ELSEVIER

Environmental Modelling & Software 17 (2002) 219–228

Environmental Modelling & Software

www.elsevier.com/locate/envsoft

Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks

S.A. Abdul-Wahab [a,*], S.M. Al-Alawi [b]

[a] *Mechanical and Industrial Engineering Department, Sultan Qaboos University, PO Box 33, Al Khoud, Muscat, Oman*

[b] *Electrical and Electronics Engineering Department, Sultan Qaboos University, PO Box 33, Al Khoud, Muscat, Oman*

Received 6 February 2001; received in revised form 14 June 2001; accepted 7 September 2001

Abstract

This work deals specifically with the use of a neural network for ozone modelling in the lower atmosphere. The development of a neural network model is presented to predict the tropospheric (surface or ground) ozone concentrations as a function of meteorological conditions and various air quality parameters. The development of the model was based on the realization that the prediction of ozone from a theoretical basis (i.e. detailed atmospheric diffusion model) is difficult. In contrast, neural networks are useful for modelling because of their ability to be trained using historical data and because of their capability for modelling highly non-linear relationships. The network was trained using summer meteorological and air quality data when the ozone concentrations are the highest. The data were collected from an urban atmosphere. The site was selected to represent a typical residential area with high traffic influences. Three neural network models were developed. The main emphasis of the first model has been placed on studying the factors that control the ozone concentrations during a 24-hour period (daylight and night hours were included). The second model was developed to study the factors that regulate the ozone concentrations during daylight hours at which higher concentrations of ozone were recorded. The third model was developed to predict daily maximum ozone levels. The predictions of the models were found to be consistent with observations. A partitioning method of the connection weights of the network was used to study the relative percent contribution of each of the input variables. The contribution of meteorology on the ozone concentration variation was found to fall within the range 33.15–40.64%. It was also found that nitrogen oxide, sulfur dioxide, relative humidity, non-methane hydrocarbon and nitrogen dioxide have the most effect on the predicted ozone concentrations. In addition, temperature played an important role while solar radiation had a lower effect than expected. The results of this study indicate that the artificial neural network (ANN) is a promising method for air pollution modelling.

Keywords: Ozone; Artificial neural networks; Kuwait

1. Introduction

Ozone (tropospheric ozone) can have a negative impact on the environment and public health when present in the lower atmosphere in sufficient quantities. In establishing ambient air quality standards, regulations have been introduced to set limits on the emissions of pollutants in such a way that they cannot exceed prescribed maximum values (Maynard, 1984; EPA, 1999). To achieve these limits, consideration was given to mathematical and computer modelling of air pollution. Ozone, however, is unique among pollutants because it is not emitted directly into the air. It is a secondary pollutant that results from complex chemical reactions in the atmosphere. It results when the primary pollutants nitrogen oxides (NO_x) and non-methane hydrocarbons (NMHC) interact under the action of sunlight. Therefore, the primary pollutants NO_x and NMHC are referred to as ozone precursors. There are thousands of sources of NMHC and NO_x. To track and predict ozone, one must create an understanding of not only ozone itself but also the conditions that contribute to its formation. In addition, ozone concentrations are strongly linked to meteorological conditions. Land–sea breezes also influence ozone concentrations at coastal sites. To predict ozone concentrations, it is necessary to apply a model

* Corresponding author. Tel.: +968-515-350; fax: +968-513-416.
E-mail address: sabah1@squ.edu.om (S.A. Abdul-Wahab).

1364-8152/02/$ - see front matter
PII: S1364-8152(01)00077-9

that describes and understands the complex relationships between ozone concentrations and the many variables that cause or hinder ozone production.

From 1978 to 1997, forecasts were based on the one-hour National Ambient Air Quality Standard (NAAQS) for ozone, which was 0.12 parts per million (ppm) (EPA, 1999). In 1997, the US Environmental Protection Agency (EPA) revised the NAAQS to reflect more recent health-effects studies that suggest that respiratory damage can occur at lower ozone concentrations. Under the revised standard, regions exceed the NAAQS when the three-year average of the annual fourth highest eight-hour average ozone concentration is above 0.08 ppm. More regions will have daily maximum eight-hour ozone concentrations that exceed the level of the revised NAAQS than the old standard, and more agencies may need to model ozone (EPA, 1999). Accordingly, both deterministic and statistical models have been developed to better understand ozone production (Topcu et al., 1993).

Deterministic models (i.e. theoretical or detailed atmospheric diffusion models) are based on a fundamental mathematical description of atmospheric processes in which effects are generated by causes (Zannetti 1983, 1994). Such models aim to resolve the underlying chemical and physical equations that control pollutant concentrations and therefore require detailed emission data and meteorological conditions for the region of interest. An excellent example is the urban airshed model (UAM) (Zannetti 1983, 1994; Johnson, 1991). This model can be used to obtain an accurate picture of the factors involved in ozone production. However, the model is highly sophisticated because it requires a high level of human resources and intense data input (Johnson, 1991; Azzi et al., 1995). There are generally severe limitations in both spatial and temporal accuracy of the data. In addition, some input data are not easily acquired by environmental protection agencies or local industries. This means that if these inputs are unknown, then the application of the UAM is problematic. Therefore, it is much more practical to rely on statistical models.

Statistical models are based on semi-empirical statistical relations among available data and measurements. They do not necessarily reveal any relation between cause and effect. They attempt to determine the underlying relationship between sets of input data (predictors) and targets (predictands). Examples of statistical models are correlation analysis (Abdul-Wahab et al., 1996) and time series analysis (Hsu, 1992). However, the complex and sometimes non-linear relationships of multiple variables can make statistical models awkward and complicated (Comrie, 1997). Therefore, it is expected that they will under-perform when used to model the relationship between ozone and the other variables that are extremely non-linear.

