



## Systematic Approach for the Prediction of Ground-Level Air Pollution (around an Industrial Port) Using an Artificial Neural Network

Mahad S. Baawain<sup>1\*</sup>, Aisha S. Al-Serih<sup>2</sup>

<sup>1</sup> *Department of Civil and Architectural Engineering, Sultan Qaboos University, P.O. Box 33, Alkhoud 123, Muscat, Sultanate of Oman*

<sup>2</sup> *Department of Biology, Sultan Qaboos University, P.O. Box 36, Alkhoud 123, Muscat, Sultanate of Oman*

---

### ABSTRACT

The prediction of air pollution levels is critical to enable proper precautions to be taken before and during certain events. In this paper a rigorous method of preparing air quality data is proposed to achieve more accurate air pollution prediction models based on an artificial neural network (ANN). The models consider the prediction of daily concentrations of various ground-level air pollutants, namely CO, PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub>, SO<sub>2</sub>, H<sub>2</sub>S, and O<sub>3</sub>, which were measured by an ambient air quality monitoring station in Ghadafan village, located 700 m downwind of the emissions of Sohar Industrial Port on the Al-Batinah coast of Oman. The training of the models is based on the multi-layer perceptron (MLP) method with the Back-Propagation (BP) algorithm. The results show very good agreement between the actual and predicted concentrations, as the values of the coefficient of multiple determinations (R<sup>2</sup>) for all ANN models exceeded 0.70. The results also show the importance of temperature in the daily variations of O<sub>3</sub>, SO<sub>2</sub>, and NO<sub>x</sub>, whilst the wind speed and wind direction play significant roles in the daily variations of NO, CO, NO<sub>2</sub>, and H<sub>2</sub>S. PM<sub>10</sub> concentrations are influenced by almost all the measured meteorological parameters.

**Keywords:** Air pollution; Industrial port; Artificial neural network; Oman.

---

### INTRODUCTION

Atmospheric pollution sourced by industrial activities is of a principal concern worldwide. Primary and secondary pollutants can be of a main concern to degrade the ambient air quality in areas adjacent to the industrial sites and put people at risk of daily exposure to them (Baawain *et al.*, 2007). Such impacts dedicate the importance of ground-level air pollution forecasting as an effective alarming system that would allow time to generate a particular response in the case of severe episodes (Bishop, 1995). Moreover, developing a satisfactory alarming system is vital to supply the local environmental agencies with inputs into decisions regarding abatement measures and air quality management.

Even so, predicting air quality or developing an alarming system is not a very simple task due to the fact that incomplete or lack of reliable environmental data often come across in environmental research. This situation may be a result of insufficient sampling, mistakes in measurements and obvious mistakes in data acquisition (Junninen *et al.*,

2004). Whatever the case, discontinuities in data represent an important obstacle for time series proposition schemes that are usually need relentless data as a status for satisfactory effectiveness (Sahin *et al.*, 2011).

Therefore, a general modeling approach that can deal with discontinuous and noise in data as well as capturing the complex interactions within data with satisfactory efficiency is necessary for obtaining reliable forecasting outcomes. Artificial Neural Network (ANN) models seem to be a good choice for the reason that they have been found to perform remarkably well in capturing complex interactions within the given input parameters (Baawain *et al.*, 2007).

A review of the available literature illustrated that ANNs have been applied successfully to predict the ground-level air pollution. Chan and Jian (2013), de Gennaro *et al.* (2013), Cheng *et al.* (2012), Gobakisa *et al.* (2011), Kurt and Oktay (2010) have shown that neural networks are promising tools for air quality prediction in comparison with other statistical models like regression-based models. Moreover, ANNs, in particular the multilayer perceptron (MLP), perform better when dealing with highly non-linear systems such as the pollution-weather phenomenon (Gardner and Dorling, 1998; Abdul-Wahab and Al-Alawi, 2008). ANNs are capable of predicting pollution episodes reasonably well (Gardner and Dorling, 1999; Dutot *et al.*, 2007; Moustris *et al.*, 2010). In addition, it has been

---

\* Corresponding author.

Tel.: +968 24143122; Fax: +968 24414012  
E-mail address: msab@squ.edu.om

demonstrated that tests using previous lagged concentrations as inputs to the model make better ANN predictions when compared with models based on no previous lagged concentrations (Gardner and Dorling, 1999; Ballester *et al.*, 2002; Tecer, 2007; Cai *et al.*, 2009). Furthermore, ANN models are glimpsed as promising tools for middle long-term forecasts in time scales of days, despite the fact that expanding the forecasting period reduces the prediction correctness of the models (Moustris *et al.*, 2010).

However, the successful performance of ANNs is immensely affected by the quality of the input data which consequently influence the success of the forecasting. Considering this, the present study conceived the possibility that a more rigorous approach in preparing the data and systematic methodology focused on dealing with limitations like missing data and noise would provide a better air quality prediction using ANN.

An opportunity arose to test this approach using data from a single ambient air quality monitoring point within a regional community adjacent to a main new industrial complex at Sohar Industrial Port (SIP) on Al-Batinah coast of Oman. The study will employ the proposed systematic methodology to develop ANN prediction models for daily concentration of ground-level pollutants, including CO, PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub>, SO<sub>2</sub>, H<sub>2</sub>S, and O<sub>3</sub> measured in the area adjacent to Sohar Industrial Port.

### Artificial Neural Networks

Artificial Neural Networks are computing systems, motivated by biological models, and made up of a number of easy and highly interconnected processing components, which process information by its dynamic state response to external inputs (Nelson and Illingworth, 1991). The processing components called neurons are organized into inter-connected layers (Nelson and Illingworth, 1991; Fausett, 1994; Haykin, 1998). The number of layers in the neural network can vary from a single layer to multiple layers. The layer that obtains the inputs from the external environment is called the input layer. It typically presents no function other than the buffering of the input level. The network outputs are generated from the output layer. Hidden layers, on the other hand are sometimes linked to a “black box” within which the input data are mapped into outputs utilizing suitable activation function(s).

