SHORT COMMUNICATION

Artificial neural network forecast application for fine particulate matter concentration using meteorological data

M. Memarianfard, A.M. Hatami*, M. Memarianfard

Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran

Received 24 September 2016; revised 20 December 2016; accepted 28 December 2016; available online 1 June 2017

ABSTRACT: Most parts of the urban areas are faced with the problem of floating fine particulate matter. Therefore, it is crucial to estimate the amounts of fine particulate matter concentrations through the urban atmosphere. In this research, an artificial neural network technique was utilized to model the PM$_{2.5}$ dispersion in Tehran City. Factors which are influencing the predicted value consist of weather-related and air pollution-related data, i.e. wind speed, humidity, temperature, SO$_2$, CO, NO$_2$, and PM$_{2.5}$ as target values. These factors have been considered in 19 measuring stations (zones) over urban area across Tehran City during four years, from March 2011 to March 2015. The results indicate that the network with hidden layer including six neurons at training epoch 113, has the best performance with the lowest error value (MSE=0.049438) on considering PM$_{2.5}$ concentrations across metropolitan areas in Tehran. Furthermore, the “R” value for regression analysis of training, validation, test, and all data are 0.65898, 0.6419, 0.54027, and 0.62331, respectively. This study also represents the artificial neural networks have satisfactory implemented for resolving complex patterns in the field of air pollution.

KEYWORDS: Air pollution; Artificial neural network (ANN); Meteorological data; PM$_{2.5}$ concentration; Tehran City

INTRODUCTION

Different kinds of airborne microbes suffer different effects because of meteorological parameters as well as influence their abundance (Fierer et al., 2008; Oliveira et al., 2009; Pyrri and Kapsanaki-Gotsi 2011; Li et al., 2011; Raisi et al., 2013). In reality, any airborne microbe has got certain effects that can be specifically considered in a special solution which concerns that microbe. In this research, to ease the solution all airborne microbes are assumed together an average effect which must be obtained. There are three main meteorological parameters effects as temperature ($t$), wind speed ($w$) and humidity ($h$), which have been defined and measured well and assessed in the paper. There are great concerns regarding visibility reduction, adverse health effects, and climate change due to atmospheric fine particle as PM$_{2.5}$ (with an aerodynamic diameter of 2.5 microns or less) and some other effects (Pope et al., 2002; Molina and Molina, 2004; Forster et al., 2007; Wang et al., 2016). PM$_{2.5}$ has high potential to adsorb or condense more toxic air pollutants such as organic compounds, metals, etc. and pose greater health risks, when its surface areas are increased (Oberdorster et al., 2005). Rapid process of industrializing, urbanization, energy utilization and the associated population growth in urban areas has caused serious particulate matter pollution in many cities worldwide, from Asian mega cities to even modern cities in Europe and America. (Han et al., 2014; Hu et al., 2014; Huang et al., 2014; Lary et al., 2014; Lin et al., 2014; Liu et al., 2013; Song et al., 2015; Zhang et al., 2012; Trizio et al., 2016). Based on the results of about 1600 cities of 91
countries, the world’s average PM$_{2.5}$ concentration is 28.4–56.8 μg/m$^3$ during 2008 to 2013, ranging from 26 to 208 μg/m$^3$ (WHO, 2014; WHO, 2015).

Tehran as the capital of Iran with approximately 8.5 million inhabitants is plagued by severe air pollution (Lotfabadi, 2014). In the past few years, due to urbanization, industrialization and population growth in Tehran, the issue of air pollution especially the PM has become extremely crucial. In recent years, the artificial neural technics (ANN) were started to be utilized for forecasting particulate matter concentrations (Perez and Reyes, 2002; Kukkonen et al., 2003; Lu et al., 2003; Ordieres et al., 2005; Zhou et al., 2014; Feng et al., 2015; Mohammadizadeh et al., 2016). The ability of ANN to change easily to suit the different situations has directed, to their use in the majority of scientific fields. Some applications of ANN in the atmospheric sciences during the 1990s were indicated in the studies that Gardner and Dorling had done (Gardner and Dorling, 1998). A variety of PM$_{2.5}$ sources such as power stations, transportation, natural disasters and heating systems in the residences have brought about considerable difficulties for PM$_{2.5}$ evaluation. In this research, the back-propagation learning algorithm for modeling and prediction was utilized. An error estimation technique has been conducted analysis upon indicating “R” and “IA”, the number of neurons on the hidden layer and epochs have been estimated. The study has been carried out in Tehran City capital of Iran and the data has been employed and performed in period of 2011 to 2015.

**MATERIALS AND METHODS**

The daily recorded data set was provided from the urban air recording stations in Tehran. The Geographical map and air pollution monitoring station locations in Tehran are shown in Fig. 1. The stations in Tehran districts are constantly monitoring air and report the daily data, which parts of recorded data are used in this paper. The recorded data set is composed of nitrogen dioxide, sulfur dioxide, carbon monoxide, PM$_{2.5}$, wind speed, temperature, and relative humidity (Qin et al., 2014).