Other statistical approaches frequently used include several artificial neural network implementations (Boznar et al., 1993; Ruiz-Suárez et al., 1995; Elkamel et al., 2001). The use of these artificial intelligence-based networks has been shown to give acceptable results for atmospheric pollution forecasting of pollutants such as SO_2, ozone and benzopyrene. Ozone in the lower atmosphere is a complex non-linear process. Therefore, the neural network is a well-suited method for modelling this process since it allows for non-linear relationships between variables. Neural networks, by their unique structure, possess the ability to learn non-linear relationships with limited prior knowledge about the process structure. They are therefore useful for evaluating the ozone problem at a particular location. In this paper, neural network modelling was used to predict ozone concentration levels.

2. Artificial neural network concepts

Artificial neural network (ANN) models are computer programs that are designed to emulate human information processing capabilities such as knowledge processing, speech, prediction, classifications, and control. The ability of ANN systems to spontaneously learn from examples, “reason” over inexact and fuzzy data, and to provide adequate and rapid responses to new information not previously stored in memory has generated increasing acceptance for this technology in various engineering fields and, when applied, has demonstrated remarkable success (Simpson, 1990; Elkamel et al., 2001).

The major building block for any ANN architecture is the processing element or neuron. These neurons are located in one of three types of layers: the input layer, the hidden layer, or the output layer (Fig. 1). The input neurons receive data from the outside environment, the hidden neurons receive signals from all of the neurons in the preceding layer, and the output neurons send information back to the external environment. These neurons are connected together by a line of communication called connection. Stanley (1990) indicated that the way in which the neurons are connected to each other in a network typology has a great effect on the operation and performance of the network. ANN models come in a variety of typologies or paradigms. Simpson (1990) provides a coherent description of 27 different popular ANN paradigms and presents comparative analyses, applications, and implementations of these paradigms. Of these, the most frequently used is the backpropagation paradigm (Rumelhart and McClelland, 1986). Detailed descriptions on the use of ANNs in environmental modelling can be found in Maier and Dandy (2000).

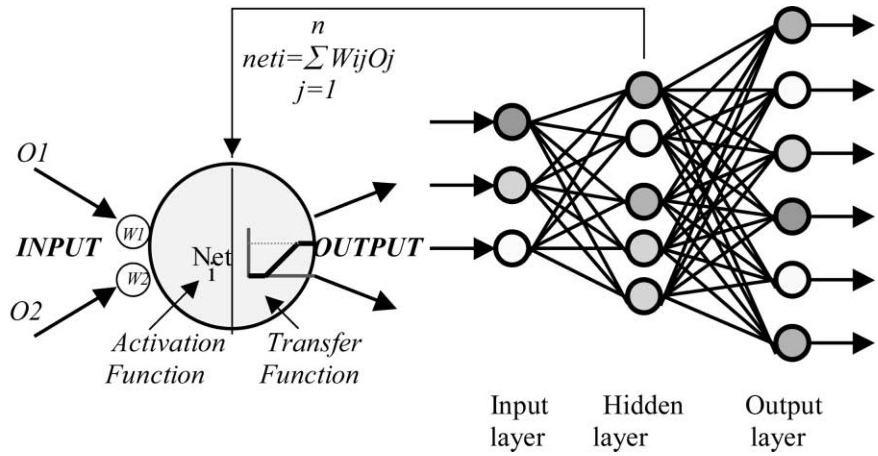


Fig. 1. A typical feedforward architecture.

3. Artificial neural networks and ozone modelling

Neural network models have the potential to describe highly non-linear relationships such as those controlling ozone production. Therefore, the application of artificial neural networks to ozone modelling has recently become available to capture those non-linear features of the relationship that a conventional statistical technique (e.g. regression models) might overlook (Comrie, 1997). Although they are relatively new and not yet widely used for this purpose, neural network models have proved to be a useful and cost-effective means for studying the relationship between ozone and other variables.

Gardner (1996) used the neural network to investigate the importance of local meteorology in determining the surface ozone concentration on an hourly basis. The purely meteorological input data were taken from Weybourne, a coastal site in north Norfolk. The data included hourly observations of temperature, humidity, irradiance, wind speed, direction and ozone concentrations for an entire year. Interestingly, Gardner's model did not involve any chemical data as input to the model. The model showed that over a period of a year, 48% of the ozone variation can be attributed to changing meteorological conditions. Any remaining variability was attributed to other causes such as chemical interaction between hydrocarbons and oxides of nitrogen.

Crowe and DeFries (1996) applied neural networks to predict ozone concentrations in southeast Texas, near Houston. The input data consisted of hourly meteorological parameters, nitrogen oxides and seven hydrocarbon species based on carbon bond four chemistry. Three neural network models were developed. The predictor variables for the first model consisted of five meteorological parameters for the same hour as the ozone measurement and also for six time delays to account for possible effects of transport and chemical reactions. The second model consisted of the same meteorological variables but included the species NO and NO_x. The third model dropped the time-lagged variables but added seven hydrocarbon species based on carbon bond four chemistry. The models showed progressively better predictive capability as evidenced by increasing R^2-values from 0.7 (first model) to 0.8 (second model) to 0.91 (third model). The authors reported that selected hydrocarbon species are more sensitive predictors of hourly ozone. They found that increasing olefins were associated with decreasing ozone and that increasing paraffin concentrations were associated with sharply increasing ozone levels.

Capone (1996) applied neural network technology to predict downwind hourly ozone data in the Baton Rouge area in the USA by using a more complicated network in which data from two downwind sites were employed as predictors. The model, which consisted of hourly meteorological and NO_x measurements at each site, was successful at predicting hourly ozone patterns. The Capone model did not involve any hydrocarbon species.