## METHODS

### Area Description

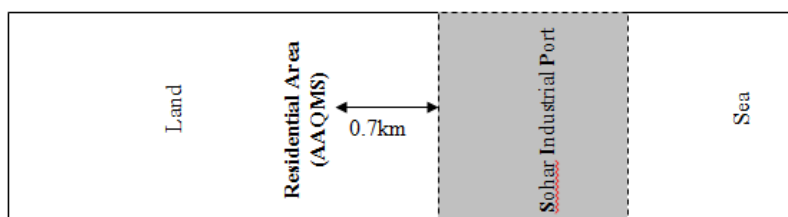
The ambient air quality data used in the study were

collected by the Ministry for Environment and Climate Affairs (MECA) from a typical urban and industrial area adjacent to the Port of Sohar. Sohar (24°22'N, 56°45'E) is among one of the most rapidly growing cities in Oman outside the capital city of Muscat and is located 220 km to the North West. The Sohar Industrial Port (SIP) is some 35 km to the north of the city of Sohar and covers a total area of 13.87 km<sup>2</sup>. The SIP features three major industrial clusters namely petrochemicals, base metal processing and cargo handling. A cooling water and waste water processing and logistic elements have also been built into the port. Most of these industries are involved in emitting a variety of air pollutants in the surrounding atmosphere like CO, PM<sub>10</sub>, NO, NO<sub>2</sub>, NO<sub>x</sub>, SO<sub>2</sub>, H<sub>2</sub>S, VOC, H<sub>x</sub>C<sub>x</sub> or O<sub>3</sub>. The area around the boundaries of SIP is featured by the follow:

- Residential areas of more than 4,000 people (2003 census, Ministry of National Economy) occur in proximity to the south, southeastern and northwestern boundaries of SIP. Most of these settlements are located within a distance of 500 meters from the boundaries of Sohar Industrial Port.
- Agricultural areas which are confined to a narrow strip of 61,500 ha on either sides of the Batinah highway. Generally, the agricultural zones do not extend to the coastline because of saline water intrusion. Additionally, a strip of farm lands approximately 300 meters wide has been purchased as a buffer for the industrial area beyond which are active farmlands and villages.
- Seven schools are distributed to the south, southeastern and northwestern boundaries of the SIP. A total of 4,600 students and staff (2010/2011 academic year census) are at risk of daily exposure to industrial emissions from Sohar Industrial Port.
- A mangrove system exists approximately 2.5 km to northwest of the SIP. It covers an area of 35.2 ha. According to Omani Ministry of Environment and Climate Affairs (2010), this system is ranked as the twelfth largest mangrove system in Oman and one of only four notable mangrove sites along the Batinah coast.

### Data Collection

Hourly records of air quality parameters along with meteorological parameters were acquired from the Ministry of the Environment and Climate Affairs. The data were obtained from the Mobile Air Quality Monitoring Station (MAQMS) located in Ghadafan village (refer to Fig. 1); about 700 m western of the Sohar Industrial Port. The Ministry of the Environment and Climate Affairs chose that particular sampling site to assess the air quality of the nearest residential area located downwind to the emissions



**Fig. 1.** Location of ambient air quality monitoring station in relation to Sohar Industrial Port.

of the Industrial Port.

The mobile station was fitted with chemical monitors and meteorological sensors. All the sensors were operating automatically. Acquired measurements covered a period of four years (2006 to 2009), based on hourly averages. Pollutants measured include PM<sub>10</sub>, O<sub>3</sub>, CO, SO<sub>2</sub>, H<sub>2</sub>S, NO, NO<sub>x</sub>, and NO<sub>2</sub>. Meteorological parameters monitored simultaneously including wind direction, wind speed, relative humidity and air temperature.

The hourly data were then processed into daily format and used to construct a data series presented as a two-dimensional table. The columns represent 13 variables consisting of date, PM<sub>10</sub>, O<sub>3</sub>, CO, SO<sub>2</sub>, H<sub>2</sub>S, NO, NO<sub>x</sub>, NO<sub>2</sub>, wind speed, wind direction, air temperature, and relative humidity. The rows of the table refer to observation date, which is represented by a day-month-year (DD-MM-YY) format. The entire table consists of 1363 rows, or 1363 days of observation.

### ANN Model Developments

This particular study design illustrated the development of ANN models into two stages: data preparation and model development. The data preparation phase included data inspection, selection, and normalization while the ANN-model development stage included data division, network design, and model validation (see Fig. 2).

### Data Preparation

Preparing data for the neural network data analysis is an important and critical step that has an immense impact on the success and performance of the neural network results (Yu *et al.*, 2006). This study data preparation was started by firstly inspecting the data set for missing data and data noise (or outliers). The entire data set covered a period from 01/01/2006 to 31/12/2009. Data inspection resulted in the omission of the period from January–March 2006 due to instrument malfunction. Additionally, some data

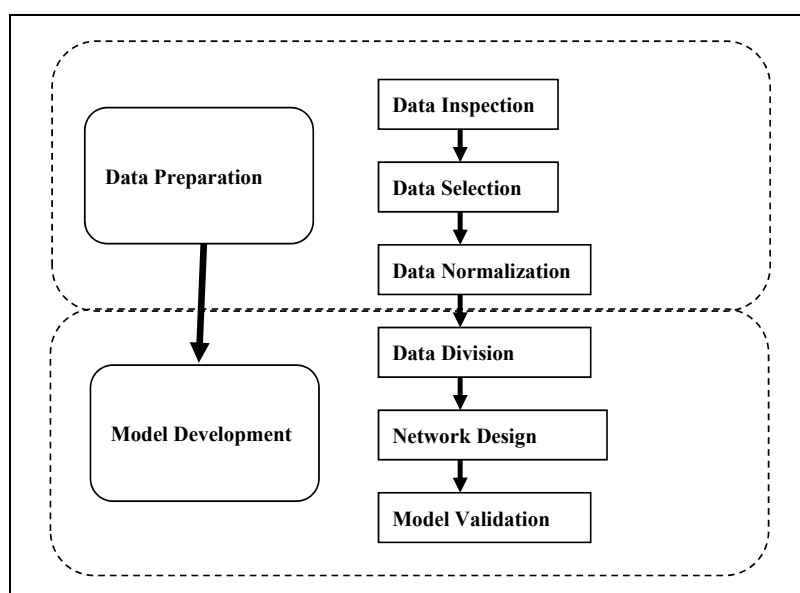
were missed due to instrument calibrations or malfunctions. Rows where errors and/or incomplete information were apparent were also entirely removed. The check of missing data resulted in a removal of 25.17% of available data. As a result, 1020 rows or days of measured observations remained to be included in the neural network models. Moreover, data inspection included the removal of discrepancies in codes or names which left small gaps within the data set. These gaps were filled by the linear interpolation method of Gupta (1999). Finally, extreme values (outliers or noisy data) were removed using the Standardized Score Method, within SPSS (Hisham, 2008). These gaps refilled by the linear interpolation method as well. Reader is encouraged to use imputation methods of missing values in air quality data sets addressed by Junninen *et al.* (2004) and Niska *et al.* (2004).

Data inspection was followed by determination of input variables for modeling. Since a separate model was to be developed for each pollutant, the selection of input variables varied from one model to another. The selection of predictor variables was based on a comprehensive review of the theoretically addressed chemical and weather processes that influence the formation and concentration of atmospheric pollutants, as described by Seinfeld and Pandis (1998), Wayne (1985) and USEPA (2003). Table 1 lists common predictor variables that influence the concentrations of each modeled pollutant.

Eventually, and in order to support the neural network to deal effectively with the data, all the input data were normalized to the range of  $\ll 0, 1 \gg$  by linear scaling. Thus, the new data encountered later by the network module would be successfully scaled if the new data are outside the given range.

