Prediction of Sulfur Dioxide (SO$_2$) Concentration Levels from the Mina Al-Fahal Refinery in Oman Using Artificial Neural Networks

[1]Sabah A. Abdul-Wahab, [2]Saleh M. Al-Alawi
[1]Mechanical and Industrial Engineering Department, Sultan Qaboos University,
P.O. Box 33, Postal code 123, Al-Khoud, Muscat, Oman
[2]Electrical and Computer Engineering Department, Sultan Qaboos University,
P.O. Box 33, Postal code 123, Al-Khoud, Muscat, Oman

Abstract: In this investigation, two Artificial Neural Network (ANN) models were applied for predicting ground-level sulfur dioxide (SO$_2$) in the Sultanate of Oman in order to provide an early warning advisory for the protection of public health. The objective of the first model (Model I) was to use ANN to predict sulfur dioxide (SO$_2$) levels at certain receptors from the Mina Al-Fahal refinery in Oman. The artificial neural network was also used for predicting the first 3 maximum SO$_2$ concentrations and their corresponding locations with respect to the refinery (Model II). The models were used to determine meteorological conditions that most affect SO$_2$ concentrations. In assessing this aspect, five meteorological parameters that are expected to affect the SO$_2$ concentrations were explored. They include wind speed, atmospheric stability class, wind direction, mixing height, and ambient temperature. The developed models showed good predictive success with, R-squared values above 0.96 indicating high accuracy for both the models development and generalization capability. The meteorological variables with the greatest influence on SO$_2$ concentrations were also identified. It was found that wind direction was the variable most important to Model I while wind direction, stability, and wind speed were the highest contributing variables in Model II. The investigation indicated that the ANN models were well-suited for modelling SO$_2$ levels. Additionally, the ANN models can be extended for other applications in which non-linear relationships are observed.

Key words: Air pollution, ANN, SO$_2$, models, refinery, Sultanate of Oman

INTRODUCTION

Sulfur dioxide (SO$_2$) is considered one of the indicators of air quality. It is formed primarily from the combustion of sulphur-containing fuels and can affect the health of people and have a negative impact on the environment. Modelling of environmental parameters is a basis for a better understanding and prevention of air pollution.

However, modelling SO$_2$ is a complex task which has drawn the attention of many scientists all over the world since the early 1960s. The literature on this subject showed that there was no universal approach for modelling SO$_2$ and, hence, several modelling approaches were proposed. One approach is to use the atmospheric diffusion model to predict future pollutant concentrations[1]. This approach simulates the airflow pattern and pollutant concentration by solving a highly coupled, non-linear, partial differential equation set. Such an approach demands huge computing costs which sometimes cause difficulties in computational convergence, especially for treating large space cases, and experimental validation, which is even more expensive and difficult to achieve due to the scaling inconsistency[2].

A second approach in facilitating the prediction of pollutant concentrations is by developing statistical models. In recent years, the general trend was to use more statistical methods instead of traditional deterministic modelling. Statistical models were used to determine the underlying relationship between a set of input variables and the targets[3-5]. These models assume that the relationship between the variables is statistical in nature. Such models require information about the distribution of data which is generally not known previously[6]. The open literature showed that statistical approaches are frequently considered for short-term forecasting applied to real-time control of

Corresponding Author: Sabah A. Abdul-Wahab, Mechanical and Industrial Engineering Department, Sultan Qaboos University, P.O. Box 33, Postal code 123, Al-Koud, Muscat, Oman

emissions or to air quality assessment[7,8]. These methods have some advantages over deterministic approaches. First, they do not need data about emissions as they are based only on the air quality and meteorological measurements. Second, they are often simpler in their structure than deterministic models, and they can more easily be implemented and used by non-experts. However, the statistical models are not portable from site to site since they are developed and calibrated on local data.

Recently, statistical models including those based on artificial neural networks (ANN) the third approach, were used to model pollutant concentrations with promising results[6,9-17]. ANNs can capture the link between the input data and the corresponding output data. By their unique structure, artificial neural networks possess the ability to learn non-linear relationships with limited prior knowledge about the process structure[18,19]. These models provide a better alternative than regression statistical models because of their computational efficiency. Additionally, the ANN methods can also be used in combination with traditional deterministic modelling techniques[20]. The capabilities and disadvantages of neural networks were described by Boznar and Mlakar[21] for the cases of wind forecasting, reconstruction, and SO_2 pollution forecasting.