A study by the University of Arizona (Comrie, 1997) used data from eight cities around the United States to compare regression models and neural networks under a variety of climate and ozone regimes. The ozone data used were the daily maximum one-hour concentrations for the months of May through September over a five-year period. A comparison between the two methods indicated that neural network techniques are somewhat (but not dramatically) better than regression models for daily ozone prediction.

Elkamel et al. (2001) illustrated the successful use of a neural network to predict ozone concentrations using both meteorological and chemical data. The network was trained using data collected near an industrial area in Kuwait for a period of 60 days. The performance of the neural network model was compared against linear and non-linear regression models. The study indicated that neural network models consistently gave better predictions.

Hence, neural network models have the potential to describe highly non-linear relationships such as those controlling ozone production. Therefore, the application of artificial neural networks to ozone modelling is recently becoming available to capture those non-linear features of the relationship that a conventional statistical technique might overlook. The work reported in this paper deals with the use of the neural network as a method for ozone modelling and for predicting ozone concentrations as a function of meteorological conditions and other primary pollutants. The study focused specifically on identification of the factors that regulate the ozone levels during daylight hours and during periods of high ozone concentration. We examined the relative percent contribution of each input variable. Furthermore, the percentage of ozone variation due to meteorological factors and other pollutants was also investigated. The purpose of this work was to come up with intelligent evaluation of the ozone situation without many available data. This will help to provide the decision makers with a rapid tool to make an initial judgement of environmental situations using the limited data available.

4. Materials and methods

4.1. Area description

The state of Kuwait covers an area of approximately 17,818 km^2. It is situated at the head of the Arabian Gulf between latitudes 28° and 30° north and between longitudes 46° and 48° east (Fig. 2). Iraq lies towards the northern and western boundaries of Kuwait, Saudi Arabia lies to the south, while the Arabian Gulf marks the eastern boundary. The terrain is a flat to slightly undulating desert plain. Much of the country is desert. Thus the climate is typically arid with very hot summers and relatively cold and dry winters. The summer season in Kuwait falls between May and September and the winter season between November and March. Summer temperatures may exceed 50 °C, and in January, the coldest month, temperatures range between −2.8 and 28.3 °C. Kuwait has very little rainfall, most of it occurring as light winter showers brought by westerly depressions, especially in January (Gulf Union Company, 1993). The annual rainfall varies from 10 to 370 mm. Dust and sandstorms are common throughout the year. They are more frequent in the winter months and in midsummer (Gulf Union Company, 1993).

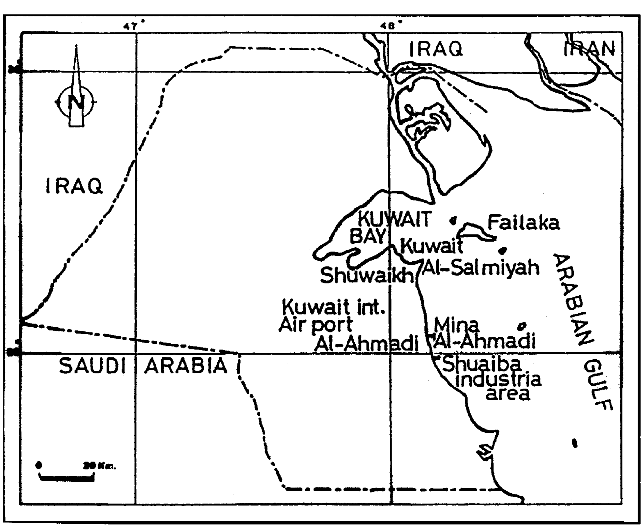


Fig. 2. Map of Kuwait State.

Relatively heavy traffic movement surrounds the area of study at Khaldiya residential area (Figs. 3 and 4) and therefore it is mainly affected by the pollutants that are discharged from the traffic load in view of the proximity of major highways. The wind rose over a one-year period (1997) is shown in Fig. 3. Most of the prevailing wind is from the west and the northwest. The monitoring site was situated downwind from the Shuwaikh industrial area and the Shuwaikh power plant in case the levels of pollution released from them were significant.

4.2. Data collection

Kuwait University's mobile air pollution laboratory was used in the study. The location of the mobile laboratory at Khaldiya was selected as the sampling site on the basis of the availability of power and security and the topography of the area. Care was taken that no high buildings or trees were present within 500 metres of the site. The mobile laboratory was fitted with chemical monitors and meteorological sensors. All the sensors were operated automatically. Measurements were

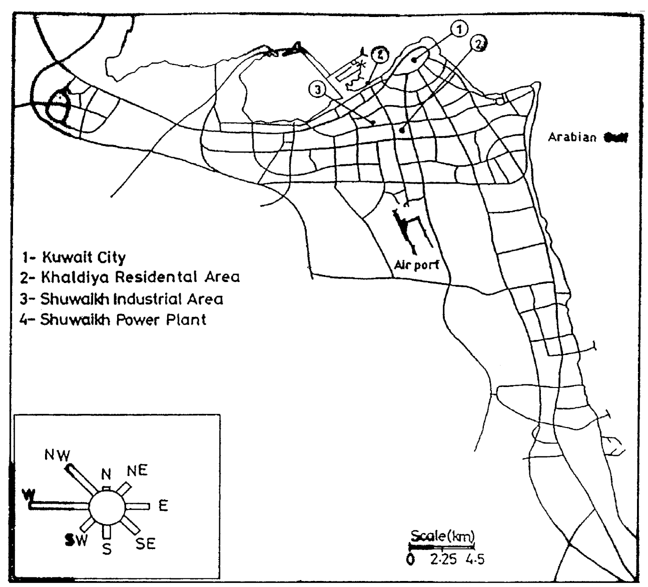


Fig. 3. Location of Khaldiya residential area in relation to Kuwait City and other areas of Kuwait. A wind rose is superimposed on the figure.

