential of the approach as well as the validity of the choice of the TLFN architecture as the basis for a PM10 prediction system.

Future work will include further analysis of the contribution of each input variable to the system's performance and the application of the system on a much larger dataset spanning over many years of measurements. Certain aspects of the network's architecture also need further investigation such as the size of the network's short-term memory (number of memory taps). This will become more feasible with the use of the larger dataset.

Once this research work is complete, the fully developed TLFN will be deployed at the Laboratory of Atmospheric Pollution and Environmental Physics of the Technological Education Institute of West Macedonia. The system will receive measurements from multiple sources in real-time mode and give a prediction for the maximum daily PM10 concentration of the following day.

References

[1] Boznar, M., Lesjak, M., and Mlakar, P., A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment Vol. 27B, No. 2, pp. 221-230, Pergamon Press Ltd. 1993.

[2] Chaloulakou, A., Grivas, G., and Spyrellis, N., Neural network and multiple regression models for PM10 prediction in Athens: a comparative assessment. Journal of the Air & Waste Management Association 53:1183-1190, 2003.

[3] Haykin, S., and Van Veen, B., Signals and Systems. Wiley, 1998.

[4] Haykin, S., Neural networks – a comprehensive foundation (2nd edition). Prentice Hall, 1999.

[5] Kelsall, J.E., Samet, J.M., Zeger, S.L., and Xu J., Air pollution and mortality in Philadelphia, 1974-1988. Am J Epidemiol, 146(9):750-762, 1998.

[6] Lu, W.Z., Fan, H.Y., Leung, A.Y.T., and Wong, C.K., Analysis of pollutant levels in central Hong Kong applying neural network method with particle swarm optimization. Environmental Monitoring and Assessment 79: 217-230, Kluwer Academic Publishers, 2002.

[7] Oppenheim, A.V., and Schafer, R.W., Discrete-time signal processing. Englewood Cliffs, Prentice Hall, 1989.

[8] Perez, P., and Reyes, J., Prediction of maximum of 24-h average of PM10 concentrations 30h in advance in Santiago, Chile. Atmospheric Environment 36: 4555-4561, Elsevier Science Ltd. 2002.

[9] Principe, J.C., Euliano, N.R., and Lefebvre, W.C., Neural and adaptive systems – fundamentals through simulations. John Wiley & Sons, 2000.

[10] Schwartz J., Dockey, D.W., and Neas, L.M., Is daily mortality associated specifically with fine particles? *J Air Waste Manage Assoc*; 46:927-939, 1996.

[11] Triantafyllou, A.G., and Kassomenos, P.A., Aspects of atmospheric flow and dispersion of air pollutants in a mountainous basin. The Science of the Total Environment 297: 85-103, Elsevier Science B.V. 2002.

[12] Triantafyllou, A.G., Kiros, E.S., and Evagelopoulos, V.G., Respirable particulate matter at an urban and nearby industrial location: concentrations and variability and synoptic weather conditions during high pollution episodes. Journal of the Air & Waste Management Association 52:174-185, 2002.

[13] Triantafyllou, A.G., Levels and trend of suspended particles around large lignite power stations. Environmental Monitoring and Assessment 89: 15-34, Kluwer Academic Publishers, 2003.

[14] Triantafyllou, A.G., PM10 pollution episodes as a function of synoptic climatology in a mountainous industrial area. Environmental Pollution 112: 491-500, Elsevier Science Ltd., 2001.

[15] Wang, W., Lu, W., Wang, X., and Leung, A.Y.T., Prediction of maximum daily ozone level using combined neural network and statistical characteristics. Environmental International 29: 555-562, Elsevier Science Ltd., 2003.