

## Modeling Vehicle Air Pollution Attributes by Artificial Neural Network

M. Vaziri<sup>1</sup>

The municipalities of large Iranian cities require regular air pollutant emission monitoring for motor vehicles. The recorded data include information about commonly regulated air pollutants and motor vehicle characteristics which were randomly selected and studied from the Tehran database. Regression analysis was found ineffective in modeling the relationships among air pollutant, meteorological and motor vehicle characteristics. However, the artificial neural network, ANN, modeling seem to be a useful tool for prediction of air pollutants. The developed ANN models were found to be superior to the developed regression models. Application of the study approach and results could improve motor vehicle air quality monitoring and control activities for urban areas.

### INTRODUCTION

Air quality is a dynamic and complex environmental phenomenon exhibiting large temporal and spatial variation. Clean and dry air consist of 78.09% nitrogen, 20.94% oxygen and the remaining 0.97% as a mixture of other gases including argon, neon, carbon dioxide, helium, krypton, xenon and nitrous oxides. Air pollutants are emitted by a large variety of anthropogenic and natural sources. Anthropogenic sources tend to be more concentrated in urban areas, whereas natural sources are more dispersed in nature. Human activities cause a vast number of polluting substances to be emitted into the air [1-5]. The list of pollutants includes common combustion products and a large array of industrial products of specialized processes. Air pollution has significant local, regional and global adverse environmental impacts. The emissions accumulate in the air and produce concentrations that are measured as pollutant density. The relation between emissions and concentrations is often complex and influenced by several factors including characteristics of emissions and topographical and meteorological attributes. Vehicle emission are highly variable and related to several factors including vehicle type and age, ambient temperature, altitude of the city, the operating cycle and driving pattern. The operating cycle consists of starts and stops, speed changes, and idling time. Diesel engines emissions are better than gasoline engines in

some air pollutant categories, but they are generally worse on particulate and oxides of nitrogen [6-11].

Recently, air pollution has become an issue of worldwide concern and a steadily growing problem for urban areas in developing countries. Urban growth is accompanied by increased traffic and energy consumption. The number of automobiles is increasing at a rate which is even more rapid than urban population growth [12-14]. Internal combustion engines usually provide the propulsion for cars and its operation involves a very rapid batch-burning process. After ignition, the flame progresses in the combustion chamber until it cools and stops burning or is quenched as it nears the chamber wall. This process leaves a layer of unburned hydrocarbon next to the wall, a portion of which subsequently mixes with the burned charge and escapes with the exhaust. In operation, the air/fuel ratio entering the engine is not usually the optimum theoretical mixture. The air/fuel flow is such that no two cylinders or cycles get exactly the same mixture homogeneously distributed throughout the combustion chamber. Thus, combustion is not complete, particularly when the mixture is fuel rich. As a result of these and other factors, a complex mixture of exhaust products is emitted from the automobile exhaust-pipe. In addition to water, oxygen, nitrogen and carbon dioxide, these products include carbon monoxide, unburned hydrocarbons, partially burned hydrocarbons, hydrogen, oxides of nitrogen and various particulates such as lead and sulfur compounds [15]. Some kind of testing is required in order to determine accurately whether emission standards are being

1. Department of Civil Engineering, Sharif University of Technology, Tehran, I.R. Iran.

met. The permanent and mobile monitoring devices can measure the exhaust emission at different engine operating regimes.

A wide variety of options are available for controlling emissions varying from direct mechanical control of emissions from vehicles to broader demand and traffic management measures aimed at curbing the use of motor vehicles. Major approaches for road traffic air pollution management include application of alternate energy source and mechanic technology, precombustion control, combustion modification, post-combustion control, energy conservation and broader transportation system management.

