Chapter 2
An Application of Artificial Neural Networks for the Prediction of Surface Ozone Concentrations in Malaysia

Negar Banan, Mohd Talib Latif, Liew Juneng and Md. Firoz Khan

Abstract In this study, an artificial neural network (ANN) was used to extract the complex relationships among divergent parameters that have the capabilities to predict O_3 concentrations which serve as an input to meteorological conditions and precursor concentrations. The ANN was trained using samples of daily maximum data provided by the Malaysian Department of the Environment (DOE) over a period of 9 year (2003–2011) in the towns of Gombak and Shah Alam in Malaysia. Furthermore, surface O_3 concentrations from the two locations (Gombak and Shah Alam) were estimated using surface meteorological variable as predictors for the ANN. Based on the results, it can be deduced that the relationship between the parameters and the O_3 concentrations are highly complex and non-linear. Analysis of the regression based model results between Gombak and Shah Alam were evaluated using the ANN. Based on the sample results it was confirmed that Shah Alam has the highest regression result of $R = 0.64$ in comparison with Gombak station. The inference drawn from this study shows that neural network model consistently gives superior predictions.

Keywords Artificial neural network · Surface ozone · Meteorological factors · Regression

N. Banan (✉) · M. T. Latif · M. F. Khan
School of Environmental and Natural Resource Sciences, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
e-mail: artibanan@yahoo.com

L. Juneng
Research Center for Tropical Climate Change System (IKLIM), Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia

A. Z. Aris et al. (eds.), *From Sources to Solution*,
DOI: 10.1007/978-981-4560-70-2_2,

Highlights

- Artificial neural network model successfully predicts O_3 concentration.
- O_3 precursors and meteorological factors influenced predicted O_3 concentration.
- The proposed algorithm exhibits a high accuracy model.

Introduction

Ozone (O_3) is a major constituent of smog because it is readily released into the atmosphere as a result of photochemical oxidation and hence, a heavy air pollutant. Many studies have shown that the high presence of O_3 is capable of causing severe health respiratory disease (Huang et al. 2012; Neidell and Kinney 2010). Furthermore, it has been proved that there is a positive correlation between exposure to O_3 and decrease in lung function (Highfill and Coasta 1995). At high O_3 concentration, vegetation and forests are affected because of the phytotoxic nature of O_3. Higher concentration above 40 ppbv can be suggested to be harmful to the plants yield, biomass production, vitality and stress tolerance of forest trees (Fuhrer et al. 1997). It has been suggested that excessive level of O_3 may affect the ability of the forest to seize carbon in an event of excess of CO_2 in the future (Banan et al. 2013).

Modeling of O_3 concentrations is not a trivial procedure due to the intrinsic relationship between pollutants and meteorological variables (Sousa et al. 2007). However, different regression approaches have been adopted for forecasting surface O_3 with the predictor parameters varying from a few to a large sample inputs. Studies conducted by Wang et al. (2003) using statistical characteristics of O_3 to enable selection of appropriate predictors of daily maximum O_3 levels and highlighted the inclusion of parameters that affect both photochemical production and atmospheric accumulation of O_3 when forecasting O_3 levels. In this paper, ANN is used for the prediction of the daily maximum O_3 concentration in urban areas using the precursors and meteorological variables.

Materials and Methods

This experiment is carried out with two datasets collected from the air quality monitoring sites at the Department of the Environment (DOE) in Malaysia, which is managed by a private company, Alam Sekitar Sdn Bhd (ASMA). The daily maximum concentration of air pollutants such as; O_3, NO, NO_2, NO_x, CO, PM_{10} and NMHC and daily maximum meteorological variables (Ambient temperature (AT), Humidity (H), Wind speed (WS) and wind direction (WD)) were used. The logarithm of daily maximum O_3 concentrations of two stations, namely Gombak

Table 2.1 Characteristics of the datasets

Datasets	No. of attributes	Training set 70 %	Testing set 20 %	Validation set 10 %
Gombak	10	1019	291	146
Shah Alam	10	1732	495	247

and Shah Alam stations, that are located in Klang Valley in the middle of the Malaysian Peninsula, were derived and used for modeling. The concentration data were sourced from a 9-year period data samples (2003–2011). All the data were used to evaluate the predicting performance of the modeling data. The characteristics of the datasets used are summarized in Table 2.1.