Many researchers in the literature have used the ANN approach for tropospheric ozone forecasting. The most widespread ANN design is a Multilayer Perceptron (MLP) with a learning procedure based on the backpropagation algorithm. Abdul-Wahab and Al-Alawi used backpropagation MLP models to test the relationship between ozone and other variables. Their investigation indicated the potential of the MLP approach for capturing non-linear interactions. On the other hand, there are limited applications of ANN models for SO_2 prediction that are cited in the literature. Boznar et al.[9] applied a neural network model to predict SO_2 concentrations at Sostanj. The used inputs included historical data on observed SO_2 concentrations and meteorological parameters such as wind speed and direction. A feedforward network with a backpropagation learning algorithm was applied. The major finding was that ANN models can be applied to predict ground-level concentrations of pollutants in complex terrains. However, the results were not compared with those obtained using any other modelling techniques. For the same area, Mlakar and Boznar[22] used an artificial neural network as a method of short-term air pollution prediction. The method was applied to the complex orography around the coal-fired thermal power plant at Sostanj.

Gardner and Dorling[12] presented a review on the application of artificial neural networks in the atmospheric sciences. The applications of multilayer perceptions in atmospheric sciences were discussed in detail. Gardner and Dorling[12] concluded that ANN generally give as good or better results compared with the statistical linear method, especially where the problem being analyzed included non-linear behaviour[23].

Chelani et al.[24] predicted SO_2 concentration by using an ANN, and the predicted values were compared with the measured concentrations at three sites in Delhi. A multivariate regression model was also used for comparison with the results obtained by using the neural network model. The study results indicated that the neural network was able to give better predictions with less residual mean square error than those given by multivariate regression models.

Nunnari et al.[25] inter-compared several statistical techniques for modelling SO_2 concentration at a point by using neural networks, fuzzy logic, generalized additive techniques, and other recently proposed statistical approaches. They found that artificial neural network-based models and neuro-fuzzy models were the most promising. The results of the modelling inter-comparison was planned to be used to give guidelines for designing a warning system for air quality assessment.

The control of SO_2 and the associated prediction of its levels are needed to take the preventive actions during the episodes of high SO_2 concentrations. In this paper, an artificial neural network was used for predicting SO_2 levels at certain receptors from the Mina Al-Fahal refinery in the Sultanate of Oman (Model I). The particular aim was to relate the SO_2 concentrations to meteorological parameters. In this study, five meteorological parameters that are expected to affect the SO_2 concentrations were investigated. These include wind speed, atmospheric stability class, wind direction, mixing height, and ambient temperature. The artificial neural network was also used for predicting the first 3 maximum SO_2 concentrations and their corresponding locations with respect to the refinery (Model II).

MATERIALS AND METHODS

Area's description: The Mina Al-Fahal refinery in Oman is used as a case study in this paper. The refinery, constructed in 1982, is situated on a plot area of 170,000 m^2 located in the industrial hub of Mina Al-Fahal. The location of the site was selected based on its

proximity to both the natural resources supply and the main consumers. The refinery was originally built to process 50,000 BPD of Oman's crude oil to meet the domestic requirements of petroleum products. However, the capacity was enhanced to 80,000 BPD in 1987 through an extensive revamp to meet the growing demand of petroleum products in the country. The refinery complex, as a whole, contains the refinery itself, its tank farm, containing 24 vertical storage tanks, wastewater treatment plant, and gas turbine electricity generating sets. The refinery's offices and laboratory are located approximately 400 m to the east of the refinery. The location of the surrounding residential areas in the vicinity of the Mina Al-Fahal area is shown in Fig. 1. Residential areas to the west and south of the refinery are within 500 m of the site boundary and therefore could be at risk from atmospheric emissions. More details about the site can be found elsewhere[19].