### Model Development

The ANN models in this study were developed using the NeuroShell2 (NS2) software from Ward Systems Group Inc. NeuroShell2 is a software program that mimics the human



**Fig. 2.** Two stage development procedures for the Artificial Neural Network Modeling undertaken in this study.

**Table 1.** Common predictor variables proposed for each model.

Model Inputs	Model Output
CO, NO, NO <sub>2</sub> , NO <sub>x</sub> , WS, WD, T, H	O <sub>3</sub>
WS, WD, T, H	PM <sub>10</sub>
H <sub>2</sub> S, WS, WD, H, T	SO <sub>2</sub>
WS, WD, H, T	H <sub>2</sub> S
CO, NO, NO <sub>2</sub> , WS, WD, T, H	NO <sub>x</sub>
CO, O <sub>3</sub> , NO, NO <sub>x</sub> , WS, WD, T, H	NO <sub>2</sub>
CO, NO <sub>x</sub> , NO <sub>2</sub> , WS, WD, H, T	NO
WS, WD, H, T	CO

brain's ability to classify patterns or to make predictions or decisions based upon past experience. NS2 is designed to assist developers who have a minimum of the specialized knowledge required to build Neural Network Models. NS2 is able to learn patterns from training data and make its own predictions (or decisions) when presented with new unseen data.

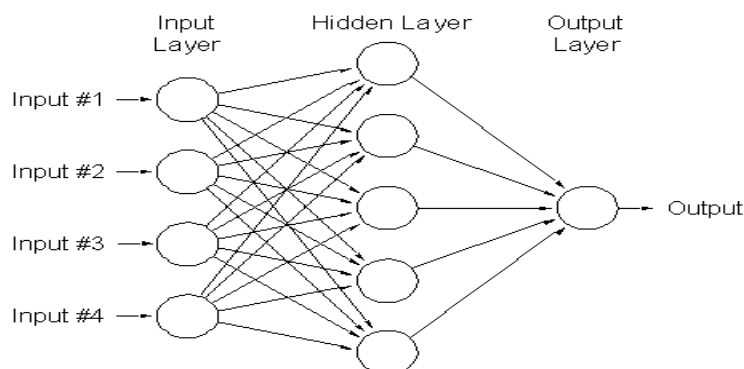
Before running ANN models, the previously prepared data set was split into two subsets for neural network learning. Because, there is no universal rule to determine the size of subsets, the data set for this project was randomly divided into a ratio of 3:1 between training and testing sets, respectively. The definition of these sets used in this study will be as follows:

- Training set: is the largest group of all subsets and used to educate/train the network, through adjusting the weights of links and changing the number of hidden neurons, to come up with the best fit between the actual and predicted output.
- Testing set: is the group of data given to the network still in the training phase which will be used later by the network to test the accuracy of the results by evaluating the minimum error. The testing set prevents overtraining networks so they will generalize well when provided with new data.
- The feed-forward back-propagation (BP) multi-layer preceptor (MLP) neural network architecture was selected for the air quality modeling undertaken here. Fig. 3 depicts the basic elements of the MLP-BP standard networks as used in the current study, in which the basic building block of standard nets is the simulated neurons

which are interconnected into a network of neurons that are eventually arranged into three layers such as input, hidden and output layer. Each link between two neurons from different consecutive layers in the MLP-BP is assigned a weight that defines the nature of the relationship between the neurons. The neuron's output is multiplied by the weight before being used as input to the neuron in the following layer. Each neuron in the hidden and output layers sums all of the received inputs to be used as an output value according to a predefined transfer function. The input data, during the training of a network, are propagated in a feed-forward manner to produce output data according to the weights and transfer function. The prediction error is then determined from the difference between the produced output and the actual output. The weights of the links are adjusted to minimize the prediction errors according to the training algorithm being used. The network is considered well trained when the sum of all the errors in the network reaches a global minimum (Baawain *et al.*, 2005).

The main objective of this study was to build up ANN simulations of typical air pollutant concentrations as measured by MECA from 2006 to 2009. Thus, separate neural network model was developed for each pollutant (PM<sub>10</sub>, O<sub>3</sub>, CO, SO<sub>2</sub>, H<sub>2</sub>S, NO, NO<sub>x</sub>, and NO<sub>2</sub>). For each model chosen types of pollutants as well as meteorological parameters were selected as input parameters as already summarized in Table 1. The models used the previous conditions of all input parameters in order to predict the concentration of each pollutant on the next day. As an example, Fig. 4 shows the network structure used to build up the O<sub>3</sub> model.

ANN models were run in a supervised manner based on trial and error technique by which several alternative adjustments were incorporated to improve the model performance. Those adjustments included alternative use of different weight update patterns, different number of neurons in the hidden layer, different numbers of learning events (epochs), as well as employment of different activation functions. The training of the network continued until achieved the highest correlation between the genuine and predicted output which is expressed by *coefficient of multiple determinations* ( $R^2$ ) and normalized root means square error (NRMSE). Accordingly, a perfect fit would outcome in

**Fig. 3.** Basic elements of MLP standard network.

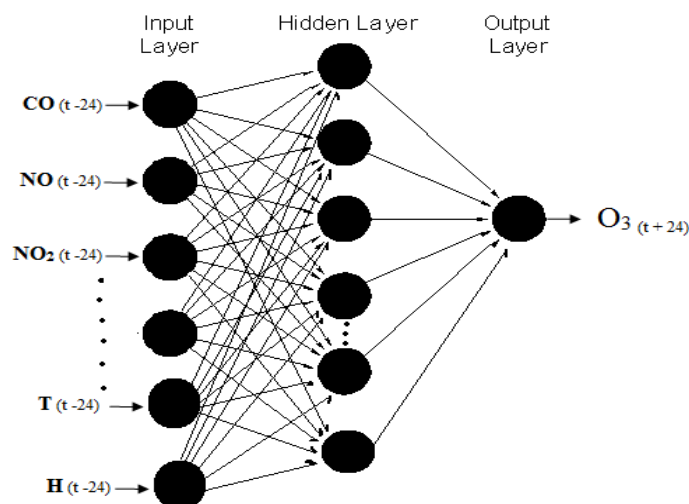


Fig. 4. Network structure used to build up O<sub>3</sub> model.

an R squared value of 1, a very good fit beside 1 and a very poor fit less than 0. On the other side, the smaller the value of NRMSE, the better is the performance of the model.

To validate the models, hidden test set of data were used for each model to evaluate the efficiency of the trained ANN models in generalizing the problem when dealing with new unseen data. The predicted results were then compared with the actual values and that was expressed by calculating R<sup>2</sup>.

