

## Forecasting the Tropospheric Ozone using Artificial Neural Network Modelling Approach: A Case Study of Megacity Madras, India

Anurag Kandya<sup>1\*</sup>, Shiva Nagendra SM<sup>2</sup> and Vivek Kumar Tiwari<sup>1</sup>

<sup>1</sup>Department of Civil Engineering, Institute of Technology, Nirma University, Ahmedabad, India

<sup>2</sup>Department of Civil Engineering, Indian Institute of Technology Madras, Chennai, India

### Abstract

Ozone, which is a secondary pollutant in the troposphere, is very injurious to human health causing irritation of respiratory system, reducing lung capacity, etc. Adversely affecting the plant growth and deteriorating the materials. Thus it is of prime importance to predict the ozone concentration so that effective mitigation strategies can be adopted. As the formation of the tropospheric ozone is dependent on various meteorological parameters and the concentration of various other air pollutants, it is necessary to consider this dependence aspect in the modelling approach. In this background, the present paper puts forward the study of short-term prediction of tropospheric ozone concentration using Artificial Neural Network (ANN) modelling approach for a busy traffic junction of Madras city, one of the four megacities of India. 8-hourly averaged values of 11 air pollutants concentrations and 6 meteorological parameters were used for the study. The respective data was collected at a busy traffic junction of the city for a period of 19 months i.e. during September 2008–March 2010. 70% of the data was used for training the ANN models while the remaining of 30% data was used for validating them. By changing the neural architecture, 34 ANN models were formulated which were statistically analyzed. Based on the encouraging results ( $d=0.80$ ,  $r=0.69$ , etc.), the paper puts forward the suitability of ANN modelling approach for the short-term prediction of tropospheric ozone concentration.

**Keywords:** Ozone concentration, Artificial neural network technique, Air quality modelling

### Introduction

With the onset of development, there is a marked deterioration in the air quality witnessed across the world. A link between poor air quality and mortality has been recognized for centuries, becoming particularly manifest as urbanization and industrialization intensified [1-3]. Due to various legislative reforms, the situation improved in the developed countries, however it is still grim in the other parts of the world. The source of air pollution was formerly dominated by widespread coal and other solid fuel burning plus other toxic emission from largely unregulated industrial processes [3], and even today 70% of the global emissions come from the burning of the fossil fuels [4]. There is recognition of also the multitude of sources and the complex atmospheric chemistry now contributing to poor air quality and of wider scale of influence [5].

In the contemporary context, the adverse impact of the tropospheric ozone on human health and environment has been brought out by World Health Organization [6] and Environmental Protection Agency [7]. Looking to the health implications of higher ozone concentration, recently the European Commission has introduced new legislative measures establishing ozone thresholds. Accordingly, the people are informed of ozone concentration above  $180 \mu\text{g}/\text{m}^3$  and they are warned if the concentration exceeds  $240 \mu\text{g}/\text{m}^3$  [8]. In India, the maximum permissible limit for ozone concentration is  $100 \mu\text{g}/\text{m}^3$  and  $180 \mu\text{g}/\text{m}^3$  for 8 hour and 1 hour average respectively [9].

Tropospheric ozone being a secondary air pollutant is not directly emitted in the atmosphere. It is formed due to the reactions between the oxides of nitrogen ( $\text{NO}_x$ , which is the sum of NO i.e. nitric oxide, and  $\text{NO}_2$  i.e. nitrogen dioxide) and volatile organic compounds (VOC) [10,11]. These reactions are driven by the various meteorological parameters like solar radiation, relative humidity, etc. In order to characterize the ozone formation, much effort has been devoted in experimentally investigating and simulating air pollution chemistry [12-14]. Artificial Neural Network (ANN) models have been successfully used for the forecasting of a wide range of pollutants and their concentrations at various time scales [15-19]. The findings of numerous research studies also exhibit that the performance of ANNs

is generally superior in comparison to traditional statistical methods, such as multiple regression, classification and regression trees and autoregressive models [20-22]. Neural network models have the potential to describe the highly non-linear relationships such as those controlling ozone productions [23]. Crowe and Defries undertook the task of simulating the ozone concentration using ANN technique and evolved a model having high predictive capability with the value of  $R^2$  as 0.91 [24] while Airekal, a research team formed by Agirre, Anta and Barron has elaborated on and evaluated prognostic models based on mathematical computations method such as ANN for the real time forecast of hourly ozone levels up to eight hours in advance and daily maximum ozone levels at several stations in the Air Quality Monitoring Network of the Basque Country [25-26].

