Evaluating Visibility Range on Air Pollution using NARX Neural Network

Tohid Irani¹, Hamid Amiri²*, Hedieh Deyhim¹

¹ Department of Civil Engineering, School of Civil Engineering Environmental, PNU University of Shiraz, Shiraz, Iran
² Department of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran

Received: 02/11/2020  Accepted: 21/02/2021  Published: 20/06/2021

Abstract
Evaluating air visibility range is considered as one of the apparent criteria of air quality. Haze air as a conclusion of air pollution causes unpleasant breathing, psychological effects, and visibility restriction. In this study, NARX neural network applied to determine air visibility restriction factors. Data of air quality control station of Baghshomal, Rastebazar and Abresan in Tabriz City, Iran used which include PM$_{2.5}$, PM$_{10}$, NO$_2$, SO$_2$, O$_3$ and CO for duration of four year since 2013 to 2017 that considered as independent variables. NARX neural network created to find each pollutant relation to visibility restriction and networks used for simulation to analysis network results in consequtuses condition. The results showed that PM$_{10}$ pollutant has the most influence on-air visibility with $R=0.9$ in the train,$R=0.728$ in test and $R=0.75$ in validation process. Also error results of the PM$_{10}$ obtained as MSE=0.054. Moreover simulation results demonstrated least area integral between curves according to ascending order for six pollutant factors and verified PM$_{10}$ accuracy in NARX network simulation. The total result as study conclusion verified NARX neural network efficiency to evaluate air visibility range while using air pollutant parameters.

Keywords: Visibility range, Air pollution, Air quality index

1 Introduction
Air pollution makes important influences on human's health and life over world cities (1). The increasing trend of air pollution concentration is so dangerous and warning. Continental condition, geographical location, meteorological states, population grow and human activities are the main reasons to produce air pollution. Fossil fuels are an important polluting factor in environment through human activities. Nevertheless the fossil fuels are preferred rather than other fuels in industry operations due to availability and low price. One of major reasons for air pollution relative to vehicle fuels uses in urban traffic and inefficient transportation system (2). These ingredients cause air pollution emission as PM$_{10}$, NO$_x$, SO$_2$, O$_3$ and CO that affects Air quality index (AQI). The Air quality index obtain by gathering air quality station's data analyzes. Air pollution problems arise when the rate of concentrations exceed normal level. State of air pollution pass to more warning condition as the criteria exceeds maximum standard levels. So it requires to analysis and modeling the emission treatment and predicts the condition that will suppose to occur again. The major propors will be reducing risk of air pollution dangerous conditions to normal condition. To proper prediction and analyses modeling for air pollution different approaches such as mathematical emission models (3-4-5), linear models (6-7) and hybrid models (8) has been used and this paper extend time series NARX neural network operation with simulation analysis consist of evaluating of NARX neural network accuracy to determine most influence of pollutant on visibility restriction. In this study EPA air quality index used to determine the states of pollution rate. Air pollution obstacle light radiation through the air and it cause haze air with visibility range restriction especially while it is smoggy day or inversion condition. Haze air impresses psychological effect such as disappointments (9) and respiratory problems. These conditions arise with high AQI rate per hour or day, so demonstrate high emission rate or high stable state for pollution emission. Using neural network for air pollution prediction and modeling are common method because of neural network easement and application that is fast and inexpensive.

Wang and Bai in a study find it efficient and useful to predict AQI using NARX neural networks (10). Also Piroddi and Pisoni used NARX neural network and simulation to predict peak air pollution and results showed efficiency for NARX neural network (11). Johnson, Schmid and Dinakaran in another study used simulink for dynamic air quality modeling for interior spaces in Alaska (12) to predict radon and PM$_{2.5}$ in indoor of homes and it demonstrate high ability of simulation to predict emissions. Ghaffarpasand et al (21) studied the interaction of desert dust and urban pollution in the particulate matter (PM) of an urban area of Isfahan, Iran, where anthropogenic and natural emission sources substantially contributed to the urban emission inventory. Results showed that PM$_{1.1-2.1}$ dominated by the dust