## RESULTS AND DISCUSSION

### ANN Models Performance

The MLP-BP architectures that yielded the best ANN models are summarized in Table 2. Generally, the best performing ANN models involved the use of a logistic activation function in the hidden layer, the exception being SO<sub>2</sub> which worked best with a Gaussian function. Additionally, most trained networks worked best using the Turbo-Prop weight update with rotation pattern. The results showed excellent performance for the developed networks of SO<sub>2</sub>, H<sub>2</sub>S, and CO according to values of R<sup>2</sup> (0.94, 0.93, and 0.90, respectively) and NRMSE (3.2%, 8.7% and 12.1%, respectively). Very good results were obtained for O<sub>3</sub>, PM<sub>10</sub> and NO<sub>2</sub> networks with R<sup>2</sup> of 0.88, 0.82, and 0.84, respectively, and NRMSE of 13.9%, 17.4% and 16.7%, respectively. The performance of the networks developed for NO<sub>x</sub> and NO were relatively lower than the other models as evident from the values of R<sup>2</sup> (0.73, and 0.73, respectively) and NRMSE (19.2% and 24.3%, respectively). Table 3 shows the sensitivity of network to training cycles (epochs) when keeping the hidden neurons and activation function as shown in Table 2. It can be seen that all developed networks are relatively stable for a range of ± 20% of training cycles as evident from the obtained R<sup>2</sup> values. Figs. 5 to 12 provide visual presentation of the obtained training and testing results for all developed networks.

The MLP-BP ANN architecture that yielded best results for O<sub>3</sub> prediction (1-day ahead) consisted of one hidden layer

with 39 hidden neurons using logistic activation function. This network was trained with 1045 epochs with Turbo-Prop weight updates and rotation pattern selection. The performance of ANN model for O<sub>3</sub> prediction (1-day ahead) was very good as the Coefficient of Multiple Determination (R<sup>2</sup>) was 0.94 for the training set and 0.88 for the testing set (see Fig. 5). This result indicates that approximately 94% and 88% of the variability in the ozone concentrations, for both training and testing data, could be explained by the selected input variables used for the model development, namely: CO, NO, NO<sub>2</sub>, NO<sub>x</sub>, wind speed, wind direction, temperature, and humidity. The best performance for O<sub>3</sub> is probably due to the mechanism of O<sub>3</sub> production. Fig. 13 shows the significant contribution of ambient air temperature and humidity which are directly related to the solar radiation and are, therefore, highly important for ozone production (Wallace and Hobbs, 1977; Seinfeld and Pandis, 1998). Yet, this performance can be reasonably improved by introducing the measures of precursor emissions, like VOCs, necessary for O<sub>3</sub> formation in the troposphere (Wallace and Hobbs 1977; Goody 1995; Seinfeld and Pandis 1998; USEPA 2003).

ANN models aiming to predict PM<sub>10</sub> concentrations (1-day ahead) consisted of one hidden layer with 47 hidden neurons using logistic activation function. This network was trained with 1200 epochs with Turbo-Prop weight updates and rotation pattern selection. Fig. 6 shows the obtained R<sup>2</sup>, 0.77 for the training set and 0.82 for testing data, which explains the importance of selected input variables including wind speed; wind direction, temperature and humidity.

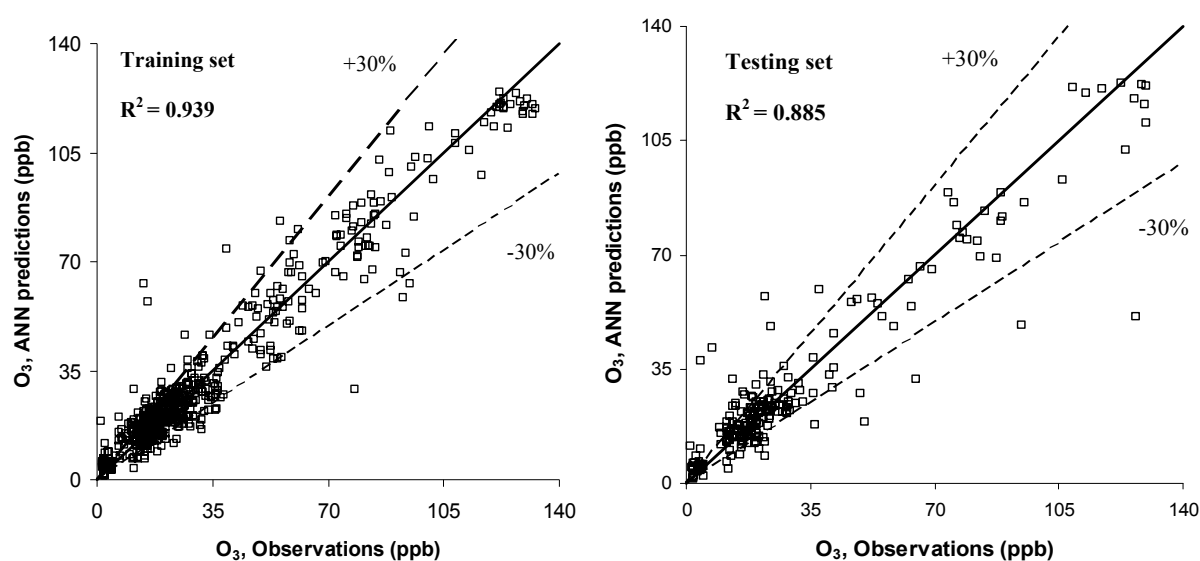
The best ANN results for SO<sub>2</sub> prediction (1-day ahead) consisted of one hidden layer with 31 hidden neurons using Gaussian activation function. This network was trained with 8707 epochs with Turbo-Prop weight updates and rotation pattern selection. The ANN performance for SO<sub>2</sub> prediction (1-day ahead) was excellent with R<sup>2</sup> values of 1.00 and 0.94 for training and testing sets, respectively (Fig. 7). The selected input variables for SO<sub>2</sub> prediction, including H<sub>2</sub>S, wind speed, wind direction, temperature, and humidity, played a major role in influencing the variability of SO<sub>2</sub> levels. The good performance of SO<sub>2</sub> prediction is

**Table 2.** Best architectures for ANN models.

Model	Hidden Neurons No.	Activation Function	Epochs	R <sup>2</sup> Testing	NRMSE Testing
O <sub>3</sub>	39	Logistic	1045	0.88	13.9%
PM <sub>10</sub>	47	Logistic	54	0.82	17.4%
SO <sub>2</sub>	31	Gaussian	8707	0.94	3.2%
H <sub>2</sub> S	34	Logistic	98955	0.93	8.7%
NO <sub>x</sub>	49	Logistic	12758	0.73	19.2%
NO <sub>2</sub>	50	Logistic	2281	0.84	16.7%
NO	33	Logistic	171740	0.73	24.3%
CO	49	Logistic	2884	0.90	12.1%

**Table 3.** Sensitivity analysis of the networks to training cycles.

Model	Epochs	R <sup>2</sup>		Epochs	R <sup>2</sup>	
		Testing	Testing		Testing	Testing
O <sub>3</sub>	850	0.83	0.88	1045	0.88	0.86
PM <sub>10</sub>	40	0.79	0.82	54	0.82	0.81
SO <sub>2</sub>	7000	0.89	0.94	8707	0.94	0.91
H <sub>2</sub> S	8000	0.86	0.93	98955	0.93	0.89
NO <sub>x</sub>	10200	0.70	0.73	12758	0.73	0.69
NO <sub>2</sub>	1800	0.82	0.84	2281	0.84	0.83
NO	137000	0.69	0.73	171740	0.73	0.72
CO	2300	0.88	0.9	2884	0.9	0.89

**Fig. 5.** ANN predicted versus observation values for O<sub>3</sub>.

possibly due to the strong influence of the selected input variables, especially H<sub>2</sub>S which has a high probability in the formation of SO<sub>2</sub> (Wayne, 1985; Goody, 1995).