According to the data, during 1461 days from 21 March 2011 to 20 March 2015, as presented in Table 1, including descriptive statistics of PM$_{2.5}$, SO$_2$, NO$_2$, CO, temperature, humidity, and wind speed. Moreover, the corresponding data histograms are shown in Fig. 2.
Neural network approach

The most ANN-based studies in the field of air pollution have suggested the back-propagation learning algorithm for modeling and prediction. In this algorithm, the data split into three parts:

1-assessing data set: which forms the bulk of the data that can be used for the training purposes?

2-examining data set: this data can be used to examine the performance quality of the trained model.

3-validating data set: this data can be used to validate the model.

Fig. 3 shows the architectural plan of the neural network proposed model with 6 predictor variables. The proposed model has been developed with the aim of presenting minimum possible error through predicting forecast computations, upon 6 entering parameters as input, 6 middle layers (hidden) parameters which are lead to one output parameter (no. 13) indicated as 6:6:1 scheme. It should be noted that the combination of SO$_2$, NO$_2$ and CO (in gaseous state) are named as precursors (input data) which are assessed upon archive correlations in ANN motor to create the amount of new particles as secondary PM$_{2.5}$ output.

**Table 1:** The descriptive statistics of daily air pollutants and meteorological parameters as input data (from 21 March 2011 to 20 March 2015)

<table>
<thead>
<tr>
<th>Statistic parameters</th>
<th>Temperature</th>
<th>Wind speed</th>
<th>Humidity</th>
<th>SO$_2$</th>
<th>NO$_2$</th>
<th>CO</th>
<th>PM$_{2.5}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>1461</td>
<td>1461</td>
<td>1461</td>
<td>1461</td>
<td>1461</td>
<td>1461</td>
<td>1461</td>
</tr>
<tr>
<td>Mean</td>
<td>18.419</td>
<td>3.065</td>
<td>34.631</td>
<td>30.42</td>
<td>56.72</td>
<td>38.49</td>
<td>97.80</td>
</tr>
<tr>
<td>Std. Error of Mean</td>
<td>0.25994</td>
<td>0.047403</td>
<td>0.469062</td>
<td>0.204</td>
<td>0.303</td>
<td>0.242</td>
<td>0.666</td>
</tr>
<tr>
<td>Median</td>
<td>18.950</td>
<td>2.750</td>
<td>30.250</td>
<td>30.00</td>
<td>57.00</td>
<td>37.00</td>
<td>96.00</td>
</tr>
<tr>
<td>Variance</td>
<td>98.722</td>
<td>3.283</td>
<td>321.448</td>
<td>60.818</td>
<td>134.055</td>
<td>85.706</td>
<td>647.975</td>
</tr>
<tr>
<td>Skewness</td>
<td>-0.141</td>
<td>9.711</td>
<td>0.932</td>
<td>2.080</td>
<td>-0.015</td>
<td>0.540</td>
<td>0.449</td>
</tr>
<tr>
<td>Std. Error of Skewness</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
<td>0.064</td>
</tr>
<tr>
<td>Kurtosis</td>
<td>-1.249</td>
<td>215.742</td>
<td>0.202</td>
<td>9.782</td>
<td>0.291</td>
<td>0.098</td>
<td>0.565</td>
</tr>
<tr>
<td>Std. Error of Kurtosis</td>
<td>0.128</td>
<td>0.128</td>
<td>0.128</td>
<td>0.128</td>
<td>0.128</td>
<td>0.128</td>
<td>0.128</td>
</tr>
<tr>
<td>Minimum</td>
<td>-2.025</td>
<td>0.500</td>
<td>9.00</td>
<td>16.00</td>
<td>20.00</td>
<td>19.00</td>
<td>26.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>36.625</td>
<td>46.000</td>
<td>93.625</td>
<td>88.00</td>
<td>97.00</td>
<td>77.00</td>
<td>204.00</td>
</tr>
</tbody>
</table>

**Fig. 2:** (A) The histograms of measured PM$_{2.5}$, NO$_2$, SO$_2$, and CO (24h. averages).
(B) The histograms of measured humidity, wind speed, and temperature (24h. averages)
Error and indicator ‘MSE’ & ‘R²’

For the performance of the ANN-based models to assess the accuracy of the estimation, there are several parameters, i.e. error and indicator: ‘MSE’ & ‘R²’, which are calculated according to Eqs. 1 and 2.

\[
(MSE) = \frac{1}{N} \sum_{i=1}^{N} [P_i - O_i]^2 \quad (1)
\]

\[
(R^2) = \frac{\sum_{i=1}^{N} [P_i - \bar{O}]^2}{\sum_{i=1}^{N} [O_i - \bar{O}]^2} \quad (2)
\]

Where N, Oᵢ, Pᵢ and \(\bar{O}\) are the number of observations, observed value, predicted value, and average value, respectively.