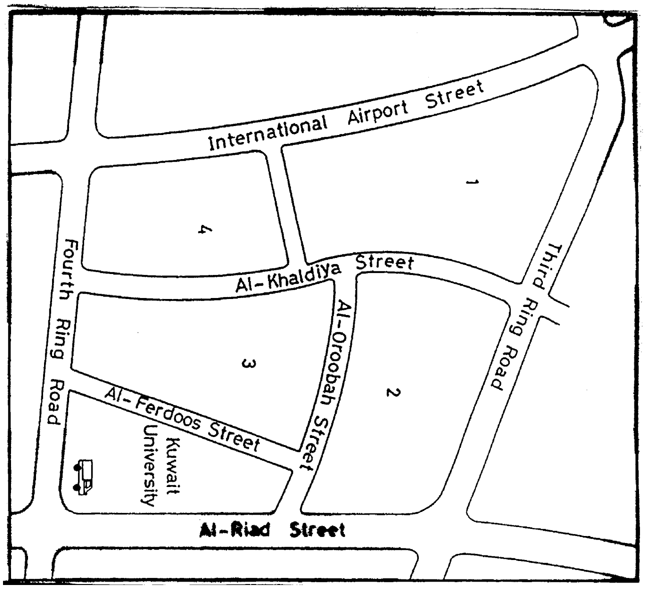


Fig. 4. The site of the mobile air pollution monitoring laboratory and structures in the immediate vicinity of Khaldiya residential area.

recorded every five minutes. Pollutants measured include CH_4, NMHC, CO, CO_2, NO, NO_2, SO_2 and O_3. Meteorological parameters monitored simultaneously included wind speed and direction, air temperature, relative humidity and solar radiation.

Methane and non-methane hydrocarbons were measured by gas chromatography using a flame ionization detector (Model MAS-1030A, Mine Safety Appliances Company) which had a detection limit of 0.05 ppm. Carbon monoxide and carbon dioxide concentrations were measured based on non-dispersive infrared absorption (Model 48 and Model 41/41H of Thermo Environmental Instruments, respectively). The detection limits for carbon monoxide and carbon dioxide were 0.1 ppmv and 5 ppb, respectively. The NO_x concentrations were measured with a detection limit of 0.5 ppb (Thermo Environmental Instrument, Model 42). SO_2 concentrations were measured by using Model 43A with a detection limit of 1 ppb (Thermo Environmental Instruments, Pulsed fluorescent). The ozone concentrations were measured by using a non-dispersive UV photometer (Monitors Labs, Model ML 9812) with a detection limit of 1.0 ppbv. Suspended dust was measured gravimetrically (TEOM Series 1400a). This was a real-time device used for assessing particulate concentration for sizes smaller than 10 μm in diameter. It was a filter-based mass monitor which was composed of a TEOM sensor and a TEOM control unit. It had a detection limit of 5 μg/m^3. Details of the mobile laboratory's meteorological sensors were given in previous studies (Abdul-Wahab et al. 1996, 2000; Bouhamra and Abdul-Wahab, 1999; Elkamel et al., 2001).

In terms of its operation, the mobile laboratory is characterized by the following: sampling inlets were located on top of the laboratory 10 metres above the ground; all the monitors were controlled by an intelligent data logger; automatic zero and span calibrations were performed using a calibration gas once every 23 hours; the Envicom software was used to record the data and then the Envaid software was used for editing and processing. A quality check was performed by examining the data in graphical form.

5. Results and discussion

A detailed analysis of ozone real-time monitoring data collected by the mobile laboratory indicated that the Khaldiya residential area was occasionally subjected to ambient ozone concentrations exceeding the NAAQS of 80 ppb. The results confirmed that high ozone events occur mainly in summer, which is in line with the results of other investigators (Salop et al., 1983; Bower et al., 1989; Poulid et al., 1991; Varshney and Aggarwal, 1992; Lorenzini et al., 1994).

On the basis of these findings, only the air quality data that were collected in June (i.e. summer) was selected for neural network evaluation. Table 1 shows the variables measured by the mobile laboratory and their minimum, maximum and mean values recorded. It can be seen that the monthly mean ozone concentration in June was 22.66 ppb. The level ranged from 0.0 to 108.5 ppb. The maximum ozone concentration observed during the study period was 108.5 ppb.

It was decided that three ANN models would be developed for this work. Thirteen variables were selected as inputs; CH_4, NMHC, CO, CO_2, NO, NO_2, SO_2, temperature (TEMP), relative humidity (RH), wind speed (WS), wind direction (WD), solar radiation (SOLAR) and dust. The selected output for these models was the ozone (O_3) concentration. Hence, for the development of the ANN architecture of these models, 13 neurons were used for input and one neuron for the output as

Table 1
Variables measured in the Khaldiya residential area during June 1997

Variable	Minimum	Maximum	Average
CH_4 (ppm)	1.6	3.12	1.74
NMHC (ppm)	0.04	6.19	0.619
CO (ppm)	0.0	18.6	2.76
CO_2 (ppm)	344	475	362.99
NO (ppb)	4	921	105.29
NO_2 (ppb)	1	154	44.44
SO_2 (ppb)	1	290	12.296
O_3 (ppb)	0	108.5	22.66
Wind speed (m/s)	0.4382	5.169	2.097
Temperature (°C)	28.85	48.05	38.03
Relative humidity (%)	7.8	69.1	21.61
Solar energy (kW/m^2)	0.018	0.9657	0.304
Suspended dust (μg/m^3)	0	1293	99.83

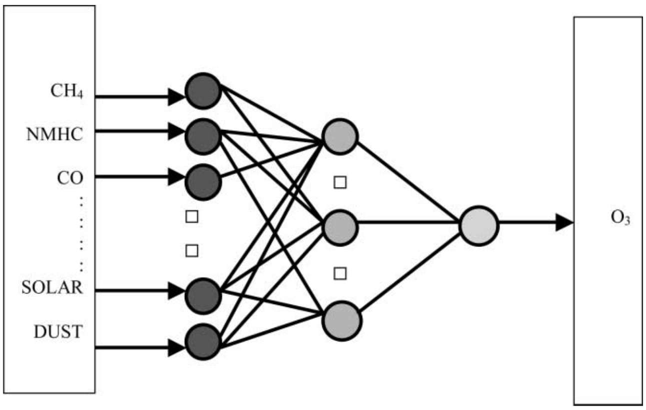


Fig. 5. The ANN architecture for the environmental models.

shown in Fig. 5. Data for the above variables were then examined for error outliers and missing values using the NeuroShell software utilities, and these records were removed from the data set.