The MLP-BP ANN architecture that yielded best results for H<sub>2</sub>S prediction (1-day ahead) consisted of one hidden layer with 34 hidden neurons using the logistic activation function. This network was trained with 98955 epochs with Turbo-Prop weight updates and rotation pattern selection. The ANN model performed very well for H<sub>2</sub>S prediction (1-day ahead). The R<sup>2</sup> ranged from 0.91 for the training set to 0.93 for testing set (see Fig. 8), is presumably reflective of a high dependence of H<sub>2</sub>S concentrations on the meteorological conditions, namely: wind speed, wind

direction, temperature, and humidity, selected as inputs for H<sub>2</sub>S modeling.

The best MLP-BP ANN used to predict NO<sub>x</sub> concentrations (1-day ahead) consisted of one hidden layer with 49 hidden neurons using the logistic activation function. This network was trained with 12,758 epochs with Turbo-Prop weight updates and rotation pattern selection. The ANN satisfactorily predicted values of NO<sub>x</sub> concentrations resulted from complex relationships between CO, NO, NO<sub>2</sub> concentrations and meteorological conditions such as wind speed wind direction, temperature, and humidity, with the R<sup>2</sup> ranging from 0.76 to 0.73 (Fig. 9).

The highest performance for NO<sub>2</sub> prediction (1-day ahead)

was obtained using one hidden layer with 50 hidden neurons applying the logistic activation function. This network was

trained with 2281 epochs with Turbo-Prop weight updates and rotation pattern selection. The performance of ANN

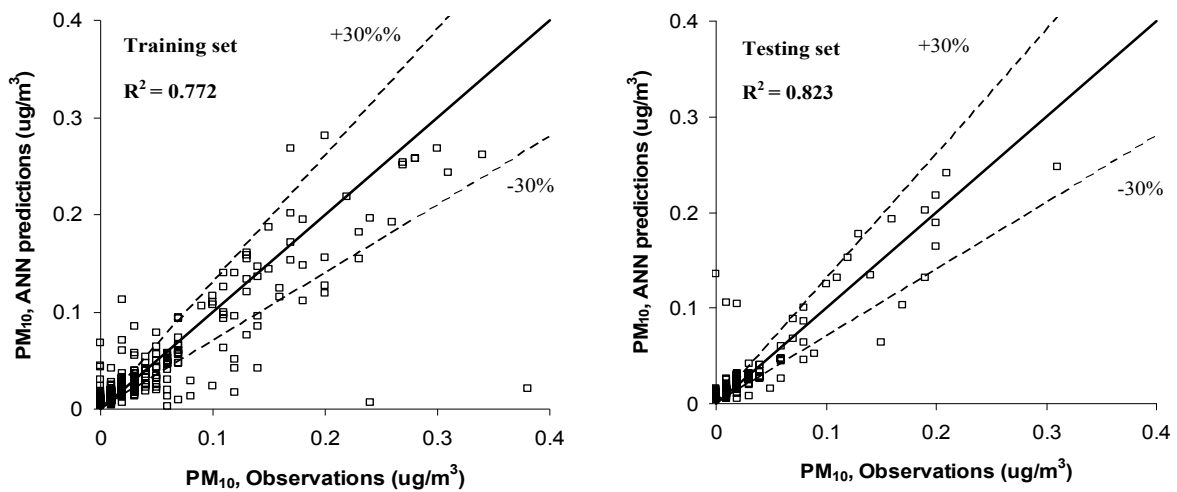


Fig. 6. ANN predicted versus observation values for PM<sub>10</sub>.

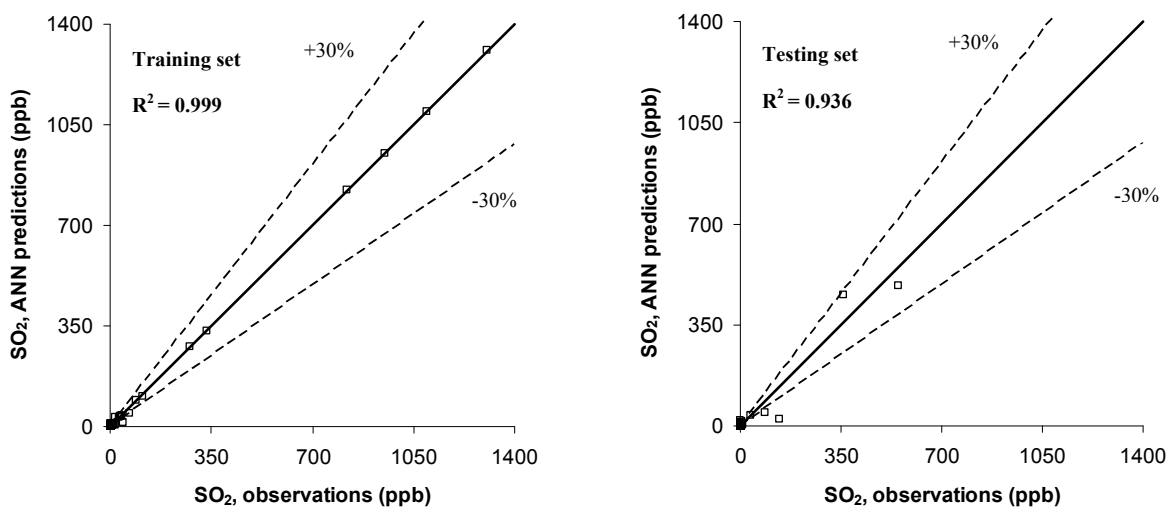


Fig. 7. ANN predicted versus observation values for SO<sub>2</sub>.

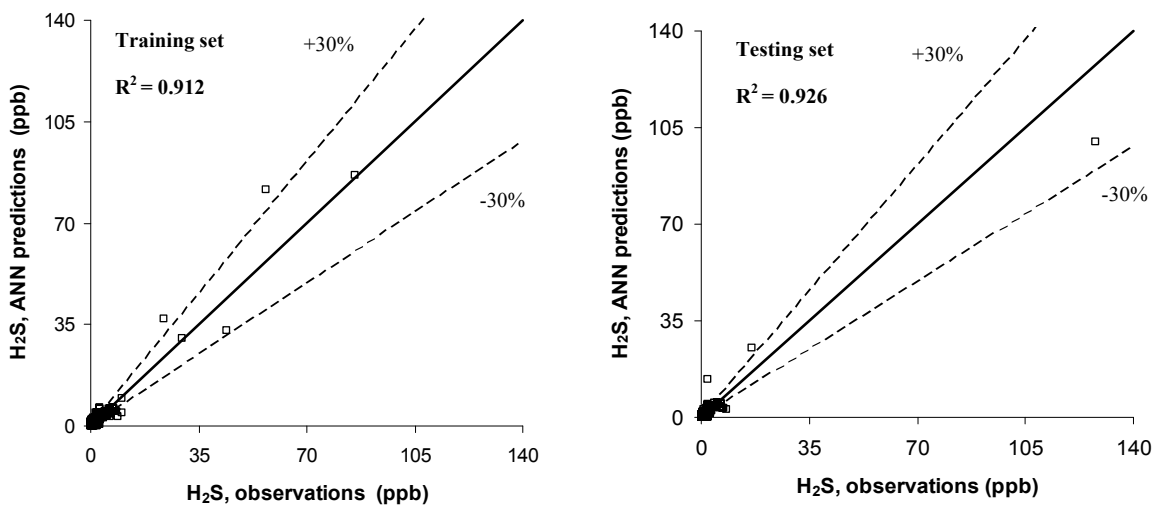


Fig. 8. ANN predicted versus observation values for H<sub>2</sub>S.