*Corresponding author: Hamid Amiri, Department of Civil and Environmental Engineering, Tarbiat Modares University, Tehran, Iran. E-mail: hamid64amiri@gmail.com
particles, and particles originated by the fuel combustion have a significant contribution to the fraction PM$_{10}$. In another research Ghaffarpasand, Nadi and Davarshahmazari (22) try to apply a new approach to address the data missing issue of the detailed air monitoring dataset of Tehran 2018 (hourly data of over 15 monitoring sites) and to study the short-term effects of anthropogenic/natural activities on the criteria air pollutants (CAP) in the urban and rural areas. Results show that the vehicular emissions were the predominant contributors to the CAP concentrations, accounting for an average of 45%, and desert dust and industries around the city were ranked as the subsequent principal CAP sources for the summer season. In other paper Ghaffarpasand, Beddows, Ropkins and Pope (23) studied reports upon and analyses vehicle emissions measured by the Emissions Detecting and Reporting (EDAR) system, a Vehicle Emissions Remote Sensing System (VERSS) type device, used in five UK based field campaigns in 2016 and 2017. Ghaffarpasand et al (24) in other paper aimed to establish a detailed exhaustive emission inventory of urban motor vehicles in a medium-sized city, i.e. Isfahan. The results showed that EFs of all pollutants emitted from vehicle exhausts usually decrease with vehicle speed. Also, gasoline vehicles have larger CO and VOCs EFs than the others, whereas vehicles fueled by diesel have the largest NO$_x$, SO$_x$, and PM$_{10}$ EFs. Ghaffarpasand (25) research over combined effects of hydrodynamic partial slip, thermal radiation due to solar energy, and nanoparticles volume fraction on natural convection and entropy generation of atmospheric ultra aerosols sample evaluated. Results show that the partial slip usually eliminates the effect of mechanical forces provoked by the moving lids.

This study used NARX neural networks to evaluate air visibility prerequisite pollutant with priorities and using simulation as complement to results and network operation. This paper include four section describe NARX neural network operation to predict visibility range which restricted with air pollution emission and it show pollutant influence on this confinements in a raw from high effect to low effect. Also simulation applied to analysis the visual relation of pollutants and visibility restriction due to conspicuous symmetry and correspondence.

2 Materials and methods

This study tries to analyze visibility restriction major factors and ingredients. To find a guideline show the relations of visibility restriction factors and which circumstance cause that a structure defined as Figure 1 which demonstrate incentives of visibility range restriction metological parameters as rain, wind and temperature are indirect parameters that influence air pollution at first and then confine visibility horizons due to pollutions. So these factors could reduce circumstance in other way to better condition. Figure 1 also depicts inputs and output structure and relations subsequently. According to this structure the inputs considered as pollutant and air quality station emission data's used for each pollutant versus output which was found with if_nested algorithm according to AQI corresponded value.

Nonlinear Auto Regressive with External input (NARX) neural network used to predict air pollution effect on visibility which regarded inputs as air pollution factors and output as visibility range, correlation of inputs and output found on mathlab simulink (13) platform. Also correlation of each input and output depicted as conspicuous chart which show the accuracy of each relation between each input versus haze air or visibility output. For simulation accuracy a sinusoidal signal source used then reaction to network operation depicted as a stepped signal. To define visibility range base on AQI also dedicating high priority amount to everyday haze air condition according to high concentration of pollution factor conditional if_nested algorithm used. In this study procedures repeated for each input parameter then once repeated for whole inputs used together versus one output.

![Visibility range relation structure](image)

**Figure 1: Visibility range relation structure**

2.1 Data

For researching on haze air and visibility, Tabriz cities Baghshomal, Rastebazar and Abresan air quality control stations data used as daily average amount for duration of 1,461 day, since 2013 to 2017 and it includes six parameter of PM$_{2.5}$, PM$_{10}$, NO$_x$, SO$_x$, O$_3$ and CO, as inputs data and air pollution visibility restriction as output. Visibility range determined base on EPA standards and relation to AQI. Inputs of each neural network choice for three value of each pollutant factors belongs to three air quality control station and output data belongs to visibility values base on EPA air quality index average, that defined as maximum concentration.
Each value of average visibility range compared to AQI amounts to define six condition according to six pollutant value, and shown as six circumstance of ‘a’, ‘b’, ‘c’, ‘d’, ‘e’, ‘f’ from clean state with highest visibility range to hazardous state with least visibility range. To define and dedicating air pollution data’s to visibility range according to EPA standard if nested algorithm used. First high value of each pollutant per day defined and evaluated in six class of quality then high class of every pollutant compared to each other to set circumstance that is more polluted, in other word from ‘a’ to ‘f’ it arise the air pollution and restriction on visibility grows. Figure 2 show algorithm used to dedicate true value of visibility range for every day as AQI standard.