The first model assessed the factors affecting the ozone concentration during a 24-hour period. Due to equipment calibration or maintenance, one or more of the variables may not have been measured at a given time. In other words, the matrix of data collected has missing entries. In such cases, the entire row was eliminated so that rows containing partial data were not considered in the data analysis. The number of complete data points recorded was 4797 records.

The diurnal profile of ozone given in Fig. 6 shows that the mean level of ozone was quite high during the daylight hours and quite low in the absence of sun. With this in mind, the second ANN model was developed in such a way that it focused only on the observations recorded during daylight hours at which higher concentrations of ozone were recorded. The number of complete data points included in this model was 1630 records.

Regulations have been introduced to set limits on the ozone concentrations in such a way that they cannot exceed the NAAQS of 80 ppb. Therefore, the third ANN model was developed for predicting daily maximum

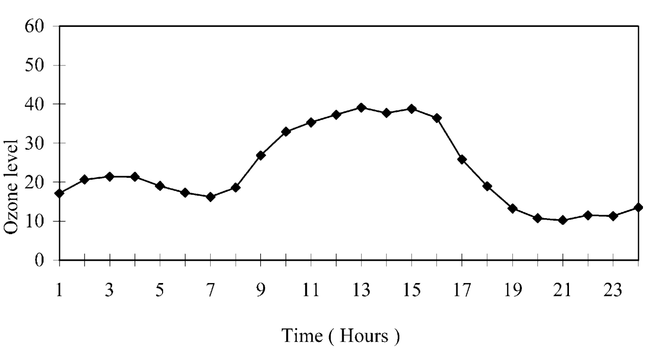


Fig. 6. Mean distribution of ozone concentration (ppb) in Khaldiya residential area as a function of time of day.

ozone levels. The prime objective of this model was to determine the conditions that favour the production of high ozone concentrations. The number of total records included in this model was 19. However, the limitations of the limited number of data should be also recognized.

5.1. Network development

Prior to conducting the network training operation for model I using the backpropagation paradigm (BP), a training set consisting of 4797 cases was obtained from the preprocessed environmental data. They were then divided into two distinct sets, a training set and a testing set, using a random process. The training set consisted of 85% of the data sets collected, or 4078 cases, that were then used for training the model. The remaining 15%, or 719 cases, were excluded from the training set and used later, once the model was developed, to test the model's performance. Typical examples of the different training patterns used as part of the training data set are shown in Table 2. The training process of this ANN model was performed using a NeuroShell™ simulator developed by Ward Systems Group. The simulator iteratively adjusts the weights until the error between the output data and the actual data (observed) is minimized. Several network factors such as momentum, learning rate, number of hidden nodes, and thresholds were tried during the training process to improve network generalization and prediction accuracy. The best results were obtained (Table 3) when the network completed 13,427 epochs in 9:30:01 hours of continuous training (i.e. the error is minimized). As shown by Fig. 7, there was a close fit between actual values and the model's prediction. The R^2-value for the trained model was 0.9812. This result indicates that approximately 98% of the variability in the ozone concentration levels (the dependent variable) could be explained by the selected independent variables and the data used for model development. Having trained the network successfully, the next step was to test the network to judge its performance and to determine how well the predicted results agree with the observed results.

The same procedures used in model I was then used in the development of models II and III. The training and testing sets, however, consisted of 1386 cases (85%) and 244 cases (15%) for model II, and 16 cases (85%) and 3 cases (15%) for model III, respectively. The best results for model II were obtained when the network completed 19,384 epochs in 8:30:05 hours of continuous training, while for model III, the best result was obtained in 1:22:33 hours after completing 1,054,327 epochs (Table 3). The R^2-values for models II and III were 0.9622 and 0.9067, respectively, as shown in Fig. 7.

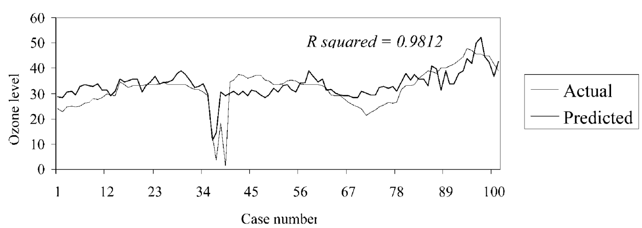
Table 2
Typical training patterns used for model development

Input													Output
CH_4 (ppm)	NMHC (ppm)	CO (ppm)	CO_2 (ppm)	NO (ppb)	NO_2 (ppb)	SO_2 (ppb)	WS (m/s)	WD (deg)	TEMP (°C)	RH (%)	SOLAR (kW/m^2)	DUST (μg/m^3)	O_3 (ppb)
1.64	0.2	0.75	352	9	15.5	6	1.39	214	41.95	13.5	0.338	90	42
1.76	0.83	2.03	363	95	77.5	16	0.97	312.1	41.86	13.1	0.651	142.5	14.5
1.66	0.51	1.96	359	78	51	9	3.53	97.2	44.29	14.2	0.493	67.5	11
1.66	0.12	0.56	357	9.5	23.5	3	2.16	279.5	41.68	13.8	0.865	40	61
1.69	0.09	1.31	359	12	33	7	2.27	302.6	41.81	16.4	0.1391	132.5	48.5
1.62	0.33	1.74	361	19	29.5	5	1.28	281.2	43.23	13.3	0.8561	507.5	38
1.67	0.44	1.83	358	66	64.5	6	2.88	88.2	42.79	14.5	0.7761	47.5	29