The present study is undertaken to fill the absence of the state-of-the-art forecast models for the short-term prediction of ozone concentration for Chennai city which is one of the four megacities of India. Having a population of 4.6 million, Chennai has witnessed a marked deterioration in the air quality in the past decade. In the present study, multiplayer ANN technique is adopted for developing an air quality forecast model for the short-term prediction of ozone concentration which could be used for issuing air quality alerts to the people thereby helping them to take proper measures at their end.

### Study Area and Data Used

Chennai is the fourth most populous metropolitan area and

**\*Corresponding author:** Anurag Kandya, Department of Civil Engineering, Institute of Technology, Nirma University, Ahmedabad, India, E-mail: [akandya@yahoo.com](mailto:akandya@yahoo.com)

**Received** September 12, 2012; **Accepted** October 26, 2012; **Published** November 06, 2012

**Citation:** Kandya A, Shiva Nagendra SM, Tiwari VK (2012) Forecasting the Tropospheric Ozone using Artificial Neural Network Modelling Approach: A Case Study of Megacity Madras, India. J Civil Environ Eng S1:006. doi:10.4172/2165-784X.S1-006

**Copyright:** © 2012 Kandya A, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

the fifth most populous city of India. It is situated on the southeast coast of India in the northeast of Tamil Nadu on a flat coastal plain known as the Eastern Coastal Plains. It has a tropical climate, specifically a tropical wet and dry climate and the weather is hot and humid for most of the year. Though the population of Chennai is around 4.6 Million, the urban agglomeration of metropolitan Chennai has an estimated population over 8.2 million people.

Air quality in Chennai city of India is mainly monitored by Central Pollution Control Board (CPCB) under the Ministry of Environment and Forest of Government of India. Criteria pollutants like  $\text{SO}_x$ ,  $\text{NO}_x$ , SPM and RSPM are monitored at seven sites across the city viz. IITM, Mylapore, Triplicane, Saidapet, Adyar, R K Nagar and Ambattur. Looking to the serious health implications of tropospheric Ozone, from 2008 onwards, its monitoring was initiated at IITM site. For the present study, 19 months data (starting from September 2008 to March 2010) comprising the various meteorological parameters and concentration of various air pollutants were obtained for the IITM site from CPCB. The IITM site is located at the Velachari gate of the Indian Institute of Technology Madras Campus on the roadside of a major arterial route (Sardar Patel Road). Figure 1 shows the geographical details of the monitoring site. The traffic flow on the Sardar Patel road is around 170,000 vehicles per day. Braking is very frequent near the measurement site due to the presence of traffic lights coupled with the intersections, located one at 50 meters to the west and another at 500 meters to the east. Many premier institutes such as Cancer Institute, hospitals, Central Leather Research Institute, Anna University, Central polytechnic, software companies and tourist spots (Children's park, Kamarajar Mandapam and Gandhi Mandapam) along with dense commercial and residential areas are also located in the adjoining areas. In addition, many heavy industries are also located about 30 km away in the NNE direction from the monitoring site. The 8-hourly  $\text{NO}_x$  concentration was generally in the range of 20-60  $\mu\text{g}/\text{m}^3$  however many times the concentration crossed the levels of 80  $\mu\text{g}/\text{m}^3$ . Thus the monitoring site duly represents the city region having intense human activity and having heavy traffic flow.