The air emissions in the refinery are mainly due to the quantities of fuel gas burned which results in the emission of various pollutants to the atmosphere, including SO_2. The main sources for SO_2 generation in the study area included five gas turbine generators, two steam generators, six heaters, three reboilers and one flare. In all these systems, gaseous fuels were combusted and the resulting flue gases that contain various pollutants were released into the atmosphere through a total of 17 stacks. The composition of the fuel gas used in the combustion contained 78.9 % H_2, 3.7 % CH_4, 5.1 % C_2H_6, 7.9 % C_3H_8, 4.1 % C_4H_{10}, 0.1 % C_5H_{12}, and 0.2 % H_2S. In the combustion, air was used as the source of oxygen. When this fuel is burned, carbon in the fuel reacted to form CO and/or CO_2, hydrogen formed H_2O, and sulphur formed SO_2. At temperatures greater than approximately 1800 ºC, some of the nitrogen in the air reacted to form oxides of nitrogen. The emission rate of SO_2 that resulted from the combustion from each stack was calculated by using material balances. Detailed information about the stacks and their SO_2 emission rates can be found in Abdul-Wahab et al.[19].

Development of ANN models: It was decided that two ANN models would be developed for this work. Table 1 shows the input and output for each model. The first model assessed the factors affecting the SO_2 concentrations at specific selected receptors. The second ANN model was developed for predicting the first 3 maximum SO_2 concentrations and their corresponding locations.
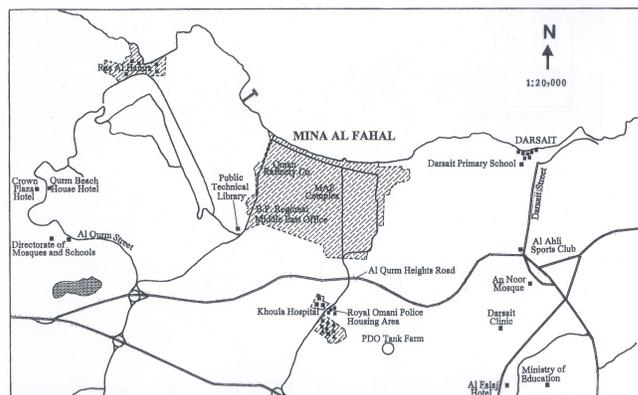

Fig. 1: Location of the surrounding residential areas in the vicinity of Mina Al-Fahal area

Table 1: Inputs and outputs for Model I and Model II

Model I	
Inputs	Outputs
Wind direction	SO_2 concentration
Wind speed	
Temperature	
Stability	
Mixing height	
X	
Y	

Model II	
Inputs	Outputs
Wind direction	X1
Wind speed	Y1
Temperature	Maximum1 of SO_2 concentration
Stability	X2
Mixing height	Y2
	Maximum2 of SO_2 concentration
	X3
	Y3
	Maximum3 of SO_2 concentration

The first and the most critical step in developing an effective ANN model is input and output definition and data preparation. This includes identifying variables of interest, gathering the relevant data and inspecting them for possible errors, missing values, and outliers; hence, the data for developing the two SO_2 models was inspected carefully since data accuracy is vital for the development of efficient models. The Feed Forward network using the Back-Propagation (BP) algorithm was used to develop these two models. The BP algorithm uses the supervised training technique.
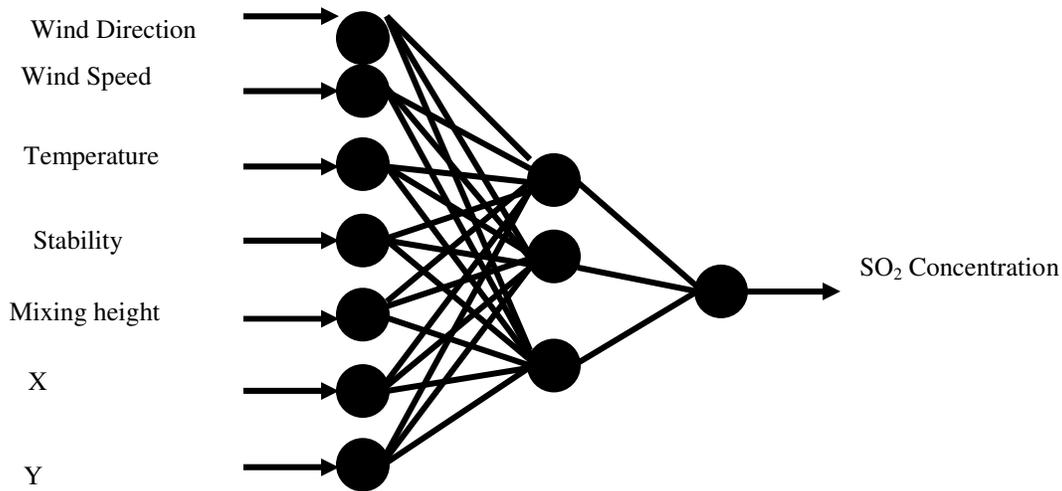
Fig. 2: Architecture of the developed ANN Model