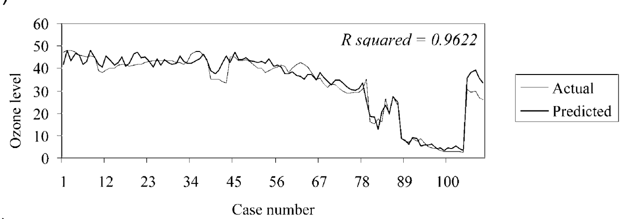
Table 3
Developed environmental model results

Description	Model I	Model II	Model III
Number of training sets	4078	1386	16
Number of testing sets	719	244	3
Elapsed time for model development	9:30:01	8:30:05	1:22:33
Number of epochs	13427	19384	1054327
R-squared	0.9812	0.9622	0.9067
Mean squared error	6.619	14.030	20.998
Mean absolute error	1.829	2.66	2.049
Correlation coefficient, *r*	0.9906	0.9811	0.9563

(a)



(b)



(c)

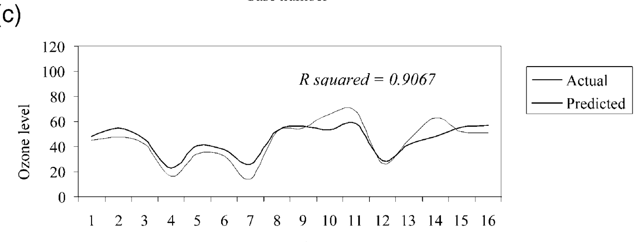


Fig. 7. Example of the results for the training for ozone (ppb) together with their actual values in the three trained models.

5.2. Model testing and validation

To validate the model's prediction capability, the test set data for each model were used to test the developed models. The resulting predictions were then compared with actual results, and statistical numerical measures were then calculated. The R^2-values for the test set in models I, II, and III were 0.9366, 0.9617, and 0.9321, respectively. The results (Fig. 8) show that the neural network model performed extremely well and therefore it was able to produce reasonably accurate predictions.

6. Interpreting variable importance

Neural network modelling can also assess the importance of each of the input variables by using the network weights. With this in mind, the method proposed by Garson (1991) for partitioning the connection weights was used. The technique involves partitioning the hidden–output connection weights of each neuron into components associated with each input neuron (Goh, 1995). The results of the calculations are shown in Fig. 9 and Table 4. The results shown in Fig. 9 are displayed as columns representing the relative importance of the various input variables while Table 4 displays the ranking of these variables.

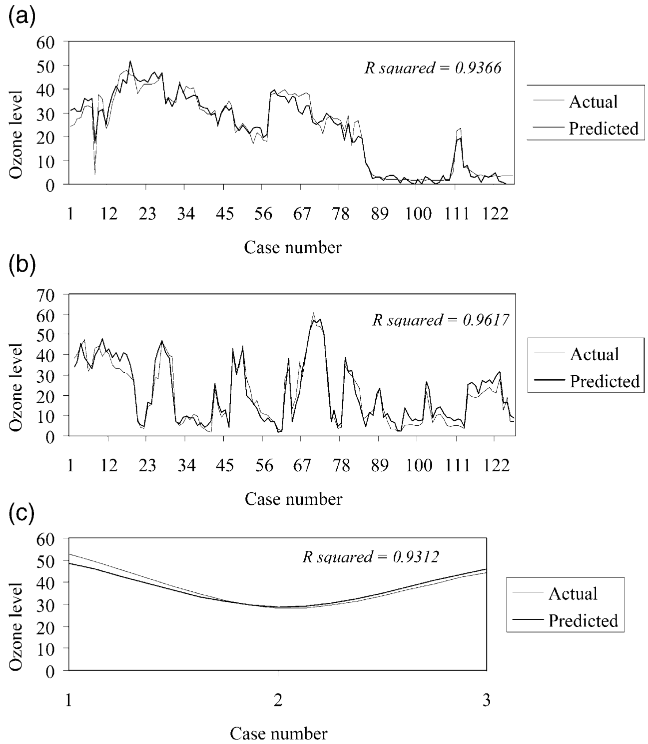


Fig. 8. Example of the results for the testing for ozone (ppb) together with their actual values in the three trained models.

It can be seen for the first model that NO, SO_2, relative humidity, NMHC and NO_2 make major contributions. The appearance of NO, NMHC and NO_2 is expected since they have been identified by many investigators as being key chemical precursors that produce ozone in the presence of sunlight. Moreover, controlled studies using environmental smog chambers have shown that ozone formation is dependent on the concentrations of its precursors as well as light intensity. Of considerable interest is the fact that solar energy appears to have a lower contribution than expected. This was found with all three models.

The significance of SO_2 on ozone concentration was also highlighted. The results of model I indicate the dependence of ozone on SO_2. However, this dependency decreased when focusing only on the results from daylight hours (models II and III). It should be noted that the importance of SO_2 and its share in the variations of ozone levels was expected because recent studies showed that they were physically and chemically coupled. It has been reported that in the presence of H_2O_2 and wet aerosols, SO_2 does participate in the chemistry of ozone. However, during day hours, SO_2 becomes a non-reactive pollutant and it participates very little in the chemistry of urban ozone (Ruiz-Suárez et al., 1995). Therefore, many studies now are directed towards modelling the impact of reduced SO_2 emissions on photochemical ozone production.