2.2 Neural network

Nonlinear Auto Regressive with External input (NARX) neural network used with 1:30 delay. NARX neural network could train to predict time series and according to past weight of time series also it could predict auto regression procedures (11). This network is a dynamic regression network includes layers to processing data it predict future amount by using time series amount (13). Equation (1) is a simple math describe for regressive procedure.

\[ X(t) = f(x(t-1)…x(t-d)) \]  \hspace{1cm} (1)

To train network 50 percent of data used, 25 percent of data for test and 25 percent of data used for validation then correlation amount and mean square error obtained for each input, last time all operation repeated for six inputs together versus one output. This type of neural network as dynamic neural network confront problems of exaggerate error due to seasonal change existing in data matrix. It turns when some data values are out of range in account of systematic or experimental error. To prevent exaggerated error over operation this outlier data must be wiped out.

Seasonal change influence time series long terms operation and reduce accuracy so moving average utilized to remove seasonal change in data with wiping out outlier data and created static form of data. Then corrected data used for time series neural network. Equation (10) used to define average moving operation. This operation remove accidental amount that make high amount of error from series then it provided to use in network.

\[ (x1+x2+…xn) / n \] \hspace{1cm} (2)

The importance of NARX network is that operation depends on answer of past values and output values. This relation shows the rule of feedback that is being use in NARX neural network structure and could improve network accuracy. Actually final output values are not same current values suppose to obtain in feed forward simple neural network indeed it regard feedback values with phase delay, hereupon it create regulated output cycle and improve prediction ability and has more similarity to real nature manner to changes e.g. if current condition relate to past conditions as delay of g, g-1, g-2, g-3 equally it expect to be more regular changes for regular periods and more predictable, in some way it observe in NARX network method.

2.3 Simulation structure

Neural network simulation which used for this study created in matlab simlink. Base on NARX neural network that was created, the settings and rules determined then proceeded to evaluation the operation of network and compare the results with external input as simulation input and receive output as conspectus figure, which include sinusoidal signal as input to network and stepped signal as output. Input and output signal between curves integral compared for each part of conspectus to analyze and compare accuracy of neural network in determining superb parameters which restrict visibility range. To harmonize input and output, signals amplitude and a gain with value of 100 used. Result of input and output merged in scope to compare versus each other. Figure 3 show NARX neural network structure in simulation. This structure show parts of network sections consist of input process section, layers section and output process. For result of simulation, the curves as input and output considered and evaluated area integral between curves used on matlab. Equation (3) describes integral of between curves area for positive and negative part of curves.

\[ \text{Area cycle} = \left| \int (f(x)-g(x)) \, dx \right| + \left| \int (f(x)-g(x)) \, dx \right| \] \hspace{1cm} (3)

![Figure 3: NARX neural network structure in simulation](image)

2.4 Air Quality Index

Standard of emission considered for outdoor and indoor air pollutions in many countries. Air quality index according to EPA used as reference index for this study also there are equations to calculate regional AQI (14-20). Equation (3) used to calculate regional air quality index (15-19).

\[ I_{p} = \frac{(I_{m}-I_{lo})}{(BP_{hi} – BP_{lo})^{*}(CR- BP_{lo})+I_{lo}} \] \hspace{1cm} (4)

Ip mention P pollution factor of AQI, Cp show pollution concentration, BPHi, BPLo respectively shows fracture point for lower and upper point of Cp and IHi, ILo are AQI amount base on BPHi and BPLo. EPA standard of air quality index (Table 2) determine six level of quality which make visibility limitation and need precaution activity for people groups with high risk of health (15-18). There are six condition to air quality state consist of Good, Moderate, Unhealthy for Sensitive Groups, Unhealthy, Very Unhealthy and Hazardous. Which one demonstrates an approximate visibility range base on EPA standard and it simply shows the linear relation between visibility range in pollution and health effect includes psychological and respiratory problems.
Table 1: AQI and visibility range