Carbon monoxide is the most widely distributed urban air pollutant and exceeds the combined mass of all other major air pollutants. Road traffic is by far the largest source of CO emissions. The adverse health effects of CO are caused by its ability to reduce the quantity of oxygen that is delivered by the blood to the tissues and, possibly, to inhibit the utilization of oxygen within the tissues. The carbon dioxide is not, itself, an air pollutant; however, it is a major contributor to the greenhouse effect, climate change and global warming. Hydrocarbons are a very important component of photochemical smog. They contribute to lung disease and have a close affinity for diesel particulates. Emission of pollutants depends on the type of engine, regime of engine operation, type and quality of fuel, year model and age of vehicle, annual mileage traveled, maintenance and inspection regime. Descriptive relations of emitted pollutants with respect to vehicle, meteorological and traffic characteristics have been extensively reported in the literature [6-14].

The objective of the research reported herein was to study relationships between emission, meteorological and vehicle characteristics for Iranian highway vehicles. The study emission data consisted of 2000 records extracted from the Tehran Air Pollution Database. The artificial neural network, ANN, and regression modeling were used to model motor vehicle air pollution emission attributes [16-23]. Due to the complexity of nonlinear relationships, multivariate statistical models capturing vehicle, meteorological and traffic effects and interactions on pollutants emission are scarce in the literature [24-31]. The developed ANN models show an alternative way of modeling these complex relationships. Application of the study approach and results could improve motor vehicle air quality monitoring and control activities for urban areas in Iran.

## DATABASE

The municipality of large Iranian urban areas require regular air pollutant emission monitoring for motor vehicles. The available data are most comprehensive for Tehran. At several monitoring stations, the permanent

monitoring devices are used to measure the exhaust emission at idle engine operating regimes. For each motor vehicle test, results of emission tests are kept in a separate record sheet. The record sheet information is processed and kept in written and electronic forms. The recorded data for 2000 motor vehicles were randomly selected from the individual record sheets of the Tehran database. For the same samples, the relevant meteorological information was extracted from the Iranian Meteorological Agency published reports [32]. Each of the 2000 study records consisted of 30 descriptive variables that reflected air pollution, motor vehicle and meteorological attributes. The relevant emission information consisted of 4 variables, namely emission density of carbon monoxide, hydrocarbons, oxygen and carbon dioxide, respectively. The relevant vehicle information consisted of 22 variables, namely case number, license plate number, testing date, validation date, testing center, model, type, mode of usage, year model, distance traveled, engine rotation, dwell angle, dwell percent, DC voltage, battery voltage, coil voltage, electric current, number of cylinders, magnet offset, engine vacuum, cycles and strobe timing, respectively. The meteorological information consisted of 4 variables, namely wind speed, rainfall, ambient temperature and humidity, respectively.

Univariate analysis of the 30 variables was mainly used for data validation [21]. The analysis also illustrated a number of interesting results. Seventy percent of the 2000 randomly selected motor vehicles were made in Iran. Ninety-five percent were automobiles. Seventy percent of them were 8 to 22 years old with a mean age of 15 years. The distance traveled showed a mean of 80,000 kilometers. Hydrocarbon emission had a mean of 389 PPM and a standard deviation of 194 PPM. To pass the test, the maximum accepted level of hydrocarbon, HC, is 500 PPM. Carbon monoxide emission, CO, showed a mean of 2.96% and a standard deviation of 1.49%. To pass the test, the maximum accepted level of carbon monoxide is 5%. The carbon monoxide concentration in clean air is normally 0.1 PPM. Oxygen emission, O<sub>2</sub>, showed a mean of 2.61% and a standard deviation of 3.45%. To pass the test, the maximum accepted level of oxygen is 3%. Carbon dioxide emission, CO<sub>2</sub>, showed a mean of 11.46% and a standard deviation of 3.14%. To pass the test, the minimum accepted level of carbon dioxide is 9%.