Looking at models I and II (Table 4), it can also be seen that the meteorological parameters with a high correlation to ozone concentrations include relative humidity and temperature. The relationship between ozone concentration and temperature can be explained on theoretical grounds. Temperature plays an enhancing role in the propagation rate of the radical chain, and has an opposite effect on the termination rate of these chains (Ruiz-Suárez et al., 1995). Relative humidity is also important because this variable may play a role in the overall reactivity of the system, either by affecting chain termination reactions or the production of wet aerosols which in turn affect the ultraviolet actinic flux. Looking at model III, it can be seen that temperature is one of the most important meteorological factors influencing the variation in ozone levels. The results of model III offer insights into the dependence of ozone on wind

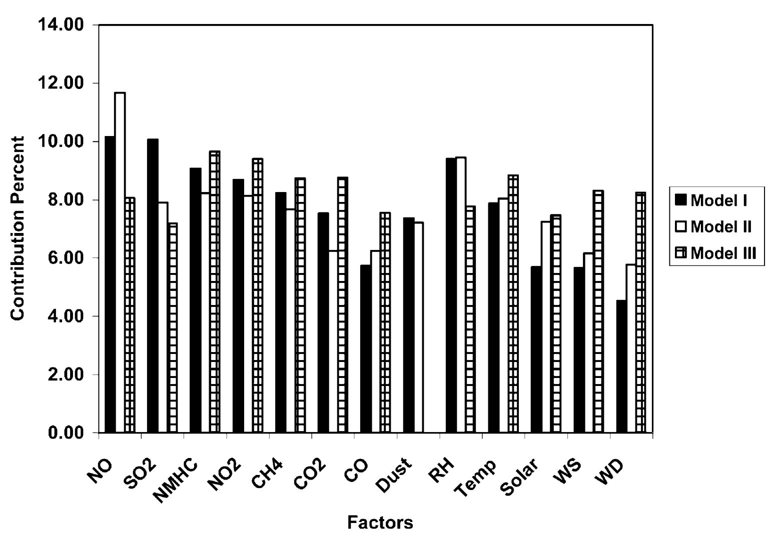


Fig. 9. The relative importance of the various input variables in the overall prediction of ozone to the ANN models.

Table 4
Percentage contribution of the various input variables to the ANN models

Factor	Model I	Rank	Model II	Rank	Model III	Rank
NO (ppb)	10.20	1	11.67	1	8.06	8
SO_2 (ppb)	10.10	2	7.9	6	7.2	12
NMHC (ppm)	9.10	4	8.23	3	9.66	1
NO_2 (ppb)	8.70	5	8.12	4	9.44	2
CH_4 (ppm)	8.20	6	7.66	7	8.73	5
CO_2 (ppm)	7.50	8	6.24	11	8.76	4
CO (ppm)	5.70	10	6.25	10	7.54	10
Dust (μg/m^3)	7.40	9	7.21	9	0	13
Relative humidity (%)	9.40	3	9.45	2	7.77	9
Temperature (°C)	7.90	7	8.04	5	8.84	3
Solar energy (kW/m^2)	5.69	11	7.25	8	7.47	11
Wind speed	5.66	12	6.16	12	8.32	6
Wind direction (degree)	4.50	13	5.77	13	8.24	7
Total	100		100		100	
Ozone variation due to meteorology	33.15		36.69		40.64	
Ozone variation due to the other pollutants	66.85		63.313		59.36	

speed and wind direction. However, wind speed and wind direction are less sensitive predictors based on the results of models I and II.

Of considerable interest is the fact that CO_2 is a major variable for model III. This is not the case for models I and II. CO and dust are not important variables for ozone forecasting in any of the three models. For CO, this is expected because it participates very little in the chemistry of urban ozone (Ruiz-Suárez et al., 1995). It has been reported in previous studies that CO is the least reactive pollutant measured. It may be transported in the atmosphere for long distances without reacting with other species. CO can therefore be used for indirect quantification of wind drift.

The results shown in Table 4 also indicate ozone concentration variability as a result of meteorological conditions and other pollutants. For model I, the meteorological conditions describe 33.15% of the variation in ozone concentration. The remaining variability could then be attributed to chemical data. However, in the second model 36.69% of the variation in ozone levels was explained by the influence of meteorological conditions, while for model III, the contribution of meteorological conditions was even higher, at 40.64%.

7. Conclusion

A neural network approach was used to explore the complex relationship between ozone and other variables based on ambient air monitoring measurements. The results offer an insight into the dependence of ozone concentrations on other primary pollutant concentrations and meteorological conditions.

It was found that the models' predictions and the real observations were consistent. The relative importance of the various input variables was also investigated. The results indicated the dependence of ozone concentrations on the other pollutants and on meteorological conditions. The contribution of meteorology on the ozone concentration variations was found to fall within the range 33.15–40.64%. The remaining variability was attributed to chemical pollutants. It was determined that relative humidity is the meteorological parameter with the highest contribution to ozone variations. It was also found that nitrogen oxide, sulfur dioxide, relative humidity, non-methane hydrocarbon and nitrogen dioxide have the most effect on the predicted ozone concentrations. In addition, temperature played an important role while solar energy played a less important role than expected.

We can conclude that the neural network can be used in modelling and predicting the ground level concentrations of ozone. Clearly, this study has indicated the potential of the neural network approach for capturing the non-linear interactions between ozone and other factors and for the identification of the relative importance of these factors. Neural network modelling, therefore, provides a simple means of modelling and analysis of air pollutants and could be used in conjunction with other methods.

