# Air Quality Prediction Using Artificial Neural Network

Suraya Ghazali[1] and Lokman Hakim Ismail[2]
[1]Faculty of Civil and Environmental Engineering,
Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat, Johor
[2]Department of Construction Engineering and Architecture,
Universiti Tun Hussein Onn Malaysia (UTHM), Batu Pahat, Johor
suraya1886@gmail.com[1], lokman@uthm.edu.my[2]

## Abstract

Over the last few years, the use of artificial neural networks (ANNs) has increased in many areas of engineering. Artificial neural network have been applied to many environmental engineering problems and have demonstrated some degree of success. The aim of study is to develop neural network air quality prediction model. In this study, a prediction method is developed using feed-forward neural network. Several parameters such as sulphur dioxide ($SO_2$), carbon monoxide (CO), nitrogen dioxide ($NO_2$), nitric oxide (NO), temperature, relative humidity and air velocity are considered in this study. The performance of the developed model was assessed through a measure of Mean Square Error (MSE) and value of $R^2$. From the constructed networks, the best prediction performance was observed in a model with network structure 7-20-4 with $R^2$ value of 0.57 and MSE 0.062.

**Keywords:** *Air Quality; Artificial Neural Network; Prediction; MATLAB.*

## 1. Introduction

Air pollution is an important issue nowadays, being a factor which influences both human health and activities. There are many different chemical substances that contribute to it. These chemicals come from a variety of sources. On one hand, there are natural sources such as forest fires, volcanic eruptions, wind erosion, pollen dispersal, evaporation of organic compounds and natural radioactivity. And on the other hand, the human industrial activity represents the artificial air pollution sources [1].

Among the many types of air pollutants are nitrogen oxides, sulphur oxides, carbon monoxides, ozone and organic compounds that can evaporate and enter the atmosphere. Large quantities of any air pollutant can affect the population health [1]. Based on the World Health Organization [2] reports that 2.4 million individuals die annually from causes directly attributable to air pollution, 1.5 million of these from indoor air pollution. Worldwide there are more deaths from poor air quality than from automobile accidents [3]. In order to prevent them, there have been developed regional, national and international air pollution monitoring networks, which inform people about major pollutants concentrations in real time [1].

Air pollutants exert a wide range of impacts on biological, physical, and economic systems. Their effects on human health are of particular concern [4]. The decrease in respiratory efficiency and impaired capability to transport oxygen through the blood caused by a high concentration of air pollutants may be hazardous to those having pre-existing respiratory and coronary artery disease [5]. Consequently, it has become a vital task to accurately keep track of the variation of ambient air pollution levels in urban areas [4].

With increasingly severe air pollution, it is important to predict air quality exactly for providing proper actions and controlling strategies so that the adverse effects can be minimized. In response to this concern, several studies on air quality prediction using artificial neural network have been done [4][6][7]. Unlike other modeling techniques, artificial neural networks (ANN) make no prior assumptions concerning the data distribution. ANN is capable of modeling highly non-linear relationships and can be trained to accurately generalize when presented with a new data set [8].

The strong capability of artificial neural networks in predicting fuzzy data and the successful application of this approach in various fields gives the idea of implementing ANN to predict air quality based on previous data. This research will attempt to use feed-forward neural network modeling in the prediction estimation where historical data collected over the years is used to 'train' the model. This research was carried out because of the people's lack of awareness about the real time air quality status. The prediction model will be developed by using MATLAB software.

This paper is organized as follows: Section 2 provides the materials and methods. Section 3 describes the results of the ANN modeling. The last section concludes the presented works.

## 2. Materials and Methods

### 2.1 Data Sets

The data used in this study are daily ambient air temperature, relative humidity, air velocity and daily concentration of CO, NO, $SO_2$ and $NO_2$ in Batu Pahat for 4 years period from 2001 to 2004. All of these data were provided by Malaysian Meteorological Department (MMD) and Department of Environment (DOE). The data was divided into two sets which is learning set for ANN training and testing set to verify the efficiency and correctness of the developed model.

### 2.2 Software

For development of the air quality prediction model, MATLAB Neural Network Toolbox (The MathWorks Inc. USA) [9] was used because it is flexible and easy to apply. The Neural Network Toolbox offers a broad variety of parameters for neural network development which can be chosen flexibly. The toolbox is provided with a practical user guide.