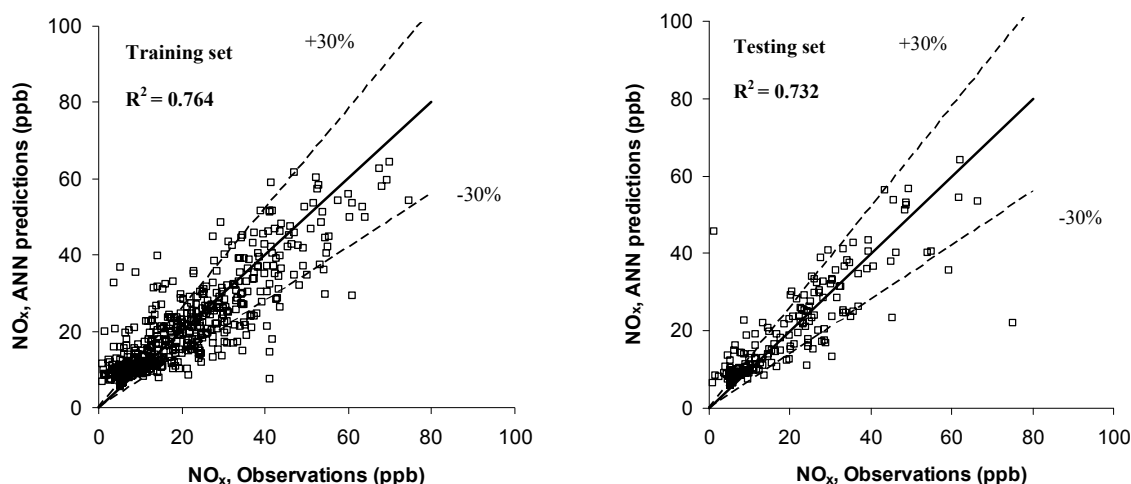


Fig. 9. ANN predicted versus observation values for  $\text{NO}_x$ .

for  $\text{NO}_2$  prediction was good as  $R^2$  values were between 0.88 and 0.84 for the training and testing sets, respectively. Fig. 10 illustrates the high performance of the developed  $\text{NO}_2$  model through the good agreement shown between measured and ANN predicted values in both training and testing data sets.

Similarly, the best MLP-BP ANN architecture for NO prediction (1-day ahead) consisted of one hidden layer with 33 hidden neurons using the logistic activation function. This network was trained with 171,746 epochs with Vanilla weight updates and random pattern selection. The ANN prediction of NO concentrations was in the range of 0.72 and 0.73 for the training and testing sets in terms of  $R^2$  performance (Fig. 11). Although these are small values, yet, they are still acceptable to explain the complex relationships between CO,  $\text{NO}_x$ , and  $\text{NO}_2$  concentrations and meteorological conditions (most importantly: wind speed, wind direction, temperature, and humidity) that influence the formation of NO.

The MLP-BP ANN architecture that yielded best results for CO prediction (1-day ahead) consisted of one hidden layer with 49 hidden neurons using logistic activation

function. This network was trained with 2884 epochs with Momentum weight updates and random pattern selection. ANN model performed very well in predicting the CO concentrations (1-day ahead) as a function of meteorological conditions, chiefly, wind speed, wind direction, temperature, and humidity (see Fig. 12). The  $R^2$  ranged from 0.86 to 0.88 and reflects the influence of meteorological conditions in regulating the variability in CO concentrations.

#### RELATIVE CONTRIBUTION OF INPUTS ON DEVELOPED NETWORKS

The contribution of inputs on modeled parameters was derived from the analysis of the weights of the trained neural networks. Fig. 13 shows the relative contributions of the inputs (chemical factors and meteorological conditions) for each modeled parameter. It can be seen that the meteorological conditions have stronger influence on the concentrations of air pollutants at the studied location. It should be noted that the location for prediction of the modeled parameters is about 700 m from the industrial area. Therefore, the meteorological conditions might have

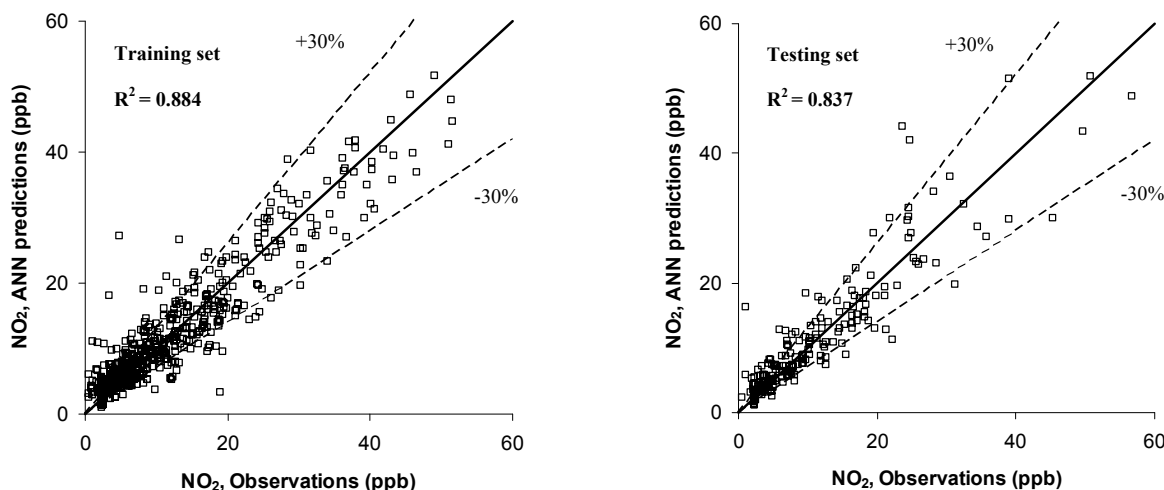


Fig. 10. ANN predicted versus observation values for  $\text{NO}_2$ .