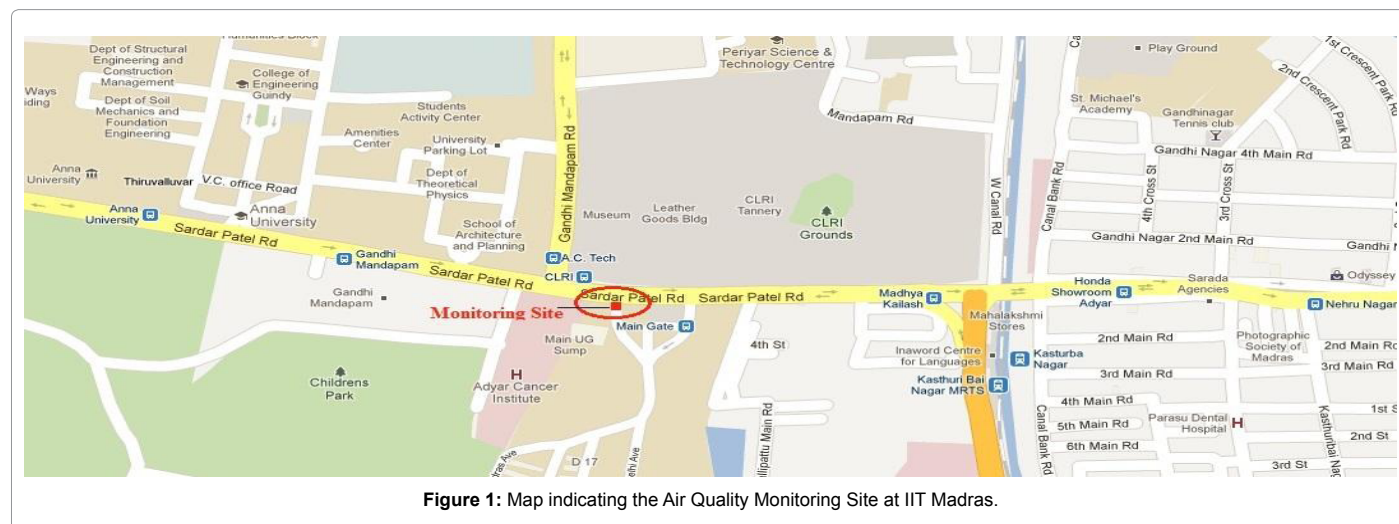
The data used in the present study consist of 8-hourly averaged concentration of air pollutants like Ozone, oxides of Nitrogen (NO and  $\text{NO}_2$ ), Sulphur-di-Oxide, Carbon Monoxide and Respirable Suspended Particulate Matter along with the concentration of hydrocarbons like Benzene, Toluene, Ethyl-benzene, m+p-Xylene and O-Xylene and 8-hourly averaged meteorological parameters like Wind speed,

Wind direction, Solar intensity, Temperature, Pressure and Relative Humidity. The extent of the availability of the usable data for the present study was around 95%.

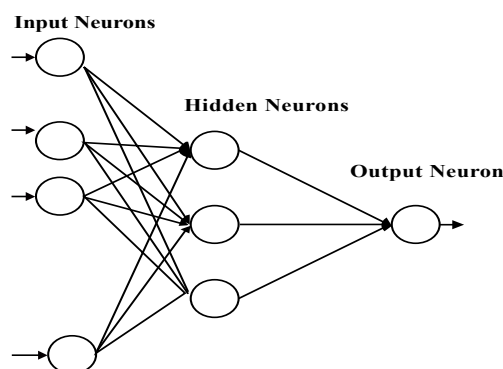
## Methodology

In the present study, multilayer ANN modelling has been adopted for predicting the tropospheric ozone concentration. Artificial Neural Networks are loosely based on biological neural systems, in that they are made up of an interconnected system of nodes (neurons) as shown in figure 2. An ANN can identify patterns in numeric data in a somewhat analogous fashion to the learning process in its biological counterpart (i.e. by accumulated experience or training) [15]. Neurons usually operate in parallel and are configured in regular architectures. They are often organized in layers, and feedback connections both within the layer and toward adjacent layers are allowed. Each feedback connection is expressed by a numerical value called as weight, which can be modified. They are basically stored in three types of layers i.e. input layer, hidden layer and output layer. The output of a node is scaled by connecting weights and fed forward to as input to the nodes in the next layer of the network implying a direction of information processing. The multilayer ANN has the ability to learn through the training. Training requires a set of training data which consists of a series of input and associated output vector. During the training, multilayer ANN is repeatedly presented with the training data and the weights are adjusted in the network from time to time till the desired error goal is reached in the output.