To develop an understanding of the basic inter-relationships among variables, a 30 × 30 correlation matrix of the database variables was investigated [21]. Although the size of the correlation coefficient matrix prevented its display herein, a derived summary for emission variables is shown in Table 1. Only the variables that are significantly correlated at a 0.05 level are shown in this table. The average significant correlation is about 24 percent which means that, on

**Table 1.** The results of the correlation analysis for emission variables.

Variable	CO	HC	O <sub>2</sub>	CO <sub>2</sub>
Year model	-0.129	-0.096	n/s	n/s
Distance traveled	n/s	n/s	n/s	-0.158
Motor rotation	n/s	n/s	0.077	n/s
Dwell angle	-0.059	n/s	n/s	n/s
Number of cylinders	0.091	n/s	-0.058	n/s
Battery voltage	n/s	-0.102	0.239	0.066
Vehicle type	n/s	n/s	n/s	0.024
Ambient temperature	n/s	-0.119	0.320	0.410
Humidity	n/s	n/s	n/s	0.049
CO	1.	0.156	-0.188	-0.223
HC	0.156	1.	-0.072	-0.051
O <sub>2</sub>	-0.188	-0.072	1.	-0.750
CO <sub>2</sub>	-0.223	-0.051	-0.750	1.

n/s : not significant at 0.05 level

average, each emission variable is significantly correlated with 24 percent of the other 29 variables. HC was found to be significantly and positively correlated with CO and negatively correlated with year model, battery voltage, ambient temperature, O<sub>2</sub> and CO<sub>2</sub> respectively. CO was found to be significantly and positively correlated with number of cylinders and HC, and negatively correlated with year model, motor dwell angle, O<sub>2</sub> and CO<sub>2</sub>, respectively. O<sub>2</sub> was found to be significantly and positively correlated with motor rotation, battery voltage and ambient temperature and negatively correlated with number of cylinders, HC, CO and CO<sub>2</sub>, respectively. CO<sub>2</sub> was found to be significantly and positively correlated with battery voltage, ambient temperature, vehicle type and humidity and negatively correlated with distance traveled, HC, CO and O<sub>2</sub>, respectively. Table 1 shows that among the three variable sets, the emission variables are mostly correlated with each other.

## REGRESSION MODELING

The method of least squares is a technique that yields the best fitting polynomial model of a postulated form to a set of observed data. The stepwise multiple linear regression analysis was used to develop emission models. The dependent variables were the four emission variables. The possible independent variables were the 22 vehicle variables and the 4 meteorological ones. Using forward stepwise linear regression analysis, several models were developed and evaluated for the study database of 2000 records [21]. Using coefficient of determination as the evaluation criterion, the selected models are reflected by the following equations:

$$HC = 2086.1 - 1.2(YR) - 5.1(TEM), \quad (1)$$

where HC is the exhaust hydrocarbon emission in part per million, PPM, *YR* is the year the vehicle was built, *TEM* is the ambient temperature in Celsius. The *t* statistics for the intercept 2086.1 is 2.0 and for the parameter estimates of *YR* and *TEM* are -2.1 and -4.1, respectively. The Root Mean Square Error (RMSE) for Equation 1 is 189 PPM.

$$CO = 31.8 - 0.02(YR) + 0.13(CYL), \quad (2)$$

where CO is the exhaust carbon monoxide emission in percent of the volume, *YR* is the year the vehicle was built and *CYL* is the number of cylinders. The *t* statistics for the intercept 31.8 is 3.4 and for the parameter estimates of *YR* and *CYL* are -3.1 and 2.2, respectively. RMSE of Equation 2 is 1.4%.

$$O_2 = -7.35 + 0.24(TEM) + 0.56(BAT), \quad (3)$$

where O<sub>2</sub> is the exhaust oxygen emission in percent of the volume, *TEM* is the ambient temperature in Celsius and *BAT* is the battery voltage. The *t* statistics for the intercept -7.35 is -5.5 and for the parameter estimates of *TEM* and *BAT* are 12.8 and 5.8, respectively. RMSE of Equation 3 is 3.1%.

$$CO_2 = 11.57 - 0.000015(KLM), \quad (4)$$

where CO<sub>2</sub> is the exhaust carbon dioxide emission in percent of the volume and *KLM* is the distance traveled. The *t* statistics for the intercept 11.57 is 9.4 and for the parameter estimate of *KLM* is -2.1. RMSE of Equation 4 is 3.4%.