ANN Model for Air Quality Prediction

The model of employed ANN is shown in Fig. 2.1. It comprises of three layers of neurons namely; input, hidden, output layers. The input layer is the first layer of neurons. Each input neural represents a separate attribute in the train/test datasets station (for example from NO to WD). The number of the inputs is equal to the number of attributes in the dataset. The number of nodes for other hidden layers is equal to the half of the number of nodes in the previous layer.

There has been recommendation that the input data be normalized between slightly offset values such as 0.1 and 0.9. One way to scale input and output variables in the interval [0.1, 0.9] as $P_n = 0.1 + (0.9–0.8)\ (P - P_{min})/(P_{max} - P_{min})$ (Xu et al. 2007).

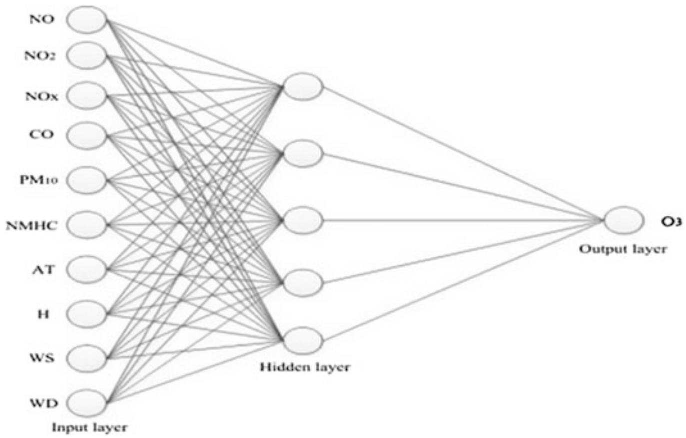

Fig. 2.1 Schematic of a neural network

Results and Discussion

In this study, artificial neural network was applied with input and targeted values normalized in the range of [0.1 and 0.9] before processing the data. The weights and biases were adjusted based on gradient-descent method in the training phase. The correlation coefficient (R) was chosen as the statistical criteria for measuring of the network performance. The result values were shown in Table 2.2.

Figure 2.2 depicts the network performance versus the number of epochs. At initialization of the network, the first values of the epochs were large due to training. Subsequently, the weights are adjusted to minimize this function which resulted in values decreasing. Meanwhile, a black dashed line is plotted which represents the best validation performance of the network. The training is stopped when the green line which represents the validation training set (network performance) cut-across with the black line.

In order to validate our hypothesis, regression analysis was performed to investigate the correlation between the actual and predicted results based on the value of correlation coefficient (R). There was a perfect fit value of R equal to 1 between the training data and the produced results. Figure 2.3 depicts the regression analysis plots of the network structure. From the regression graph, the perfect fit depicts that there is positive correlation between the predicted and targets as indicated by the solid line. The dashed line indicates the best fit produced by the algorithm.

Table 2.2 Correlation coefficient

Datasets	Correlation coefficient (R)
Gombak	0.62945
Shah Alam	0.63819

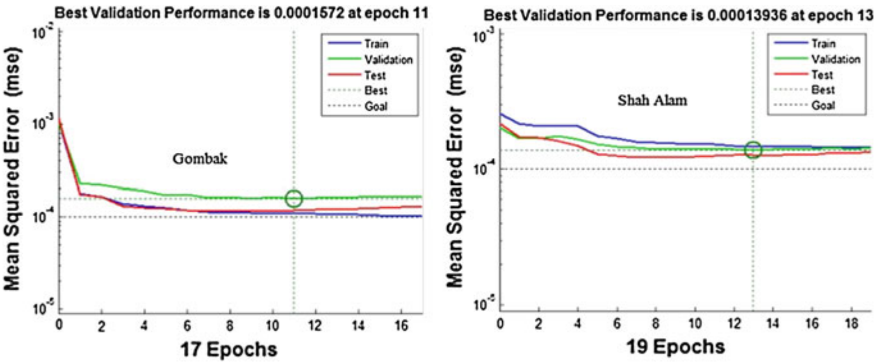

Fig. 2.2 Performance function of the network during training network structure, 10-5-1