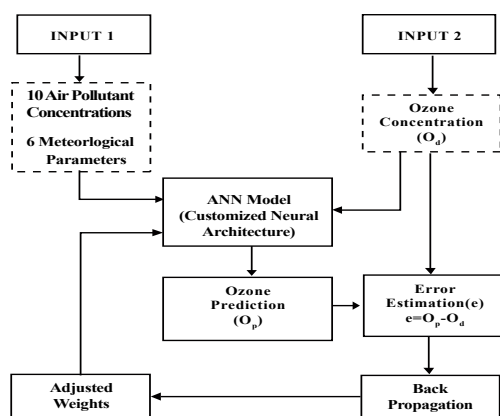
In the present study, the respective data of the 17 parameters for the duration of 19 months was split into two parts viz. 70% of the data was used in training the ANN model while the remaining 30% was used in validating the model. Contrary to the 'rule of thumb' to find the optimal number of neurons in the hidden layers used by the other researchers, the present study used the iterative approach for higher accuracy in prediction [27,19]. The architecture of the ANN model was varied by changing both the number of neurons in the hidden layer and the number of epochs which are the steps in training the model. Starting from 5 neurons, the model formulation went up to 125 neurons with an incremental factor of 10 neurons in each respective model. Based on the learning rate, each set of neurons was trained at 1000 and 2000 epochs and in few cases at 5000 epochs also. Subsequently, 34 ANN models named M-1, M-2, M-3, till M-34 respectively were formulated and tested in the present study. Figure 3 shows the sequential steps adopted for forecasting the tropospheric ozone concentration in the



**Figure 1:** Map indicating the Air Quality Monitoring Site at IIT Madras.



**Figure 2:** General Neural Network Architecture [19].



**Figure 3:** Forecasting the Ozone Concentration using ANN Modelling Approach.

present study. Neural Network Toolbox in Matlab 8.0 was used for simulating the ozone concentration. Statistical parameters like Index of Agreement (d), Coefficient of Regression (r), Fractional Bias (FB), Root Mean Square Error (RMSE), Factor of 2 (FAC 2) and Geometric Variance (VG) were used for deciding the best architecture of the ANN model [28]. For having a better comparative analysis, Air Quality Index was computed for the observed and predicted values of the ozone concentrations [29].

## Results and Discussions

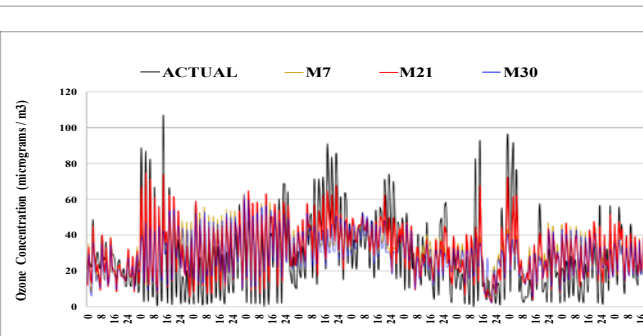
The iterative approach for formulating the architecture of the neural network model for the short-term prediction of ozone concentration resulted in encouraging results. Based on the various statistical parameters, it can be said that of the 34 models so formulated, model M-21 having the architecture of 17 input parameters, 105 neurons and 2000 epochs, has the relatively best forecasting capability. The paired analysis between the between the observed and the forecasted concentration reveal that the respective index of agreement (d) was 0.80 and coefficient of regression (r) was 0.68 which are satisfactory. Additionally, of the 34 ANN models so formulated, 10 models had the value of 'd' more than 0.66. The in-depth analysis of the modelling results revealed that with the increase in the number of neurons in the hidden layers, there was an increase in the prediction accuracy, however beyond 105 neurons the prediction accuracy decreased. Moreover, it was also observed that higher number of epochs does not necessarily lead to better predictions as evident from the present study. Comparing

model M-21 and M-31 which had the same neural architecture except the number of epochs, it was observed that model M-21 having 2000 epochs performed better than model M-31 having 5000 epochs. Table 1 shows the various statistical performance measures of model M-21 along with eight other models viz. M-3, M-7, M-8, M-19, M-20, M-29, M-30 and M-31 which are relatively close to M-21 in the forecasting capability. Figure 4 shows the time-series variation of the observed ozone concentration for the different hours of the day i.e. at 0, 8 and 16 during the evaluation period of the model along with the forecasted values by model M-21 along with model M-7 and M-30 which were relatively close to the best model. In general the forecasted ozone concentration followed the trend of the actual ozone concentration however it appears that the model M-21 slightly under-predicts the maximum values and over predicts the minimum values. Similar inferences were drawn by Comrie in his study of forecasting ozone concentration over 5 cities of United States of America viz. Phoenix, Tucson, Boston, Atlanta and Charlotte [15]. The inter-comparison of the present study with the study undertaken by Comrie is shown in table 2 and table 3. The comparative analysis reveals that the present study performs reasonably well.