In this technique, the interlayer connection weights and the processing elements' thresholds are first initialized to small random values. The network is then presented with a set of training patterns, each consisting of an example of the problem to be solved (the input) and the desired solution to this problem (the output).

These training patterns are presented repeatedly to the ANN model and the error between actual and predicted results is calculated. Weights are then adjusted by small amounts that are dictated by the General Delta Rule[26]. This adjustment is performed after each completed iteration whenever the network's computed output is different from the desired output. This process continues until weights converge to the desired error level or the output reaches an acceptable level.

RESULTS AND DISCUSSION

Prior to conducting the network's training operation using the back propagation paradigm, training sets of 9216 cases were obtained from the data set. This data set covers different situations that could possibly take place. For the purpose of model testing and validation, 15% were extracted from the given data set. Therefore, the training set consisted of 7834 cases while the testing set consisted of 1382 cases. As shown in Fig. 2, the ANN Model I used in this work consisted of 7 input nodes representing wind direction, wind

speed, temperature, stability, mixing height, X and Y. The output consisted of one node representing the SO_2 concentration. The training process was performed using the NeuroShell® simulator. After several adjustments to the network parameters, the network converged to a threshold of 0.00001 using 3 hidden nodes. The trained model prediction was in good agreement with the actual results. The R^2 value for the developed model was 0.9693 (Table 2). This indicates that approximately 96.9% of the variation in the SO_2 concentration could be explained by the selected input variables and the data used for model development.

Having trained the network successfully, the next step is to test the network in order to judge its performance and to determine whether the predicted results confirm with the actual results. The trained model is assumed to be successful if the model gives good results for the test set. Using the 1382 cases allocated for the testing set, the model-input parameters were entered consecutively for each case and a prediction for SO_2 concentration values was obtained. The results were then compared with the actual results for these cases. The statistical analysis of these results shown in Table 2 indicates that the R^2 value for the testing set was 0.9666, the mean squared error was 8.902 and the mean absolute error was 1.173. The ANN predicted and actual testing results are also shown in Fig. 3.
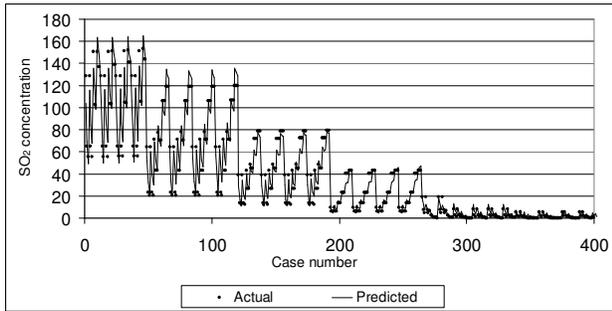
Fig. 3: Comparison of predicted and actual SO_2 concentrations for the validation data set for Model I

This high generalization capability indicates that the ANN model developed in this work can be used to model and predict the relationship between SO_2 concentration and the given input variables.

Model II, on the other hand, consists of 5 input nodes representing wind direction, wind speed, temperature, stability, and mixing height and 9 output nodes X1,Y1, Maximum 1 of SO_2 concentration, X2,Y2, Maximum 2 of SO_2 concentration, X3,Y3, Maximum 3 of SO_2 concentration, as illustrated in Fig. 4. The training set for ANN Model II consisted also of 7834 cases while the testing set consisted of 1382 cases. Similarly, the training process was performed using the NeuroShell® simulator. The trained model prediction was in good agreement with the actual results.