For Equations 1 to 4, the *t* statistics for parameter estimates showed a significant level of 0.05 or better. Equations 1 to 4 can be used for prediction; however, their large RMSE with respect to the air pollution variables' mean and standard deviation make them less appealing. The coefficient of determination for the developed models of 1 to 4 were less than 0.20, reflecting that the multiple linear regression can not efficiently capture the relationships. This also has been demonstrated by appearance of one or two variables as significant variables in the developed linear models.

## ANN MODELING

Artificial neural networks take their name from the network of nerve cells in the brain. They are inspired by knowledge from neuroscience and draw their methods to a large extent from statistical physics. Their potential applications lie in the field of computer science and engineering [16-23]. The idea of modeling the functioning of the brain draws inspiration from the fact that brain is a very robust processing body. The robustness of brain comes from the millions of neurons and their interconnections and from the parallel processing of

information. The powerful structure of brain provides it with the capability to analyze extremely complex input-output systems.

Inspired by the architecture of brain, artificial neural network is a parallel, distributed information processing system composed of many simple processing elements, which are interconnected via synaptic or weighted connections. They are powerful mathematical tools for modeling complex and sometimes intractable functions between system inputs and outputs. This is because of the fact that neural networks extract the essence of the relationship between system inputs and outputs through the data made available to them as training information. The key characteristics of an ANN include: number of processing elements in each layer, number of layers, type of transfer function and learning rule. Variety of ANN architectures, such as back-error propagation, Kohonen layer, competitive learning, Adaline and Madaline, have been used in transportation and environment modeling [16-23].

Back-error propagation refers to the method by which ANN is trained. A basic backpropagation ANN consists of three layers, namely input layer, hidden layer and output layer, all interconnected with different weights. There are no criteria for determining the appropriate number of layers and processing elements. Backpropagation ANN takes its name from how it handles error. In the training of backpropagation networks, the error information is passed from the output layer to the input layer. Element connection weights are adjusted by comparing the desired outputs with actual outputs using a mathematical rule such as gradient descent method. Delta rule is generally used as the training algorithm. The function most commonly used for error is the sum of the square of the difference between the actual and desired output layer elements' output. For a backpropagation ANN with sigmoid transfer function, the elements' outputs are defined as follows:

$$Y_j = 1/(1 + e^{-\sum_i w_{ij} Y_i}), \quad (5)$$

where  $Y_j$  is the ANN's actual output for  $j$ th element,  $w_{ij}$ , the weight of connection between  $j$ th element and the  $i$ th element in the previous layer of the  $j$ th element and  $Y_i$ , the  $i$ th input for the element  $j$  or the output of the  $i$ th element in the previous layer. In the training of backpropagation ANN, the error information is passed backward from the output layer to the input layer. A network learns by successive repetition and training based on the observed information, making fewer errors with each iteration. The most commonly used function for the errors is the sum of the squared errors of the output elements. The  $w_{ij}$ 's are adjusted based on Delta rule which is:

$$E = 0.5 \sum_j (Y_j - Y_{dj})^2, \quad (6)$$

where  $E$  is the sum of the square of errors,  $Y_j$  is defined as in Equation 5 and  $Y_{dj}$  is the ANN's desired output or the observed data for  $j$ th element. To minimize the error, take its derivative in Equation 6 with respect to  $w_{ij}$  as:

$$\partial E / \partial w_{ij} = Y_i Y_j (1 - Y_j) \Omega_j, \quad (7)$$

where  $\partial E / \partial w_{ij}$  is the derivative of  $E$  with respect to the weight between elements  $i$  and  $j$ ,  $Y_i$  and  $Y_j$  are outputs of elements  $i$  and  $j$ ,  $\Omega_j = (Y_j - Y_{dj})$  for output layer elements and  $\Omega_j = \sum_k w_{jk} Y_k (1 - Y_k) \Omega_k$  for hidden layer elements when  $k$  presents the number of elements in the next layer that element  $j$  is connected to. The error can be calculated directly for the links going into the output layer elements. For hidden layer elements, however, the derivative of Equation 7 depends on values calculated at all the layers that come after it. That is, the value  $\Omega$  must be backpropagated through the network to calculate the derivatives. For each sample pattern, a forward pass through network with some initial values for  $w_{ij}$ 's produces an output pattern. Then using Equations 5 to 7, backpropagation algorithm starts with choosing a step size  $\delta$  and then updating the  $w_{ij}$ 's with the following relation:

$$\Delta w_{ij} = -\delta Y_i Y_j (1 - Y_j) \Omega_j, \quad (8)$$

where  $\Delta w_{ij}$  is the change for  $w_{ij}$  and all other variables are defined as in Equations 5 to 7. The algorithm continues until the network is trained and the sum of the square of errors in Equation 6 becomes smaller than a prespecified error limit.

Results from the regression analysis of Equations 1 to 4 lead to the notion that there are more complex relationships between emissions, vehicle and environmental characteristics. To predict HC, CO, O<sub>2</sub> and CO<sub>2</sub> emissions, among many available options, several ANNs were trained and tested. A software called Neural Work Explorer was used to develop ANNs [22]. The developed ANNs with all the 26 vehicle and meteorological variables were found not superior to the developed ANNs in Table 1 with nine selected vehicle and meteorological variables, in regard to training convergence and testing results. The selected ANNs would use the motor vehicle and ambient characteristics as inputs to nine processing elements in the input layer. These characteristics were distance traveled, year model of vehicle, motor rotation per minute, motor dwell angle, number of cylinders, battery voltage, vehicle type, ambient temperature and humidity. As reflected in Table 1, the selected input variables are correlated with emission variables.

With 9 processing elements in the input layer, several one and two hidden layer backpropagation ANNs were trained and tested. Indeed, the actual architecture of any backpropagation ANN is problem