For quantifying the temporal variation of ozone concentration along with its health implication, Air Quality Index (AQI) analysis of the observed and predicted ozone concentration was undertaken. AQI is an index of reporting the daily air quality and tells us how clean or polluted the air is, along with its associated health impacts. Mathematically,  $AQI = \text{Pollutant Concentration} \times 100 / \text{Pollutant Standard Concentration}$ . The higher the AQI value, the greater the level of air pollution and the greater are the health concerns. The AQI analysis of the observed ozone concentration during the evaluation period of the model reveal that 85% of the time the air quality was 'good' which means that during this time period, 8-hourly ozone concentration was below 50  $\mu\text{g}/\text{m}^3$  while the remaining 15% of the time it was 'moderate' which means that the respective ozone concentration was between 51-100  $\mu\text{g}/\text{m}^3$ . In comparison to this, according to the model M-21, 88% of the time the air quality was 'good' while the remaining 12% of the time it was 'moderate'. The deviation of 3% observed in the 'moderate' category is because of the under prediction tendency of the model. Figure 5a and figure 5b show the AQI analysis of the observed and forecasted values of ozone concentration.

## Conclusions

The encouraging results of the present study strengthen the inferences and conclusions of the global studies regarding the potential of ANN modelling approach for forecasting ozone concentration and their superiority over conventional modelling approach. Many metropolitan air quality agencies which are making daily ozone



**Figure 4:** Time-series variation of 8-hourly tropospheric Ozone Concentration.



Statistical Performance Measures	Architecture of the various Air Quality (Ozone) Forecasting ANN Models (Number of parameters-Number of Neurons-Number of Epochs)								
	M-3 (17-25-1000)	M-7 (17-65-1000)	M-8 (17-75-1000)	M-19 (17-85-2000)	M-20 (17-95-2000)	M-21 (17-105-2000)	M-29 (17-85-5000)	M-30 (17-95-5000)	M-31 (17-105-5000)
R	0.66	0.66	0.66	0.66	0.66	0.69	0.68	0.66	0.68
FB	-0.15	-0.07	-0.22	0.08	-0.13	-0.13	-0.08	-0.01	-0.12
RMSE <sub>u</sub>	6.88	10.42	10.2	8.37	10.33	8.92	8.89	8.67	9.83
RMSE <sub>s</sub>	15.45	12.38	13.77	13.63	12.46	13.08	13.28	13.26	12.76
RMSE	16.91	16.18	17.14	16.00	16.18	15.83	15.98	15.84	16.11
D	0.66	0.74	0.74	0.72	0.76	0.80	0.73	0.73	0.75
FAC 2	261	269	256	273	267	269	267	273	271
VG	2.45	1.96	2.24	1.75	1.97	1.95	2.04	1.9	2.02

**Table 1:** Statistical Performance Measures of the respective ANN Models trained for forecasting the Tropospheric Ozone Concentration.

Statistical Performance Measures	Present Study (M-21)	Ozone Forecasting using Neural Network Models									
	Madras	Phoenix		Tucson		Boston		Atlanta		Charlotte	
		N1	N2	N1	N2	N1	N2	N1	N2	N1	N2
d	0.80	0.62	0.73	0.66	0.77	0.78	0.81	0.85	0.87	0.84	0.86
RMSE	15.83	14.85	13.55	9.84	8.64	15.37	14.67	16.41	15.57	13.95	13.12
RMSE <sub>u</sub>	13.08	12.71	10.59	8.24	6.43	11.76	10.67	11.01	9.99	9.53	8.34
RMSE <sub>s</sub>	8.92	7.66	8.43	5.37	5.76	9.87	10.04	12.13	11.9	10.17	10.11

