ABSTRACT

Introduction: As a metropolitan area in Iran, Tehran is exposed to damage from air pollution due to its large population and pollutants from various sources. Accordingly, research on damage induced by air pollution in this city seems necessary. The main purpose of this study was to forecast ozone in the city of Tehran. Considering the hazards of ozone (O3) gas on human health and the environment and its ascending trend over the past decades, it is also essential to study and predict its quantities in the air. Forecasting ozone in the air can be further used to prevent and control pollution by authorities.

Material and Methods: Using an analytical-applied research method, this study was to predict ozone gas in this metropolitan area via daily ozone data of air quality measurement stations, traffic variables, green space, as well as time factors such as one-day time delay. In this regard, an artificial neural network (ANN) model was employed to forecast ozone concentration using the MATLAB software.

Results: The results of the ANN model were compared with a linear regression one. Correlation coefficient and root-mean-square error (RMSE) of the ANN model were subsequently compared with $R^2=0.734$ and RMSE=0.56 as well as $R^2=0.608$ and RMSE=11.69 regression equations.

Conclusion: It was concluded that the error in the ANN model was smaller than that in the regression one. According to the results of the sensitivity analysis of the season parameters, the length of sunshine hours had the most significant effect on the amount of ozone gas in Tehran air.

Keywords: Air pollution, Ozone gas, Artificial neural network, Multivariate regression, Sensitivity analysis

1. INTRODUCTION

The problem of environmental pollution has been undoubtedly recognized today as one of the main challenges facing urban life, since it is annually causing enormous damage to the lives and finances of residents of large cities. In the meantime, the problem of air pollution is high on the list of pollution hazards and a review of the World Health Organization (WHO) statistics can provide enough evidence in this regard (1). According to a 2013 report by the International Institute of Cancer Research (IARC), air pollution and particulate matter (PM) are classified as 100% carcinogenic to humans. They are also the fourth leading cause of death in the world, as well as the seventh most common risk factor in Iran, with reference to the burden of diseases related to global air pollution (2). One method to manage and control the problem of air pollution in the city of Tehran, as a metropolis, is to utilize scientific and practical instrumentation to approximate and to forecast the increasing rate of each pollution factor (3). Recently, several models have been introduced to predict the quantities of air pollution in different cities. Artificial neural networks (ANNs) are one of the most accurate
tools in the process of optimization as such networks can process and learn simultaneously (4). The bulk of advances in ANNs are also associated with new structures and learning methods. The given networks can be further detected as the electronic model of the central nervous system (CNS) in the human brain (5). In other words, an ANN is a mathematical model that is capable of modeling and creating non-linear relationships for interpolations (6). Accordingly, many scholars in this field have conducted several studies including the one (7) in the polluted industrial areas of Slovenia. In this sense, comparing ANNs and classic regression models is common in the determination of modeling techniques (4).

2. MATERIALS AND METHODS
To fulfill this research, the data were collected from the 2015 Tehran Air Quality Control Company (AQCC), Tehran Meteorological Organization, Tehran Traffic Organization, and Tehran Gardens and Green Spaces Organization. The parameters measured in these organizations included ozone (O₃) concentration, green space information, weather parameters, such as temperature (°C), traffic parameters, e.g., length of north-south (y) and east-west (x) crossings (km), average number of vehicles at intersections, as well as time parameters such as one-day time delay (viz. ozone concentration in the previous day), two-day time delay (namely, ozone concentration in the past two days), and the day of the year (i.e., 1-365). The desired month was from the desired year (1-30) and the season was from the desired year (1-4) and the warm and cold seasons (1-2). In this research, the meteorological and pollution monitoring stations were located close to each other and the green space information as well as the traffic information of the streets near the stations. In order to predict air pollution, the given variables were studied using an ANN model, whose accuracy was measured by comparing its output and calculated indices including correlation coefficient (R²), mean absolute error (MAE), mean square error (MSE), and root-mean-square error (RMSE) (8). To perform sensitivity analysis and to calculate the output sensitivity of the ANNs, each of the variables of the model was kept constant of all variables equal to their mean. The model outputs were then calculated. This process was performed for each of the input variables and the sensitivity of the model results to each of the variables was consequently determined.