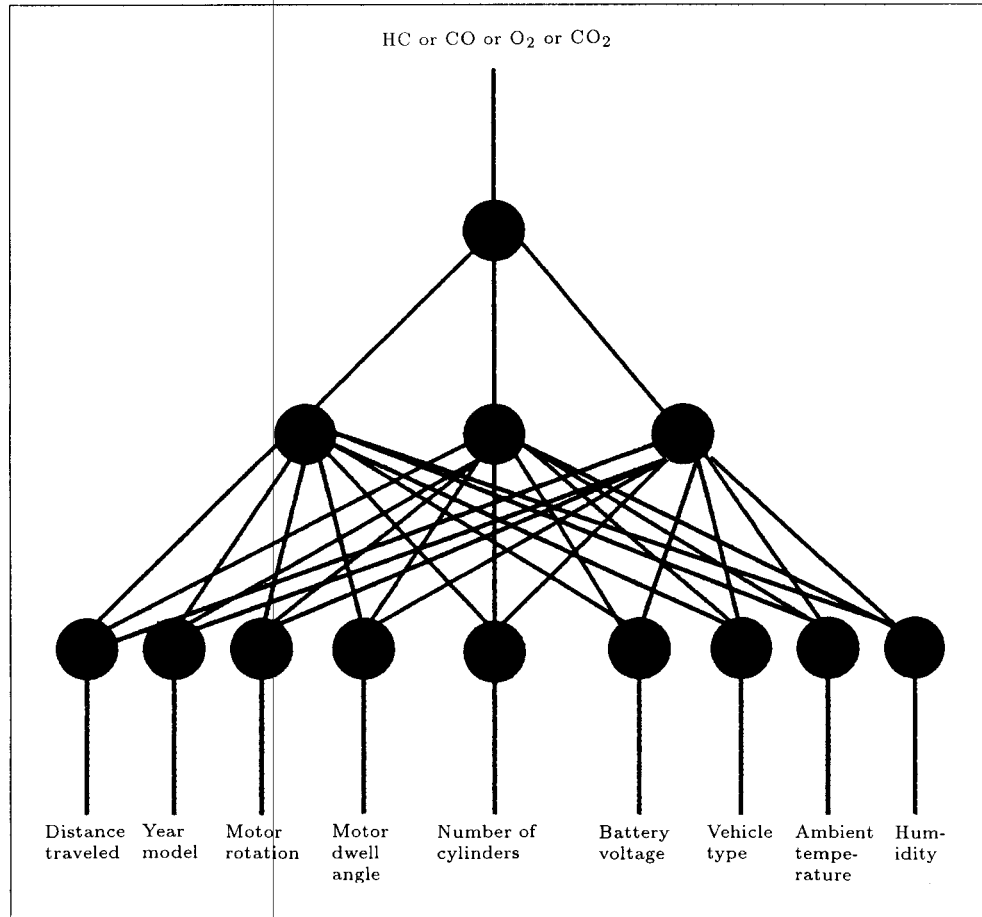


Figure 1. Structure of the selected backpropagation ANN.

dependent. The selected ANNs, which had simpler architecture and smaller RMSE, for the testing data, were the basic three layer networks shown in Figure 1. There were 3 processing elements in the hidden layer. In the output layer, the processing element provided the emission variable estimates. In this study, three types of transfer functions, namely hyperbolic tangent, sigmoid and sine were tried. For each transfer function, the training data were properly scaled by the software. The sigmoid transfer function, given by Equation 5, was selected for its superiority in training convergence and testing results. The applied learning rule to the hidden and output layers was the cumulative Delta learning rule, which accumulated the weight changes over several presentations of training examples and then applied them to the weights. The key parameters of the cumulative Delta rule include learning coefficients, momentum and epoch [19,20,22,23]. After several trials, the epoch, momentum and learning coefficients of the hidden and output layers were set at 16, 0.5, 0.4 and 0.2, respectively. The momentum and learning coefficients were gradually reduced for higher numbers of training iterations, for convergence to the preselected RMSE values. After

more than 30,000 iterations that randomly presented 1800 records as ANNs' training data, the trainings converged to RMSE of 0.1 for standardized output. In this way, four ANNs were trained and selected. The trained backpropagation ANNs were then tested with the remaining 200 records. For the testing data, RMSE of the trained ANNs and regression models are compared and summarized in Table 2. The testing data showed an average RMSE reduction of 90%, when four ANNs' predictions were compared with Equations 1 to 4. The developed ANNs were found superior to the developed regression models. For emission prediction, the developed ANNs can be used when relevant vehicle and meteorological data is available.

Table 2. RMSE of the trained ANNs and regression models.

Variables	ANN	Regression
CO	0.23%	1.92%
HC	19 PPM	178 PPM
O <sub>2</sub>	0.52%	3.13%
CO <sub>2</sub>	0.15%	2.59%

## CONCLUSION

Among motor vehicle and ambient characteristics available in the study database, 9 variables were found more effective in predicting motor vehicle exhaust air pollution emission. These variables were distance traveled, year model of vehicle, motor rotation per minute, motor dwell angle, number of cylinders, battery voltage, vehicle type, ambient temperature and humidity. The artificial neural network modeling was found to be a useful tool for predictions of motor vehicle air pollution attributes. The developed ANNs were found superior to the developed regression models. Although the trained ANNs were dependent on 2000 randomly selected records from the Tehran motor vehicle air pollution database, the same methodology may be applied to predict any motor vehicle air pollution attributes.

## ACKNOWLEDGMENT

The technical and financial support provided by Sharif University of Technology is gratefully acknowledged. The author also wishes to thank Mr. M. Moazemi for the extensive data extraction process.