The R^2 value for the training and testing sets for the output variables X1,Y1, Maximum 1 of SO_2 concentration, X2,Y2, Maximum 2 of SO_2 concentration, X3,Y3, and Maximum 3 of SO_2 concentration is shown in Table 3. All R- squared values were above 0.996, indicating high accuracy for both model development and model generalization capability. Fig. 5 illustrates the comparison of predicted and actual maximum SO_2 concentration levels.

To find the percent contribution of each of the input variables with respect to the output variables, the partitioning method of the connection weights proposed by Garson[27] was used. The method involves partitioning the hidden-output connection weights of each hidden neuron into components associated with each input neuron[28,29]. It should be noted that there are several methods for calculating artificial neural networks' output contributions. Some of these methods are (1) Partial Derivatives or 'PaD'; (2) the "weights method or the partitioning method used in the current work; (3) the "perturb" method; (4) the "Profile"

method; (5) the "Classical Stepwise"; (6) the "improved stepwise a"; (7) the "improved stepwise b". A summarized review and comparison of methods to study the contribution of variables in artificial neural network models can be found in Gevrey et al.[30]. Gevrey et al.[30] presented a simplification of the partitioning algorithm as indicated below:

- For each hidden neuron *h*, divide the absolute value of the input-hidden layer connection weight by the sum of the absolute value of the input-hidden layer connection weight of all input neurons, i.e.

For *h*=1 to *nh*,

For *i*=1 to *ni*,

$$Q_{ih} = \frac{\left|W_{ih}\right|}{\sum_{i=1}^{ni}\left|W_{ih}\right|}$$

end,

end.

- For each input neuron *i*, divide the sum of the Q_{ih} for each hidden neuron by the sum for each hidden neuron of the sum for each input neuron of Q_{ih}, multiply by 100. The relative importance of all output weights attributable to the given input variable is then obtained.

$$RI(\%) = \frac{\sum_{h=1}^{nh}Q_{ih}}{\sum_{h=1}^{nh}\sum_{i=1}^{ni}Q_{ih}} X100$$

Using this method, for Model I, it was found that the highest contributing variable that affects the SO_2 concentration level is the wind direction; while in Model II, wind direction, stability and wind speed are the highest contributors. The relative importance of the different input variables in the prediction of SO_2 concentration levels are shown in Fig. 6 for Model I, and in Table 4 for Model II.

Table 2: Developed environmental model results for Model I

Description	Training Data Set	Testing Data Set	Total Data Set
Number of sets	7834	1382	9216
R-squared	0.9693	0.9666	0.9690
Mean squared error	11.971	8.902	11.511
Mean absolute error	1.288	1.173	1.270
Correlation coefficient, r	0.9853	0.9843	0.9852

Table 3: Developed environmental model results for Model II

Model output	R-squared	Mean Squared Error	Mean Absolute error	Correlation Coefficient r
		Training Data Set=7834		
X1	0.9958	4983.085	42.682	0.9980
Y1	0.9967	7125.474	53.708	0.9984
Maximum1 of SO_2	0.9969	18.993	3.434	0.9986
X2	0.9958	15304.464	71.703	0.9980
Y2	0.9943	26052.289	100.118	0.9972
Maximum2 of SO_2	0.9973	8.281	2.15	0.9987
X3	0.9941	26381.559	118.613	0.9971
Y3	0.9970	24776.770	114.227	0.9987
Maximum3 of SO_2	0.9931	10.617	1.877	0.9966
		Testing Data Set= 1382		
X1	0.9951	6067.152	44.666	0.9976
Y1	0.9957	8977.985	57.420	0.9979
Maximum1 of SO_2	0.9972	20.064	3.509	0.9987
X2	0.9963	14601.571	70.318	0.9982
Y2	0.9944	26019.018	100.655	0.9972
Maximum2 of SO_2	0.9977	7.958	2.153	0.9989
X3	0.9934	31469.248	124.393	0.9968
Y3	0.9971	24923.211	118.831	0.9987
Maximum3 of SO_2	0.9922	13.039	1.963	0.9962