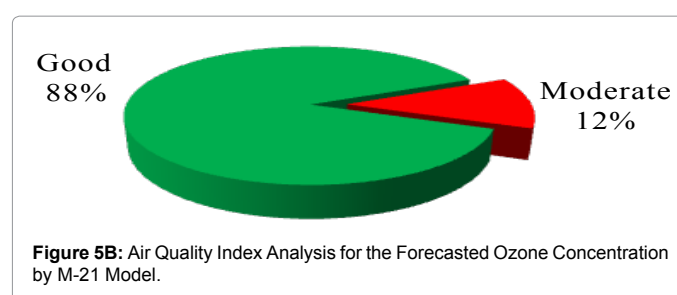
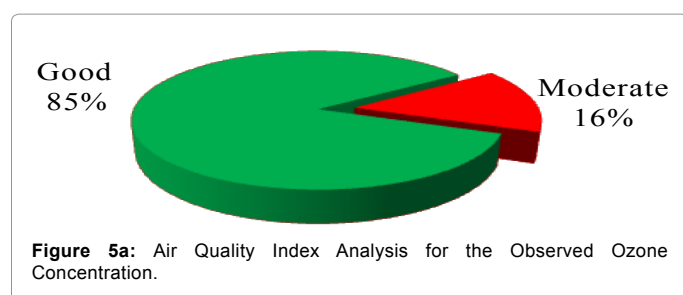
N1: Neural Network Unlagged, N2: Neural Network lagged

**Table 2:** Comparison of the present study with other Neural Network Models for Ozone Forecasting [15].

Statistical Performance Measures	Present Study (M-21)	Ozone Forecasting using Regression Models									
	Madras	Phoenix		Tucson		Boston		Atlanta		Charlotte	
		R1	R2	R1	R2	R1	R2	R1	R2	R1	R2
D	0.80	0.63	0.73	0.54	0.76	0.73	0.79	0.81	0.84	0.81	0.85
RMSE	15.83	15.00	13.55	10.65	8.82	16.6	15.55	17.53	16.48	14.46	13.44
RMSE <sub>u</sub>	13.08	12.44	10.29	9.47	6.48	12.87	11.35	13.24	11.58	10.69	9.04
RMSE <sub>s</sub>	8.92	8.36	8.80	4.86	5.97	10.45	10.61	11.44	11.67	9.71	9.92

R1: Multiple Regression Unlagged, R2: Multiple Regression lagged

**Table 3:** Comparison of the present study with Regression Models for Ozone Forecasting [15].



forecast commonly adopt multiple regression models which are based on linear relationships between weather data and other air pollution data including ozone. However, looking to the potential of the ANN models to incorporate complex and nonlinear relationships such as those controlling ozone formations, they should decisively think of exploring the possibility for a shift in the modelling methodology. In general, the forecasted ozone concentration followed the trend of the observed ozone concentration however the ANN model (M-21) slightly under-predicts the maximum values and over predicts the minimum values. The results suggest that the relationship between the meteorological parameters and the primary air pollutants with ozone might be sufficiently complex which could not be fully captured even by a relatively sophisticated model like ANN. Inputs of getter sophistication are probably required for increasing the prediction accuracy. The intra-comparison of the 34 ANN models so formulated

in the present study reveal that iterative approach should be used for formulating the neural architecture for getting higher accuracy. In the background of the rapid industrialization and urbanization which is adversely affecting climatic conditions and the air quality both at micro and macro level, alert system based on ANN technique could be used for informing the people in advance about the probable ozone concentrations in the respective areas. Such alerts would help the people to take necessary actions at the end which could safeguard their health.

#### Acknowledgements

The authors duly acknowledge Central Pollution Control Board for giving the necessary data for the present study. Support extended by Department of Civil Engineering, IIT Madras and Department of Civil Engineering, Nirma University, Ahmedabad is also duly acknowledged.