### 2.3 ANN Model for Air Quality Prediction

As network architecture, a 3-layer perceptron model as shown in **Figure 2.1** was used. The first input layer contains the input variables of the network. Here, there were seven neurons in the input layer including four pollutants which is CO, NO, $SO_2$ and $NO_2$; and three comfort variables which is $T_{air}$, RH and $V_{air.}$ The number of hidden layers and values of neurons in each hidden layer are the parameters to be chosen in the model. Therefore, one or two hidden layers and different value of neurons were chosen to optimize the ANN performance. The last layer is the output layer, which consists of the target of the prediction model. Here, $SO_2$, $NO_2$, CO and NO were used as the output variables. Hyperbolic tangent sigmoid function was used as the transfer function. The database was divided into three sections for early stopping. 70% of the data were used in training the networks, 15% were designated as the validation set, and the remaining 15% were employed in testing the networks. The mean square error (MSE) was chosen as the statistical criteria for measuring of the network performance.

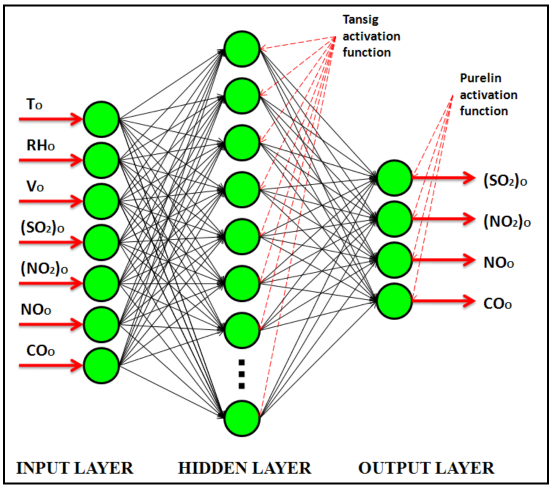



Figure 2.1: ANN architecture for air quality prediction

## 3. Results and Discussion

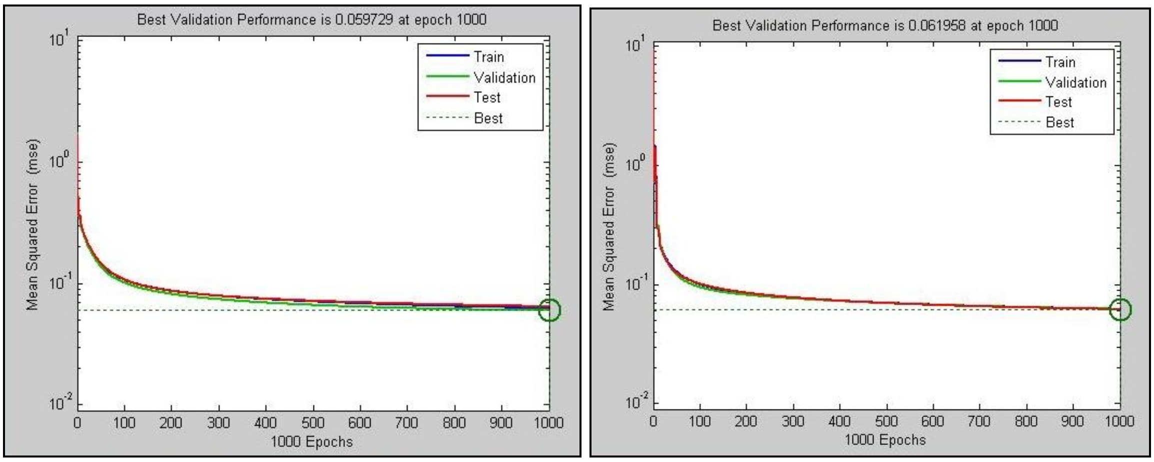
### 3.1 ANN Modeling

Feed-forward neural network have been applied in this study. The tansig and purelin functions were used for the neurons in the hidden layer and output layer respectively. The input and target values were normalized into the range of [-1,1] in the pre-processing phase. The weights and biases were adjusted based on gradient-descent back-propagation in the training phase. The mean square error was chosen as the statistical criteria for measuring of the network performance. The overview of the parameters and their values was shown in **Table 3.1**

Table 3.1: Structure and training results for the neural network models

| Net. No | Net. structure | Training function | learning | Learning rate | Momentum constant | MSE | R | $R^2$ |
|---|---|---|---|---|---|---|---|---|
| 1 | 7-10-4 | Tansig-purelin | traingdm | 0.1 | 0.6 | 0.0626 | 0.7496 | 0.562 |
| 2 | 7-20-4 | Tansig-purelin | traingdm | 0.1 | 0.6 | 0.0620 | 0.7547 | 0.570 |