Table 4: Percentage contribution of the various input variables on the outputs of Model II

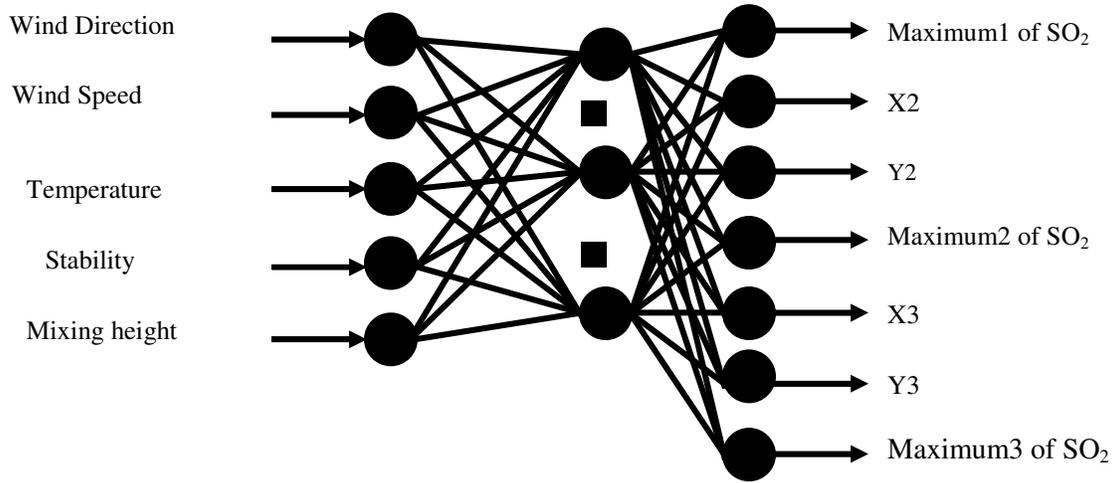| Input Parameters	Maximum1 of SO_2	Rank	Maximum2 of SO_2	Rank	Maximum3 of SO_2	Rank
Wind direction	35.82	1	31.50	1	28.69	2
Stability	28.41	2	30.56	2	31.93	1
Wind speed	25.99	3	27.19	3	26.86	3
Mixing height	8.91	4	9.19	4	10.58	4
Temperature	0.87	5	1.56	5	1.95	5

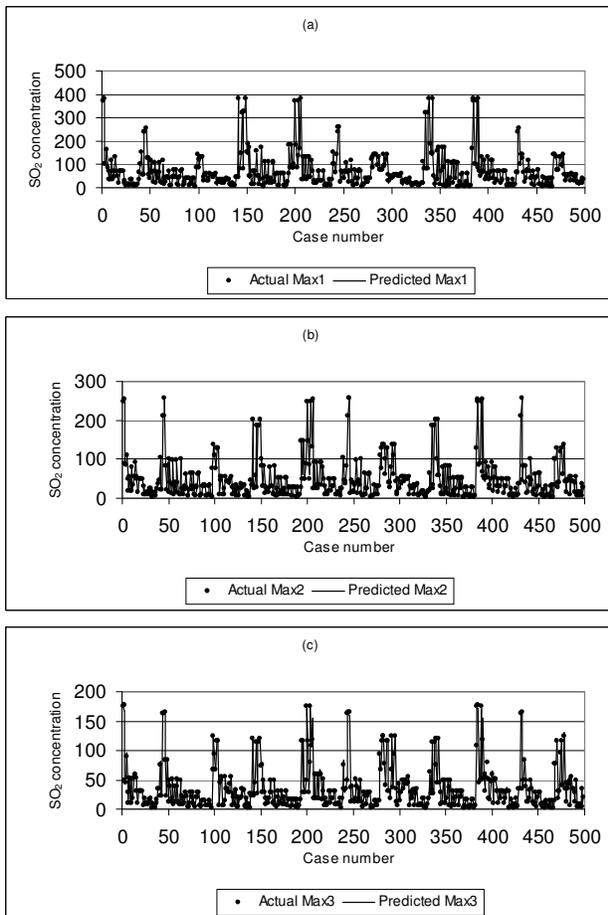Fig. 4: Architecture of the developed ANN Model II