## References

- Brimblecombe P (1987) The Big Smoke: A history of air pollution in London since medieval times. Methuen, London.
- Anderson HR (2009) Air Pollution and mortality: A History. Atmos Environ 43: 142-152.
- Mathew RH, Prashant Kumar, Roy M. Harrison (2012) Particles, air quality, policy and health. Royal Society of Chemistry.
- <http://www.ipcc.ch/pdf/assessment-report/ar4/wg3/ar4-wg3-chapter4.pdf>
- Fenger J (2009) Air Pollution in last 50 years: From local to global. Atmos Environ 43: 13-22.
- World Health Organization (2003) Health aspect of air pollution with particulate matter, ozone and Nitrogen dioxide, Report on a WHO Working Group. Germany.
- <http://www.epa.gov/air/ozonepollution/pdfs/health.pdf>
- <http://www.environ.ie/en/Environment/Atmosphere/AirQuality/AirQualityStandardsMonitoring/>
- The Gazette of India (2009) National Ambient Air Quality Standards, India.
- Seinfeld JH, Pandis SN (1998) Atmospheric chemistry and physics: from air pollution to climate change.
- <http://www.elsevierdirect.com/v2/companion.jsp?ISBN=9780122570605>
- Carter WPL, Lloyd AC, Sprung JL, Pitts JN Jr (1979a) Computer Modeling of Smog Chamber Data: Progress in Validating Detailed Mechanisms for the Photooxidation of Propene and n-Butane in Photochemical Smog. Int J Chem Kine 11: 45-101.
- Carter WPL, Winer AM, Darnal KR, Pitts JN Jr (1979b) Smog chamber studies of temperature effects in photochemical smog. Environ Sci Technol 13: 1094-1100.
- Oh SC, Yeo K (1998) Modelling and Simulation of Ozone formation from a Propene-Nitrogen Oxide Wet air mixture in a smog-chamber. Korean J Chem Eng 15: 20-27.
- Comrie AC (1997) Comparing neural networks and regression models for ozone forecasting. Journal of the Air & Waste Management Association 47: 653-663.
- Gardner MW, Dorling SR (1999) Neural network modelling and prediction of hourly NO<sub>x</sub> and NO<sub>2</sub> concentrations in urban air in London. Atmos Environ 31: 709-719.
- Hadjiiski L, Hopke PK (2000) Application of artificial neural networks to modeling and prediction of ambient ozone concentrations. J Air Waste Manag Assoc 50: 849-901.
- Kolehmainen M, Martikainen H, Ruuskanen J (2001) Neural network and periodic components used in air quality forecasting. Atmos Environ 35: 815-825.
- Nagendra SMS, Khare M (2006) Artificial neural network approach for modelling nitrogen dioxide dispersion from vehicular exhaust emissions. Ecol Model 190: 99-115.
- Yi J, VR Prybutok (1996) A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environmental Pollution 92: 349-357.
- Gardner MW, Dorling SR (2000) Statistical surface ozone models: an improved methodology to account for non-linear behavior: A summary of the AMS Workshop on Dispersion Model Performance. Atmos Environ 34: 21-34.
- Chaloulakou A, Saisana M, Spyrellis N (2003) Comparative assessment of neural networks and regression models for forecasting summertime ozone in Athens. Sci Total Environ 313: 1-13.
- Abdul-Wahab SA, Al-Alawi SM (2002) Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. Environ Modell Softw 17: 219-228.
- Crowe W, DeFries TH (1996) Use of observation based models to predict ambient ozone levels. Presentation to EPA/Office of Air Quality Planning Standards (OAQPS).
- Agirre E, Anta A, Barron LJR, Albizu MV (2007) A neural network based model to forecast hourly ozone levels in rural areas in the Basque Country. In: Air Pollution XV, WIT Transactions on Ecology and the Environment, Witpress, Southampton 109-118.
- Agirre-Basurko E, Ibarra-Berastegi G, Madariaga I (2006) Regression and multilayer perceptron-based models to forecast hourly O<sub>3</sub> and NO<sub>2</sub> levels in Bilbao area. Environ Modell Softw 21: 430-446.
- Haykin S (2001) Neural Networks: A Comprehensive Foundation. In (2<sup>nd</sup> eds) Pearson Education Inc, New Delhi, India.
- Mohan M, Kandya A, Yadav M (2011) An evaluation and comparison of the various statistical and deterministic techniques for forecasting the concentration of criteria air pollutants. Int J Environ Pollut 44: 96-105.
- Mohan M, Kandya A (2007) An Analysis of the Annual and Seasonal Trends of Air Quality Index of Delhi. Environ Monit Assess 131: 267-277.

This article was originally published in a special issue, [Urban Air Pollution: Measurements, Physicochemical characteristics, Exposure, Health and Dispersion Modelling](#) handled by Editor(s). Dr. Prashant Kumar, University of Surrey, UK