<table>
<thead>
<tr>
<th>AQI Value</th>
<th>Visibility range</th>
<th>Haze air index</th>
<th>You should take the following precautions:</th>
<th>Health effects may include:</th>
</tr>
</thead>
<tbody>
<tr>
<td>16 km or better</td>
<td>Good</td>
<td>None</td>
<td>None likely. People with preexisting heart and lung diseases may experience some discomfort. Consult with your physician.</td>
<td></td>
</tr>
<tr>
<td>8-16 km or AQI: 51-100</td>
<td>Moderate</td>
<td>None</td>
<td>Aggravation of preexisting heart and lung diseases symptoms are likely.</td>
<td></td>
</tr>
<tr>
<td>4.8-8 km or AQI: 101-150</td>
<td>Unhealthy For Sensitive Groups</td>
<td>People with pre-existing heart and lung diseases the elderly and children should limit outdoor exertion.</td>
<td>Respiratory effects in all population segments likely. Increased aggravation of preexisting heart and lung diseases symptoms. Premature mortality (death) among the most sensitive populations.</td>
<td></td>
</tr>
<tr>
<td>2.4-4.8 km or AQI: 151-200</td>
<td>Unhealthy</td>
<td>People with pre-existing heart and lung diseases the elderly and children should avoid outdoor exertion. All population segments should limit outdoor exertion.</td>
<td>Significant increase of adverse respiratory effects in all population segments likely. Significantly increased aggravation of preexisting heart and lung diseases symptoms. Premature mortality (death) among sensitive populations.</td>
<td></td>
</tr>
<tr>
<td>1.6 km or AQI: 201-300</td>
<td>Very Unhealthy</td>
<td>People with pre-existing heart and lung diseases the elderly and children should consider relocating to a &quot;clean air&quot; area. Everyone else should avoid outdoor exertion.</td>
<td>Serious risk of adverse respiratory effects in the general population. Serious aggravation of preexisting heart and lung diseases symptoms. Increased (death) among sensitive populations.</td>
<td></td>
</tr>
<tr>
<td>Less than 1.6 km or AQI: 301-500</td>
<td>Hazardous</td>
<td>Everyone should avoid any outdoor exertion. Relocate to a &quot;clean air&quot; area if necessary.</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There are some suggestion bases on the standards to prevent sensitive groups confronting high risk of pollution so visibility range make it easy and fast to recognize these conditions also it is valuable to determine reasons of air pollution and visibility restriction to set appropriate solution to decrees unpleasant visibility problem and healthy bothers. Air quality index determine for high value of pollution due to pollutants daily or hourly concentration. Table 1 describes circumstance of AQI base EPA standards for visibility range and its relation to health and pollution. Six state of AQI which regarded as color of green, yellow, orange, red, violet and brown used similarly in visibility range index based on EPA air quality index.

2.5 Visibility and haze air index

To determine visibility range index for this study EPA standard and average amount of ranges used. (16-17-18) Also to demonstrate air pollutant concentration and the visibility range index that mentioned before, if_nested algorithm used to select high value of concentration for six factors per day as haze air index and visibility state.

<table>
<thead>
<tr>
<th>AQI Value</th>
<th>Visibility range</th>
<th>Haze air index</th>
<th>AQI average</th>
</tr>
</thead>
<tbody>
<tr>
<td>More than 10 mile</td>
<td>a</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td>5 to 10 mile</td>
<td>b</td>
<td>50</td>
<td></td>
</tr>
<tr>
<td>3 to 5 mile</td>
<td>c</td>
<td>75</td>
<td></td>
</tr>
<tr>
<td>1/5 to 3 mile</td>
<td>d</td>
<td>125</td>
<td></td>
</tr>
<tr>
<td>1 mile</td>
<td>e</td>
<td>150</td>
<td></td>
</tr>
<tr>
<td>Less than 1 mile</td>
<td>f</td>
<td>175</td>
<td></td>
</tr>
</tbody>
</table>

Firstly pollution state base on high value of AQI per day for each pollutant determined then maximum amount of pollutant for every day choice with if_nested algorithm base on AQI of EPA standards. States defined as letters of ‘a, b, c, d, e, f’ to define state of air pollution relevance to visibility range. Letter of ‘a’ define as condition that the pollution values are good and air is clear and clean but with changing states to ‘f’ it turn to be more polluted and hazy and nominated as most restriction state for visibility range in ‘f’ with high pollution index. Results for if_nested algorithm described as Table 2.