Fig. 5: Comparison of predicted and actual maximum SO$_2$ concentrations for the testing data set for Model II
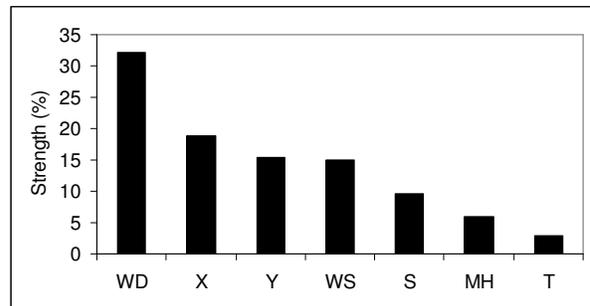

Fig. 6: The relative importance of the different inputs in the prediction of SO$_2$ in Model I

CONCLUSION

Interest in neural networks and other artificial intelligence techniques is on the increase due to their high potential. Intelligent systems make it possible to obtain expert knowledge and to discover new knowledge to improve data analysis and the decision making process, hence improving the quality of life.

This paper examines the potential of ANN models as a predictive tool for predicting SO$_2$ concentrations as a function of meteorological conditions. The objective was to use an ANN model to predict SO$_2$ levels at certain receptors in the Mina Al-Fahal refinery in the Sultanate of Oman (Model I). The results offer an insight into the dependence of SO$_2$ concentrations on meteorological parameters. Five meteorological parameters that are expected to affect

the SO_2 concentrations were investigated. These include wind speed, atmospheric stability class, wind direction, mixing height, and ambient temperature. The artificial neural network was also used for predicting the first 3 maximum SO_2 concentrations and their corresponding locations with respect to the refinery (Model II). The data fed to the neural network was divided into two sets: a training set and a testing set.

It was found that the neural network model consistently gives superior predictions and the results produced were encouraging. The R-squared values were found to be high. The R-squared for the first model was of the order of 97 % and they were of the order of 99.9 % for the second model. The relative importance of the various meteorological variables was also investigated. Clearly, this study indicated the potential of the neural network approach for capturing the non-linear interactions between SO_2 levels and meteorological variables and for the identification of the relative importance of these variables.