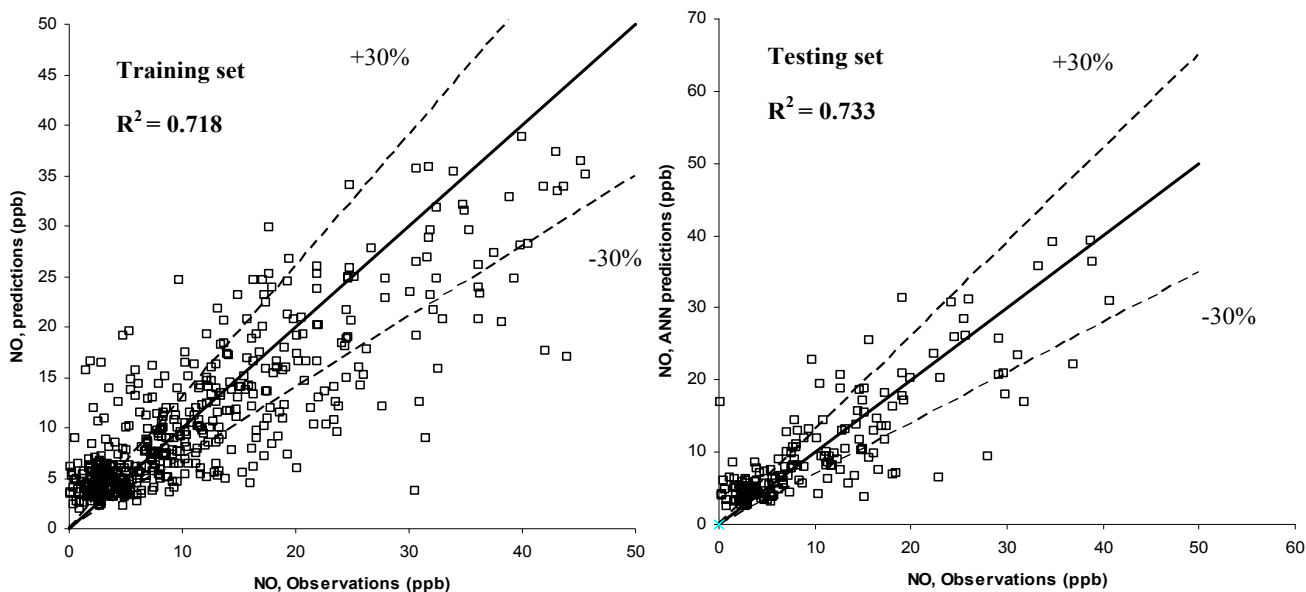


Fig. 11. ANN predicted versus observation values for NO.

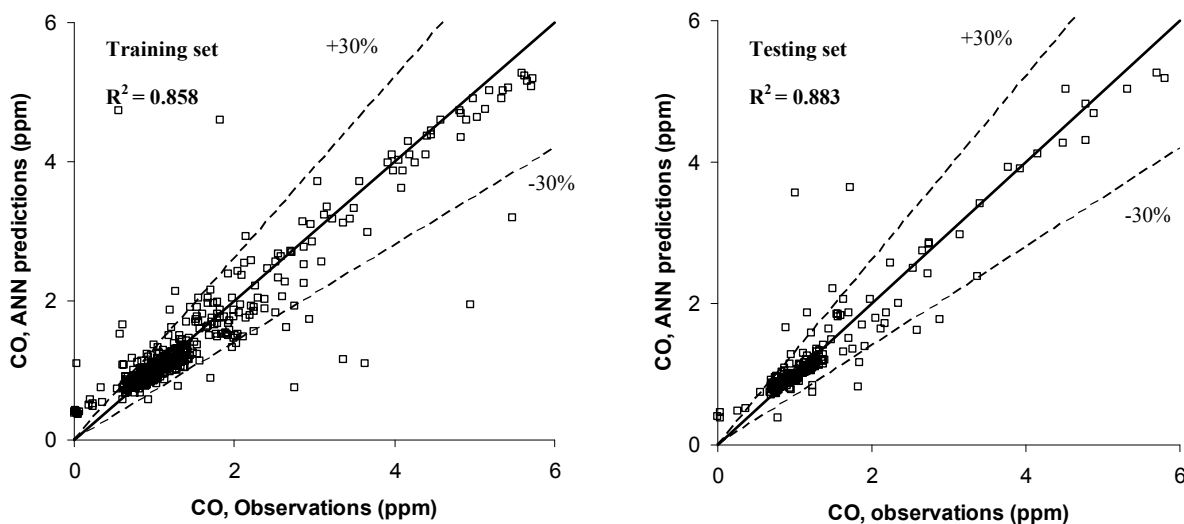


Fig. 12. ANN predicted versus observation values for CO.

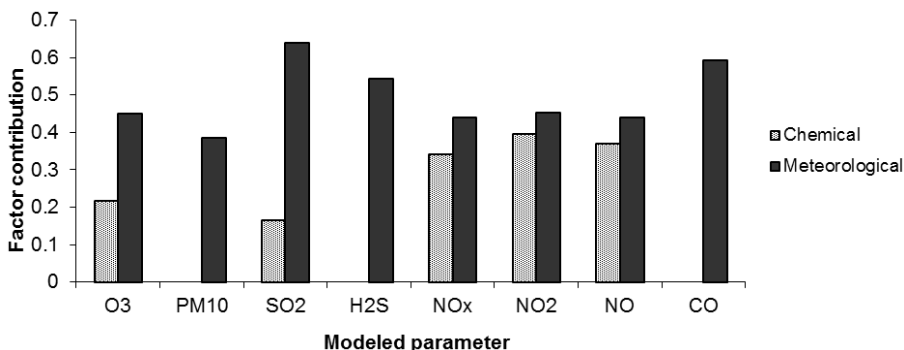


Fig. 13. Relative contribution of inputs on modeled parameters.

higher influence on the dispersion of the air pollutants from their sources to the location of interest compared to the chemical factors in this case.

**CONCLUSIONS**

This study has investigated the potential use of a

systematic approach to develop Artificial Neural Network (ANN) predicting models for the ground-level air pollution, one day in advance, at a specific receptor area nearby Sohar Industrial Port. The goal was to determine the concentration of pollutants, including O<sub>3</sub>, PM<sub>10</sub>, SO<sub>2</sub>, H<sub>2</sub>S, NO<sub>x</sub>, NO<sub>2</sub>, NO, and CO, in the atmosphere according to their relationship with the previous day air quality data and meteorological conditions.

Each pollutant was separately modeled. Each ANN model was trained using previous day conditions in order to predict the next day concentrations. The ANN models were trained by historical daily time series of air quality measurement as well as meteorological measurements. The models were developed using feed forward multi-layer perceptron (MLP) technique based on the back-propagation algorithm (BP).

This study has found that generally air pollution concentrations using the Sohar data set have been well predicted, using ANN models as the coefficient of multiple determinations (R<sup>2</sup>) was found to exceed 0.7 for both training and testing data sets. The findings of this study provide a system of air pollution forecasting (1-day ahead) for Ghadafan village located nearby Sohar Industrial Port. This would be of great benefit for environmentalists and stakeholders to create an early alert of air quality for the public so that they can take the necessary precautions.