## REFERENCES

1. Boughey, A.S., *Man and Environment*, Second Edition, MacMillan, New York, USA (1975).
2. Horne, R.A., *Chemistry of Environment*, John Wiley & Sons, New York, USA (1978).
3. Khopkar, S.M., *Environmental Pollution Analysis*, Newage International, New Delhi, India (1993).
4. Marr, I.L. and Cresser, M.S., *Environmental Chemical Analysis*, International Textbooks, New York, USA (1983).
5. Sen, A.K., *Environmental Management and Planning*, John Wiley & Sons, New York, USA (1988).
6. Horowitz, J.L., *Air Quality Analysis for Urban Transportation Planning*, The MIT Press, London, England (1982).
7. Noll, K.E., *Air Monitoring Survey Design*, Ann Arbor Science Publishers, Ann Arbor, USA (1977).
8. Perkins, H.C., *Air Pollution*, McGraw-Hill, New York, USA (1974).
9. Seinfeld, J.H., *Air Pollution: Physical and Chemical Fundamentals*, McGraw Hill, New York, USA (1975).
10. Sigrist, M.W., *Air Monitoring by Spectroscopic Techniques*, John Wiley & Sons, New York, USA (1994).
11. Stern, A.C., *Air Pollution*, Third Edition, 1-5, Academic Press, New York, USA (1977).
12. Faiz, A. and Carbajo, J., *Automotive Air Pollution Control Strategic Options for Developing Countries*, World Bank, Washington, D.C., USA (1991).
13. Sanders, C.L., *Toxicological Aspects of Energy Production*, Battelle Press, Columbus, USA (1986).
14. Theodore, L. and Buonicore, M., *Air Pollution Control Equipment*, 1&2, CRC Press, Boca Raton, USA (1988).
15. Brehob, W.M. "Mechanisms of pollution formation and control from automotive sources", *Society of Automotive Engineers*, Special Report 365, Detroit, USA (1971).
16. Caudill, M. and Butler, C., *Understanding Neural Networks: Computer Explorations*, The MIT Press, London, England (1993).
17. Chirta, S.P. "Neural net applications in chemical engineering", *AI Expert*, 7(11), pp 20-25 (1992).
18. Crainic, T.G. et al. "Parallel computing: New opportunities for transportation research", *Scientia Iranica*, 2(3), Sharif University of Technology, I.R. Iran, pp 231-244 (1995).
19. Fausett, L. *Fundamentals of Neural Networks*, Prentice-Hall, Englewood Cliffs, USA (1994).
20. Hecht-Nielsen, R., *Neurocomputing*, Addison-Wesley, New York, USA (1989).
21. Nie, N.H. et al., *Statistical Package for the Social Sciences*, McGraw-Hill, New York, USA (1975).
22. Neural-Ware Inc., *Neural Computing*, Pittsburgh, USA (1991).
23. Wasserman, P.D., *Neural Computing, Theory and Practice*, Van Nostrand Reinhold, New York, USA (1989).
24. Barth, M. et al. "Modal emissions modeling: A physical approach", *Transportation Research Record*, No 1520, pp 81-89 (1996).
25. Degobert, P., *Automobiles and Pollution*, Thomas Telford, London, England (1980).
26. Flagan, R.C. and Seinfeld, J.H., *Fundamentals of Air Pollution Engineering*, Prentice Hall, New York, USA (1988).
27. Moseholm, L. et al. "Forecasting carbon monoxide concentrations near a sheltered intersection using video traffic surveillance and neural networks", *Transportation Research, Part D: Transport and Environment*, 1D(1), pp 15-29 (1996).
28. Rallis, T., *Intercity Transport, Engineering and Planning*, John Wiley & Sons, New York, USA (1977).
29. Rao, M.N. and Rao, H.V., *Air Pollution*, McGraw Hill, New York, USA (1990).
30. Stedman, D.H. "Automobile carbon monoxide emissions", *Environmental Sciences and Technology*, 23, pp 147-149 (1989).
31. Wayne, O. et al. "Carbon monoxide exposures inside an automobile traveling on an urban arterial highway", *Journal of Air & Waste Management Association*, 44, pp 1010-1018 (1994).
32. Iranian Meteorological Agency, *Tehran Meteorological Reports*, Iranian Meteorological Agency, Tehran, I.R. Iran (1994).