REFERENCES

1. Collet, R.S. and K. Oluyemi, 1997. Air quality modelling: a technical review of mathematical approaches. Meteorological Applications, 4: 235-246.
2. Lu, W.Z., W.J. Wang, X.K. Wang, S.H. Yan and J.C. Lam, 2004. Potential assessment of a neural network model with PCA/RBF approach for forecasting pollutant trends in Mong Kok urban air, Hong Kong. Environmental Research, 96: 79-87.
3. Finzi, G. and G. Tebaldi, 1982. A mathematical model for air pollution forecast and alarm in an urban area. Atmospheric Environment, 16(9): 2055-2059.
4. Ziomass, I.C., M. Dimitrios, C.Z. Christos and F.B. Alkiviadis, 1995. Forecasting peak pollutant levels from meteorological variables. Atmospheric Environment, 29 (24): 3703-3711.
5. Shi, J.P. and R.M. Harrison, 1997. Regression modelling of hourly NOx and NO2 concentrations in urban air in London. Atmospheric Environment, 31: 4081-4094.
6. Comrie, A.C., 1997. Comparing neural networks and regression models for ozone forecasting. Journal of the Air and Waste Management Association, 47: 653-663.
7. Zannetti, P., 1990. Air Pollution Modelling: Theories, Computational Methods and Available Software. Van Nostrand Reinhold, New York.
8. Zannetti, P., 1994. Computer modelling of air pollution, science, art or fiction: Air pollution II. In: Computer Simulation, vol. 1. Computational Publications, Southampton, Boston.
9. Boznar, M., M. Lesjak and P. Mlakar, 1993. A neural network-based method for short-term predictions of ambient SO_2 concentrations in highly polluted industrial areas of complex terrain. Atmospheric Environment, 27B: 221-230.
10. Broomhead, D. and D. Lowe, 1988. Multivariable functional interpolation and adaptive networks. Complex System, 2: 321-355.
11. Gardner, M.W. and S.R. Dorling, 1996. Neural network modelling of the influence of local meteorology on surface ozone concentrations. Proceedings of 1st International Conference on GeoComputation, University of Leeds, pp. 359-370.
12. Gardner, M.W. and S.R. Dorling, 1998. Artificial neural networks (The multilayer perceptron)- a review of applications in atmospheric sciences. Atmospheric Environment, 32: 2627-2636.
13. Hadjiiski, L. and P.K. Hopke, 2000. Application of artificial neural networks to the modelling and prediction of ambient ozone concentrations. Journal Air Waste Management Association, 50: 894-901.
14. Reich, S.L., D.R. Gomez and L.E. Dawidowski, 1999. Artificial neural network for the identification of unknown air pollution sources. Atmospheric Environment, 33: 3045-3052.
15. Roadknight, C.M., G.R. Balls and G.R. Mills, 1997. Modelling complex environmental data. IEEE Trans. Neural Networks, 8: 852-861.
16. Song, X.H. and P.K. Hopke, 1996. Solving the chemical mass balance problem using an artificial neural network. Environmental Science and Technology, 30: 531-535.
17. Yi, J. and R. Prybutok, 1996. A neural network model forecasting for prediction of daily maximum ozone concentration in an industrialized urban area. Environmental Pollution, 92 (3): 349-357.
18. Elkamel, A., S.A. Abdul-Wahab, W. Bouhamra and E. Alper, 2001. Measurement and prediction of ozone levels around a heavily industrialized area: A neural network approach. Advances in Environmental Research, 5: 47-59.
19. Abdul-Wahab, S.A. and S.M. Al-Alawi, 2002. Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. Environmental Modelling & Software, 17(3): 219-228.

20. Kukkonenl, J., L. Partanen, A. Karppinen, J. Ruuskanen, H. Junninen, M. Kolehmainen, H. Niska, S. Dorling, T. Chatterton, R. Foxall and G. Cawley, 2003. Extensive evaluation of neural network models for the prediction of NO2 and PM10 concentrations, compared with a deterministic modelling system and measurements in central Helsinki. Atmospheric Environment, 37: 4539-4550.

21. Boznar, M., and P. Mlakar, 1995. Neural networks- a new mathematical tool for air pollution modelling. International conference on air pollution, Volume 1, pp 259-266.

22. Mlakar, P. and M. Boznar, 1994. Short-term air pollution prediction on the basis of artificial neural networks. International conference on air pollution, Volume 1, pp 545-552.

23. Kolehmainen, M., H. Martikainen and J. Ruuskanen, 2001. A Neural networks and periodic components used in air quality forecasting. Atmospheric Environment, 35: 815-825.

24. Chelani, A.B., C.V. Raoi, K.M. Phadke and M.Z. Hasan, 2002. Prediction of sulphur dioxide concentration using artificial neural networks. Environmental Modelling & Software, 17: 161-168.

25. Nunnari, G., S. Dorling, U. Schlink, G. Cawley, R. Foxall and T. Chatterton, 2004. Modelling SO_2 concentration at a point with statistical approaches. Environmental Modelling & Software, 19: 887-905.

26. Rumelhart, D. E. and J. L. McClelland, 1986. Parallel distribution processing: exploration in the microstructure of cognition. Vol.1, Foundations, MIT Press, Cambridge, MA.

27. Garson, G. D., 1991. Interpreting neural-network connection weights. AI Expert, 6(7): 47-51.

28. Goh, A. T. C., 1995. Back-propagation neural networks for modeling complex systems. Artificial Intelligence in Engineering, 9: 143-151.

29. Yoon, Y., T. Guimaraes and G. Swales, 1994. Integrating artificial neural networks with rule-based expert systems. Decision Support Systems, 11(5): 497-507.

30. Gevrey, M., I. Dimopoulos and S. Lek, 2003. Review and comparison of methods to study the contribution of variables in artificial neural network models. Ecological Modelling, 160:249-264.