Multiple linear regression (MLR) analysis was ultimately completed using the stepwise regression method. Therefore, the variables that had a significant effect on the dependent one were utilized. In this technique, the measured daily ozone concentration was considered as a dependent variable and the effective factors were taken into account as independent variables.

3. RESULTS AND DISCUSSION
To optimize the ANN, it was estimated using 61 neurons and two hidden layers. According to the trained network results, with two hidden layers and 30 neurons in each layer using sigmoid tangent function in both hidden layers, linear function in the output layer, optimal correlation in the network results, and homogeneity of correlation coefficients in training, validation, and data test; the total data as well as the minimum MSE of the entire data (as listed in Table 1) were found very accurate in predicting ozone concentration in Tehran air. The MLR model was further used to investigate the relationship between changes in temporal, climatic, traffic, and green space factors on ozone concentration in Tehran. The study findings revealed that prediction of ozone concentration by the stepwise regression method via the variables of temperature, sunshine hours, wind speed, one-day

<table>
<thead>
<tr>
<th>Statistic</th>
<th>R²</th>
<th>MSE</th>
<th>RMSE</th>
<th>MAE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Training</td>
<td>0.851</td>
<td>0.169</td>
<td>0.411</td>
<td>0.339</td>
</tr>
<tr>
<td>Validation</td>
<td>0.597</td>
<td>0.412</td>
<td>0.641</td>
<td>0.483</td>
</tr>
<tr>
<td>Test</td>
<td>0.558</td>
<td>0.536</td>
<td>0.732</td>
<td>0.467</td>
</tr>
<tr>
<td>All</td>
<td>0.734</td>
<td>0.32</td>
<td>0.566</td>
<td>0.388</td>
</tr>
</tbody>
</table>

Figure 1. Results of Multivariate regression model in the stepwise method in forecasting Ozone concentration in Tehran

Table 1. Results of ANN model to forecast ozone concentration in Tehran air
The first objective of this study was to determine the model of ozone density forecast in Tehran air using ANNs. The results showed that the perceptron neural network designed with two hidden layers and 30 neurons in each hidden layer and the sigmoid tangent function had a good capability to model the changes of ozone density in the city of Tehran. The model obtained in this study provided the feasibility of predicting the quantities of ozone condensation was related to the season/year variables with an effective coefficient of 8.12 and the sunshine hours having an effective coefficient of 4.19, implying that other variables had less impact on ozone concentration in Tehran air.

4. CONCLUSION
The first objective of this study was to determine the model of ozone density forecast in Tehran air using ANNs. The results showed that the perceptron neural network designed with two hidden layers and 30 neurons in each hidden layer and the sigmoid tangent function had a good capability to model the changes of ozone density in the city of Tehran. The model obtained in this study provided the feasibility of predicting the quantities of ozone condensation and enabling it to be utilized in the spatial decision system of air quality management and with the help of optimal urban transport management of Tehran metropolitan. The second objective was to compare the results of the modeling prediction of Tehran ozone density by an ANN with an MLR model. The results revealed that the ANN model with a correlation coefficient of 0.85 was more effective in predicting ozone concentration than the MLR methods with a coefficient of determination of 0.78. Besides, the third objective was to explore the governing relationships between temporal, climatic, traffic, and green space parameters and the amount of ozone concentration in Tehran air. Performing sensitivity analysis and identifying the most effective parameters also showed that the season/year parameters and the sunshine hours were prioritized in order to predict ozone concentration in Tehran air, which could be effective in managing air pollution in this metropolitan area. Another significant parameter shaping ozone concentration was the length of street x and then the warm/cold season. This helped measure the amount of ozone in the air that day, and had a major impact on managing air pollution in the city. Accordingly, management decisions can be made to prevent some effects of air pollution, such as the less vulnerable strata of society, i.e., children, the elderly, and patients with cardiovascular diseases (CVDs). Forecasting air pollution in the coming days can thus help make managerial decisions to control air pollution and to provide the basic solutions and warnings. This issue is of particular importance in forecasting air pollution in industrial and populated cities such as Tehran.

6. REFERENCES