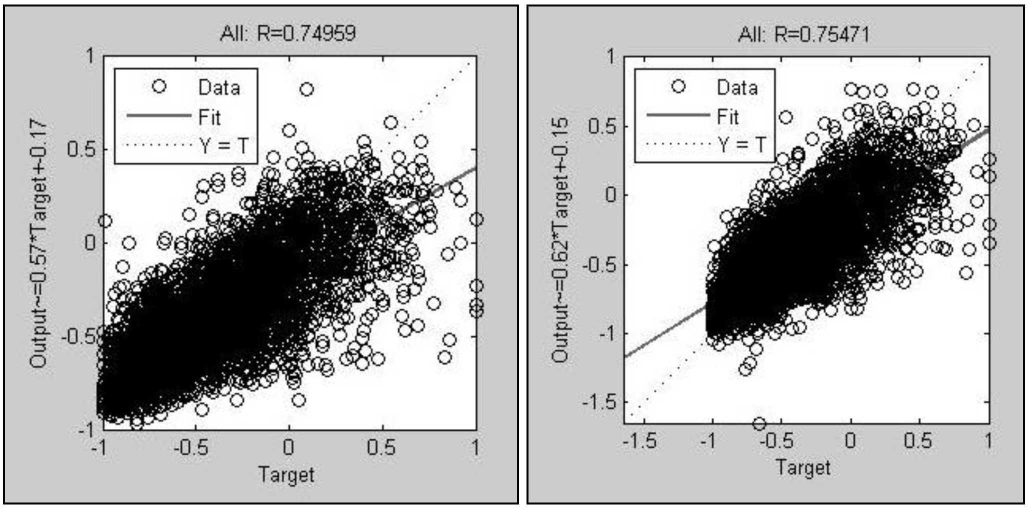
During the training, the following figure appears. It represents the network performance versus the number of epochs. The network performance starts by a large value at the first epochs and due to training, the weights are adjusted to minimize this function which makes it decreasing. Moreover, a black dashed line is plotted representing the best validation performance of the network. The training stops when the green line which represents the validation training set (network performance) intersects with the black line. The performance function of the network is shown in **Fig. 3.1**.




**(a)** **(b)**

Figure 3.1: Performance function of the network during training.
(a) Network structure, 7-10-4 (b) Network structure, 7-20-4

Regression analysis was performed to investigate the correlation between the actual and predicted results based on the value of correlation coefficient, R. The perfect fit between the training data and the produced results was indicated by the value of R which is equal to 1. **Figure 3.2** shows the regression analysis plots of the network structure. In a regression plot, the perfect fit which shows the perfect correlation between the predicted and targets is indicated by the solid line. The dashed line indicates the best fit produced by the algorithm.




**(a)** **(b)**

Figure 3.2: Regression plot analysis
(a) Network structure, 7-10-4 (b) Network structure, 7-20-4

From **Table 3.1**, it can be seen that model with network structure 7-20-4 is the best models for the air quality prediction as it yields the lowest values of MSE and a 57% coefficient of determination, $R^2$. As shown in **Figure 3.2**, the value of correlation coefficient, R for both models did not have much difference between each other. The best model chosen show a good agreement between predicted and measured values based on the value of correlation coefficient, R.

## 4. Conclusion

Based on the analysis conducted, model with neural network structure 7-20-4 produces the best performance in the prediction of air quality compared to the first model based on the values of R and the prediction accuracy. This model produces R of 0.7547 which indicate a good correlation between the targets and predicted outputs. However, this model still need an improvement to give a better result for air quality forecasting.

## 5. Acknowledgements

The author would like acknowledge the source of funding provided by the Ministry Of Higher Education (MOHE) for this research work. The provision of original data from the Malaysian Meteorological Department (MMD) and Department Of Environment (DOE), are also appreciated.

## 6. References

[1] Dragomir, E.G. (2010), “Air Quality Index Prediction using K-Nearest Neighbor Technique”, Bulletin of PG University of Ploiesti, Series Mathematics, Informatics, Physics, vol.LXII, no.1/2010, pp.103-108.
[2] WHO Member State. “Estimated deaths and DALYs attributable to selected environmental risk factors”. [Accessed 27 Aug 2009].
[3] Deleawe, S., Kusznir, J., Lamb, B. and Cook, D. (2010), “Predicting air quality in smart environments”, Journal of Ambient Intelligence and Smart Environments, Vol.2, No.2: 145-154.
[4] Barai, S.V., Dikshit, A.K. and Sharma, S. (2003), “Neural Network Models for Air Quality Prediction: A Comparative Study”.
[5] Rao, M.N., and Rao, H.V. (2000), “ Air pollution”, Tata McGraw-Hill, New Delhi.
[6] Wang, W., Xu, Z. And Lu, J.W. (2003), “Three improved neural network models for air quality forecasting”, Engineering Computations, Vol.20, No.2, pp. 192-210.
[7] Li, M. and Hassan, M.R. “Urban air pollution forecasting using artificial intelligence-based tools”.
[8] Kurt, A., Gulbagci, B., Karaca, F., Alagha, O. (2008), “An Online Air Pollution Forecasting system using neural networks”, Environment International, vol 34, Issue 5, July 2008, pp.592-598
[9] Demuth, H., Beale, M. and Hagan, M. (2009), “ Neural Networks Toolbox Manual”, Math Works Inc.