**RESULTS AND DISCUSSION**

In this paper, the ANN approach was utilized for...
modeling and the data was divided into three groups for training the network (50% of data), validating (25% of data), and testing (25% of data) the network. The function optimization technique used is the scaled conjugate gradient algorithm. The accomplishment of the model has been estimated with calculation of the mean square error (MSE) as the statistical criteria and “R”. Scattered plot of the observed and predicted concentrations of PM$_{2.5}$ using the data from 21 March 2011 to 20 March 2015 have been shown in Fig. 4. R$^2$ and MSE values were found to be 0.3885 and 0.049438, respectively.

In order to avoid over-training problem in this study, two indicators of the network were utilized, which are optimum choice of the hidden neuron numbers and error goal. According to input data, the neural network system that was required had 6 nodes on the input layer and one node on the output layer as well. Upon the error estimation method through analysis, the number of neurons on the hidden layer and epochs were estimated. The numbers of neurons were available in the hidden layer varied from 2 to 30. The MSE index was used to reach the optimum number of neurons in the hidden layer. As a result, the network layer with 6 neurons in the input layer, as a single hidden layer, including 6 hidden neurons plus a single neuron in the output layer will be able to represent best forecast.

MSE values for training, validation and test have been shown in Fig. 5(a and b) shows Schematic showing of the predicted and observed concentrations of PM$_{2.5}$ data during the four years, from 21 March 2011 to 20 March 2015.

R values for training, validation, test, and all data regression analysis have been shown in Fig. 6. The “R” value for regression analysis of training, validation, test and all data are 0.65898, 0.6419, 0.54027, and 0.62331, respectively. The correlation factor values of near 0.6 are normal with regards to random climate changing data. The correlation factor value corresponding to test may be considered as 0.54 so poor. Therefore, the ANN proposed model capability to handle such random variation can be accepted as fair, so acceptable. Actually in climatic modeling, generally, such expectation of 0.95 correlation factor may be impossible.

**CONCLUSION**

Generally, modeling of climate pollution, including several irregular trends, does not obey certain mathematical theory. The use of ANN, as a tool for predicting future climatic conditions upon the trends taken from previous measured data, can be the best method to be utilized for modeling. Furthermore, employing certain algorithm or constitutive relations, may be quite helpful in reducing the errors.

The present work indicates that the predicted power of artificial neural network models depends on several vital parameters, namely choice of six inputs data, six parameters in hidden layers, learning algorithm, and types of stopping criteria. One of the prime issues in developing optimal ANN is over-training. It arises when ever network learns the noisy details
in assessing the data. The obtained results indicate that the network with 6 hidden neurons at training epoch 113 has the best performance with the lowest MSE value (MSE=0.049438). This explains that this network prediction is closely matching with actual observation. Further, the “R” value for regression analysis of training, validation, test and all data are 0.65898, 0.6419, 0.54027, and 0.62331, respectively. Accordingly, the ANN-based approach is capable of producing accurate evaluation data set in the field of air pollution.

ACKNOWLEDGEMENT
The authors would like to acknowledge Professor S.A. Sadrnejad, the faculty member of K.N. Toosi University for his valuable supports throughout this research performance.

CONFLICT OF INTEREST
The authors declare that there are no conflicts of interest regarding the publication of this manuscript.

ABBREVIATIONS

- **ANN**: Artificial neural network
- **CO**: Carbon monoxide
- **IA**: Information Assurance
- **MSE**: Mean square error
- **N**: Number of observations
- **NO₂**: Nitrogen dioxide
- **Oi**: Observed value
- **O̅**: Average value
- **P**: Predicted value
- **PM**: Particulate matter
- **PM_{2.5}**: Particles less than or equal to 2.5 micrometers in diameter
- **PM_{10}**: Particles less than or equal to 10 micrometers in diameter
- **SO₂**: Sulfur dioxide

REFERENCES
ANN application for atmosphere PM$_{2.5}$ concentration


AUTHOR (S) BIOSKETCHES
Memarianfard, M., M.Sc., Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran. Email: memarian.m7@gmail.com
Hatami, A.M., Ph.D. Candidate, Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran. Email: amir.m.hatami@gmail.com
Memarianfard, M., Ph.D., Assistant Professor, Department of Civil Engineering, K.N. Toosi University of Technology, Tehran, Iran. Email: memarian@kntu.ac.ir

COPYRIGHTS
Copyright for this article is retained by the author(s), with publication rights granted to the GJESM Journal. This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/).

HOW TO CITE THIS ARTICLE

DOI: 10.22034/gjesm.2017.03.03.010
url: http://gjesm.net/article_23079.html

AAP_database_methods_2014.pdf?