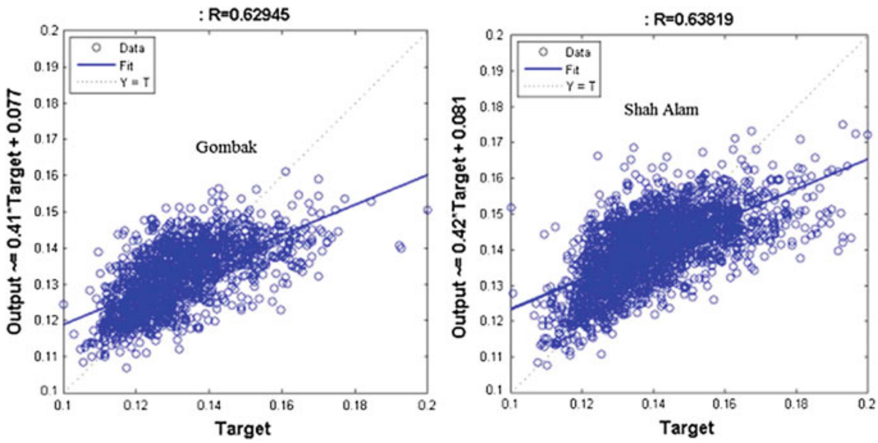

Fig. 2.3 Regression plot analysis network structure, 10-5-1

From Table 2.2, it can be seen that the model with network structure 10-5-1 is the optimal model for the air quality prediction as it yields the highest values of correlation coefficient (R). As depicted in Fig. 2.3, there is a match between predicted and measured values based on the value of correlation coefficient (R).

Conclusion

The input data to ANNs as used to forecast the O_3 concentrations is dependent on meteorological conditions and precursor concentrations. The network was trained using daily maximum data that were extracted from the Malaysian department of the Environment (DOE) during a 9-year period (2003–2011) in Malaysian towns of Gombak and Shah Alam. During the project under review, the surface O_3 concentrations were forecasted using surface meteorological variable as predictors for the artificial neural network on four experimental sites in Malaysia. The experimental results revealed that the proposed algorithm exhibits a high accuracy in predicting the O_3 concentrations. Additional study is hence needed to validate that the ANN algorithm can be used for the purpose of predicting O_3 concentrations with inputs from meteorological conditions and precursor concentrations during monsoon season.

References

Banan N, Latif MT, Juneng L, Ahmad F (2013) Characteristics of surface ozone concentrations at stations with different backgrounds in the Malaysian Peninsula. Aerosol Air Qual Res 13:1090–1106

Fuhrer J, Skarby L, Ashmore MR (1997) Critical levels for ozone effects on vegetation in Europe. Environ Pollut 97:91–106

Highfill JW, Costa DL (1995) Statistical response models for ozone exposure: their generality when applied to human spirometric and animal permeability functions of the lung. J Air Waste Manag Assoc 45:95–102

Huang HL, Lee MG, Tai JH (2012) Controlling indoor bioaerosols using a hybrid system of ozone and catalysts. Aerosol Air Qual Res 12:73–82

Neidell M, Kinney PL (2010) Estimates of the association between ozone and asthma hospitalizations that account for behavioral responses to air quality information. Environ Sci Policy 13:97–103

Sousa SIV, Mrtins FG, Alvim-Ferraz MCM, Pereira MC (2007) Multiple linear regression and artificial neural networks based on principle components to predict ozone concentrations. Environ Model Softw 22:97–103

Wang W, Lu W, Wang X, Leung AYT (2003) Prediction of maximum daily ozone level using combined neural network and statistical characteristics. Environ Int 29:555–562

Xu L, Jiandong X, Shizhong W, Yongzhen Z, Rui L (2007) Optimization of heat treatment technique of high-vanadium high-speed steel based on back-propagation neural networks. Mater Des 28:1425–1432

http://www.springer.com/978-981-4560-69-6

From Sources to Solution
Proceedings of the International Conference on Environmental Forensics 2013
Aris, A.Z.; Tengku Ismail, T.H.; Harun, R.; Abdullah, A.M.; Ishak, M.Y. (Eds.)
2014, XVII, 613 p. 123 illus., 80 illus. in color., Hardcover
ISBN: 978-981-4560-69-6