## FUTURE PROJECTIONS

The predictions of the developed models were based on the limited history of air quality and were restricted for one sampling location. However, the benefit of ANN model prediction can be improved by incorporating the following aspects:

- Routinely update the current ANN models due to the expansion of the industrial activities in the port.
- Periodic maintenance of monitoring stations in order to get more consistent data and more accurate models.
- Investigation of the validity of ANN models for different areas around Sohar Industrial Port and many other areas in Oman.
- Use emission data and episode levels definition as inputs data along with ambient air quality data for further validation of ANN predictions.
- Improve the outputs of current ANN models into indices using the Air Quality Index in order to aid the general public with simple understanding of air quality information.
- Explore the capability of ANN models in predicting air quality 2-days or 3-days in advance.

## REFERENCES

- Abdul-Wahab, S.A. and Al-Alawi, S.M. (2008). Prediction of Sulfur Dioxide (SO<sub>2</sub>) Concentration Levels from the Mina Al-Fahal Refinery in Oman Using Artificial Neural Networks. *Am. J. Environ. Sci.* 4: 473–481.
- Baawain, M.S., Nour, M.H., El-Din, A.G. and El-Din, M.G. (2005). ENSO Prediction Using SOI and NIÑO3 as Onset Indicators: Application of Artificial Neural Networks. *J. Environ. Eng. Sci.* 4: 113–121.
- Baawain, M.S., El-Din, M.G. and Smith, D.W. (2007). Artificial Neural Networks Modeling of Ozone Bubble Columns: Mass Transfer Coefficient, Gas Hold-Up, and Bubble Size. *Ozone Sci. Eng.* 29: 343–352.
- Ballester, E.B., Valls, G.C.I., Carrasco-Rodriguez, J.L., Olivas, E.S. and Valle-Tascon, S.deI. (2002). Effective 1-Day Ahead Prediction of Hourly Surface Ozone Concentrations in Eastern Spain Using Linear Models and Neural Networks. *Ecol. Modell.* 156: 27–41.
- Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford.
- Cai, M., Yin, Y. and Xie, M. (2009). Prediction of Hourly Air Pollutant Concentrations near Urban Arterials Using Artificial Neural Network Approach. *Transp. Res. Part D* 14: 32–41.
- Chan, K.Y. and Jian, L. (2013). Identification of Significant Factors for Air Pollution Levels Using a Neural Network Based Knowledge Discovery System. *Neurocomputing* 99: 564–569.
- Cheng, S., Li, L., Chen, D. and Li, J. (2012). A Neural Network Based Ensemble Approach for Improving the Accuracy of Meteorological Fields Used for Regional Air Quality Modeling. *J. Environ. Manage.* 112: 404–414.
- de Gennaro, G., Trizio, L., Di Gilio, A., Pey, J., Pérez, N., Cusack, M., Alastuey, A. and Querol, X. (2013). Neural Network Model for the Prediction of PM<sub>10</sub> Daily Concentrations in Two Sites in the Western Mediterranean. *Sci. Total Environ.* 463–464: 875–883.
- Dutot, A.L., Rynkiewicz, J., Steiner, F.E. and Rude, J. (2007). A 24-h of Ozone Peaks and Exceedance Levels Using Neural Classifiers and Weather Predictions. *Environ. Modell. Softw.* 22: 1261–1269.
- Faussett, L. (1994). *Neural Networks: Architectures, Algorithms, and Applications*, Prentice-Hall, Inc., New Jersey.
- Gardner, M.W. and Dorling, S.R. (1998). Artificial Neural Networks (The Multilayer Perceptron): A Review of Applications in the Atmospheric Sciences. *Atmos. Environ.* 32: 2627–2636.
- Gardner, M.W. and Dorling, S.R. (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.* 33: 709–719.
- Gobakisa, K., Kolokotsab, D., Synnefac, A., Saliari, M., Giannopoulouc, K. and Santamourisc, M. (2011). Development of a Model for Urban Heat Island Prediction Using Neural Network Techniques. *Sustainable Cities Soc.* 1: 104–115.
- Goody, R.M. (1995). *Principles of Atmospheric Physics and Chemistry*, Oxford University Press.
- Haykin, S. (1998). *Neural Networks: A Comprehensive Foundation* (2<sup>nd</sup> ed), Prentice-Hall, Inc, New Jersey.
- Hisham, M.B. (2008). Detecting Outliers. Available on the Internet at <http://www.hishammb.net/workshopfeb2008/outlierinspss.pdf>.
- Junninen, H., Niska, H., Tuppurainen, K., Ruuskanen, J. and Kolehmainen, M. (2004). Methods for Imputation of Missing Values in Air Quality Data Sets. *Atmos. Sci.* 38: 2895–2907.

- Kurt, A. and Oktay, A.B. (2010). Forecasting Air Pollutant Indicator Levels with Geographic Models 3 days in Advance Using Neural Networks. *Expert Syst. Appl.* 37: 7986–7992.
- Ministry of Environment and Climate Affairs (MECA) (2010). *Mangroves in the Sultanate of Oman*, 3rd Edition.
- Moustris, K.P., Ziomas, I.C. and Paliatsos, A.G. (2010). 3-Day-Ahead Forecasting of Regional Pollution Index for the Pollutants NO<sub>2</sub>, CO, SO<sub>2</sub>, and O<sub>3</sub> Using Artificial Neural Networks in Athens, Greece. *Water Air Soil Pollut.* 209: 29–43.
- Nelson, M.M. and Illingworth, W.T. (1991). *A Practical Guide to Neural Nets*, Addison-Wesely, Boston, USA.
- Niska, H., Hiltunen, T., Karppinen, A., Ruuskanen, J. and Kolehmainen, M. (2004). Evolving the Neural Network Model for Forecasting Air Pollution Time Series. *Eng. Appl. Artif. Intell.* 17: 159–167.
- Şahin, U.A., Bayat, C. and Uçan, O.N. (2011). Application of Cellular Neural Network (CNN) to the Prediction of Missing Air Pollutant Data. *Atmos. Res.* 101: 314–326.
- Seinfeld, J.H. and Pandis, S.N. (1998). *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley & Sons, Inc., Canada.
- Tecer, L.H. (2007). Prediction of SO<sub>2</sub> and PM Concentrations in a Coastal Mining Area (Zonguldak, Turkey) Using an Artificial Neural Network. *Pol. J. Environ. Stud.* 16: 633–638.
- United States Environmental Protection Agency (2003). Guidelines for Developing an Air Quality (Ozone and PM<sub>2.5</sub>) Forecasting Program, EPA-456/R-03-002.
- Wallace, J.M. and Hobbs, P.V. (1977). *Atmospheric Science: An Introductory Survey*, Academic Press, New York.
- Wayne, R.P. (1985). *Chemistry of Atmosphere: An Introduction to the Chemistry of the Atmosphere of Earth, the Planets, and their Satellites*, Clarendon Press, Oxford.
- Yu, L., Wang, S. and Lai, K.K. (2006). An Integrated Data Preparation Scheme for Neural Network Data Analysis. *IEEE Trans. Knowl. Data Eng.* 18: 217–230.

Received for review, June 11, 2013

Accepted, September 23, 2013