Finally, it is important to note that the rate of formation of ozone is a function of the nature of the hydrocarbon molecule. Hydrocarbons differ in their rate of interaction with the NO_2 photolytic cycle. Therefore, the training and testing sets used for the independent input parameters could, in principle, be substantially improved by specific hydrocarbon measurements. It would also be particularly interesting to observe the ranking of the variables and to determine whether the ranking of the variables remains constant when the selected hydrocarbon species are included as independent input parameters for the neural network approach. Therefore, it is recommended that the effect of selected hydrocarbon species should be examined in future studies. Research should also be directed towards elaborating on the effect of other meteorological data when they are added as input to the model. For example, what is the dependence of ozone concentration when the inversion layer is used as an input to the model?

References

Abdul-Wahab, S.A., Bouhamra, W., Ettouney, H., Sowerby, B., Crittenden, B.D., 1996. Development of statistical model for prediction of ozone levels in Shuaiba Industrial Area in Kuwait. Environmental Science and Pollution Research (ESPR) 3 (4), 195–204.

Abdul-Wahab, S.A., Bouhamra, W., Ettouney, H., Sowerby, B., Crittenden, B.D., 2000. Analysis of ozone pollution in the Shuaiba Industrial Area in Kuwait. The International Journal of Environmental Studies, Section B: Environmental Science and Technology 57, 207–224.

Azzi, M., Johnson, G.M., Hyde, R., Young, M., 1995. Prediction of NO_2 and O_2 and O_3 concentrations for NO_x plumes photochemically reacting in urban air. Mathematical and Computational Modelling 21 (9), 39–48.

Bouhamra, W., Abdul-Wahab, S.A., 1999. Description of outdoor air quality in a typical residential area in Kuwait. Environmental Pollution 105, 221–229.

Bower, J.S., Broughton, G.F., Dando, M.T., Stevenson, K.J., Lampert, J.E., Sweeney, B.P. et al., 1989. Surface ozone concentrations in the U.K. in 1987–1988. Atmospheric Environment 23 (9), 2003–2016.

Boznar, M., Lesjak, M., Mlakar, P., 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.

Capone, R.L., 1996. Predicting downwind air quality with a neural network. Presentation to EPA/Office of Air Quality Planning Standards (OAQPS).

Comrie, A.C., 1997. Comparing neural networks and regression models for ozone forecasting. Journal of the Air and Waste Management Association 47, 653–663.

Crowe, W., DeFries, T.H., 1996. Use of observation based models to predict ambient ozone levels. Presentation to EPA/Office of Air Quality Planning Standards (OAQPS).

Elkamel, A., Abdul-Wahab, S.A., Bouhamra, W., Alper, E., 2001. Measurement and prediction of ozone levels around a heavily industrialized area: a neural network approach. Advances in Environmental Research 5, 47–59.

EPA, 1999. Guideline for developing an ozone forecasting program. Environmental Protection Agency, EPA-454/R-99-009.

Gardner, M.W., 1996. An investigation into the importance of meteorology in determining surface ozone concentrations: a neural network approach. In: First International Conference on GeoComputation. University of Leeds, Leeds School of Geography, GMAP Ltd, Geoinformation International, Elsevier Science pp. 17–19.

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

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

Gulf Union Company, 1993. Map of Kuwait with city maps of central Kuwait and Kuwait urban areas. Gulf Union Company, Safat, Kuwait.

Hsu, K.J., 1992. Time series analysis of the interdependence among air pollutants. Atmospheric Environment 26 (4), 491–503.

Johnson, G.M., 1991. An outline of air quality measurements and interpretations and tests of airshed models made possible by application of Airtrak instrumentation. Explanatory notes and supplementary materials for seminar series held at EPA. Research Triangle Park.

Lorenzini, G., Nali, C., Panicucci, A., 1994. Surface ozone in PISA (Italy): a six year study. Atmospheric Environment 28 (19), 3155–3164.

Maier, H.R., Dandy, G.C., 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. Environmental Modelling and Software 15 (1), 101–124.

Maynard, E.S., 1984. Review of the attributes and performance of 10 rural diffusion models. American Meteorological Society 65 (6), 554–558.

Poulid, O., Dickerson, R., Doddridge, B., Holland, J., Wardell, R., Watkins, J., 1991. Trace gas concentrations and meteorology in rural Virginia: ozone and carbon monoxide. Journal of Geophysical Research 96 (D12), 22461–22475.

Ruiz-Suárez, J.C., Mayora-Ibarra, O.A., Torres-Jiménez, J., Ruiz-Suárez, L.G., 1995. Short-term ozone forecasting by artificial neural networks. Advances in Engineering Software 23, 143–149.

Rumelhart, D., McClelland, J., 1986. PDP research group parallel distributed processing. In:. Explorations in the Microstructure of Cognition, vol. 11: Foundations. Bradford Books, MIT Press, Cambridge, MA.

Salop, J., Wakelyn, N.T., Lev, G.F., Middleton, E.M., Gervin, J.C., 1983. The application of forest classification from landsat data as a basis for natural hydrocarbon emission estimation and photochemical oxidant model simulations in South Eastern Virginia. Journal of the Air Pollution Control Association 33, 17–22.

Simpson, P.K., 1990. Artificial Neural Systems: Foundations, Paradigms, Applications, and Implementations. Pergamon Press, New York.

Stanley, J., 1990. Introduction to Neural Networks, 3rd edition. Sierra Madre, California Scientific Software.

Topcu, N., Keskinler, B., Bayramoglu, M., Akcay, M., 1993. Air pollution modeling in Erzurum City. Environmental Pollution 79, 9–13.

Varshney, C.K., Aggarwal, M., 1992. Ozone pollution in the urban atmosphere of Delhi. Atmospheric Environment 26B (3), 291–294.

Zannetti, P., 1983. Air pollution monitoring system for the residential area around Shuaiba Industrial Area. Study submitted to Environmental Protection Council, Kuwait.

Zannetti, P., 1994. Computer modelling of air pollution, science, art or fiction: Air pollution II. In: Computer Simulation, vol. 1. Computational Publications, Southampton, Boston.