Table 3: Results for NARX operation of pollutants

<table>
<thead>
<tr>
<th>MSE</th>
<th>Train</th>
<th>Test</th>
<th>Validation</th>
<th>Total</th>
<th>Pollution</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.054</td>
<td>0.9</td>
<td>0.728</td>
<td>0.753</td>
<td>0.853</td>
<td>PM_{10}</td>
</tr>
<tr>
<td>0.07</td>
<td>0.85</td>
<td>0.723</td>
<td>0.704</td>
<td>0.81</td>
<td>PM_{2.5}</td>
</tr>
<tr>
<td>0.081</td>
<td>0.846</td>
<td>0.738</td>
<td>0.696</td>
<td>0.81</td>
<td>SO_{2}</td>
</tr>
<tr>
<td>0.088</td>
<td>0.825</td>
<td>0.692</td>
<td>0.705</td>
<td>0.786</td>
<td>NO_{2}</td>
</tr>
<tr>
<td>0.095</td>
<td>0.793</td>
<td>0.69</td>
<td>0.698</td>
<td>0.762</td>
<td>CO</td>
</tr>
<tr>
<td>0.098</td>
<td>0.789</td>
<td>0.656</td>
<td>0.702</td>
<td>0.755</td>
<td>O_{3}</td>
</tr>
<tr>
<td>0.06</td>
<td>0.858</td>
<td>0.741</td>
<td>0.751</td>
<td>0.82</td>
<td>TOTAL</td>
</tr>
</tbody>
</table>

3 Results and Finding

3.1 Neural network results

After normalizing data and rectifying seasonal change NARX neural network created in six states and for each pollutant include PM_{2.5}, PM_{10}, NO_{2}, SO_{2}, O_{3} and CO parameters versus visibility range amount. There are three values for each pollutant which obtained by Baghshomal, Rastebazar and Abresan air quality station data in Tabri City that regarded as input and one output value as described obtained with if_nested algorithm. Then once more for last time we repeat operation for six inputs include average value of each pollutant and one output value. Evaluation of network operation derived by correlation and mean square error analyzes. Table 3 show results of neural network training testing and validation for each pollutant and once for six
pollutants. Each network trained, tested then validated. Evaluation values in every step obtained as well, total amount of correlation and error mean square found for each order per every pollutant. Totally it showed that PM$_{10}$ had more accurate results with R=0.9 and MSE=0.054 and other pollutants effect determined respectively as PM$_{2.5}$, SO$_2$, NO$_2$ at end of list CO and O$_3$ had least effect on haze air. Also results of network operation with six pollutant input and one output showed that average amount of six pollutant effect visibility ranges strongly and the results in compare to PM$_{10}$ as single pollutant, it take second place of accuracy an important.

3.2 Simulation results

Simulation operated after setting structure and using created network for each pollutant and tested by sinusoidal signal as input. Also output regarded as stepped signal which create with trained NARX network as response. Comparison of signals depict a simulation of networks operation which most accurate network show high corresponding of signals. Figure 4 depict trained neural network operation in simulation state. Sinusoidal curve as input and stepped curve as answer output. Ascending and descending condition changes output direction in a parallel mode as answer to input, so curves correspondence like true direction and peak points could show approximate correspondency. The Y axis describes signal amplitude and X axis regard as cycle period.

![Simulation results](image)

Figure 4: Trained NARX neural network simulation chart for each pollutant and total six pollutants
To fit signals amplitude for input and output curves a gain with values of 100 used to fit peak to peak extreme point of signals. Each chart in Figure 5 depict pollutant state that influence visibility range likely to Figure 4. There is a discrete in specification of Figure 5 that integral of curves between areas considered on curves to determine input and output curves and similarity for positive and negative part of cycles. So each output and input signal integral between of two curves created that it used as correspondence value of curves. It supposed to determine high influence of pollutant on visibility range according to least integral of two curves between areas. To find areas between curves limited integrator is a method. In this paper Mathlab functions used as area calculator (13) and ascending order to find areas maximum and minimum amount. Finally areas ascending order used to determine least area integrals to most area integrals subsequently and in increasing trend from low to high area, PM$_{10}$ with least area integral and maximum correlation to visibility range circumstance defined as most important factor that influences visibility range restriction and hazy days.

Table 4: Area ascending order

<table>
<thead>
<tr>
<th>Area ascending order</th>
<th>R</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>PM$_{10}$(least area)</td>
<td>0.9</td>
<td>0.054</td>
</tr>
<tr>
<td>Total</td>
<td>0.85</td>
<td>0.06</td>
</tr>
<tr>
<td>PM$_{2.5}$</td>
<td>0.85</td>
<td>0.07</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>0.84</td>
<td>0.081</td>
</tr>
<tr>
<td>SO$_2$</td>
<td>0.82</td>
<td>0.088</td>
</tr>
<tr>
<td>O$_3$</td>
<td>0.78</td>
<td>0.098</td>
</tr>
<tr>
<td>CO</td>
<td>0.79</td>
<td>0.095</td>
</tr>
</tbody>
</table>
4 Conclusion and Discussions

Results showed that particulate matter with 10 micrometer or PM₁₀ with amount of R=0.9 for training and R=0.728 for test and R=0.753 for validation and least mean square error of MSE=0.052 qualify as high priority factor influencing air pollution, blur and visibility range that is imaginable due to particulate matter exclusivity of light radiation obstacle, these pollutant deviate light radiation especially within 10 micrometer size. More over simulation conспектuses demonstrate most relation of output to input which obtained as least area integral between curves and the results of NARX training, test and validation values confirm it. Network operation on six input order determined as second factor of visibility range restriction with R=0.85 and MSE=0.06 also with low area integral between curves according ascending order results. Then PM₂.₅ placed as third factor influencing visibility range restriction with R=0.85 and MSE=0.07 subsequently NO₂ with R=0.84 and MSE=0.081 and average area integral between curves and close to SO₂ results with average area integral between curves and R=0.81 and MSE=0.088 placed in next factors with moderate influence on visibility restriction. O₃ and CO had least correlation and most error also most area integral between curves according to ascending order results with R=0.78 and MSE=0.098 for O₃ and R=0.79 and MSE=0.095 for CO that showed CO results in network operation is better than simulation area integral between curves. Total results demonstrate high efficiency of NARX neural network in predicting visibility range due to air pollution and it clearly showed that hazardous circumstance of air pollution which confine visibility horizon harshly change with particulate matters in 10 micrometer changes so reducing particulate matters in air could reducing hazy condition to clear and clean condition. Recorded correlation values and mean square error values also simulation results verify efficiency and accuracy of network operation to predict air pollution and visibility range which correspond to Piroddi and Pisoni research results with multilayer NARX that used to predict peak air pollution so simulation showed rising function weight improve efficiency. Also Bai and Wang research with use NARX neural network to predict AQI and results showed efficiency with R=0.98 for training, R=0.97 for test and R=0.973 and MSE=0.002 correspond to this study results. It also correspond to Johnson, Schmid and Dinakaran research results over using simulink for dynamic air quality modeling for interior spaces in Alaska to predict radon and PM₁₀ in indoor of homes and it demonstrate high ability of simulation to predict emission.

Ethical issue

Authors are aware of, and comply with, best practice in publication ethics specifically with regard to authorship (avoidance of guest authorship), dual submission, and manipulation of figures, competing interests and compliance with policies on research ethics. Authors adhere to publication requirements that submitted work is original and has not been published elsewhere in any language.

Competing interests

The authors declare that there is no conflict of interest that would prejudice the impartiality of this scientific work.

Authors’ contribution

All authors of this study have a complete contribution for data collection, data analyses and manuscript writing.

Reference

13. SIMULINK Model-Based and System-Based Design. The MATHWorks Inc 3Apple Hill Drive Natick. MA01760-2098.
16. EPA. Technical Assistance Document for the Reporting of Daily Air Quality the Air Quality Index (AQI);2009.