An Outlook of Ozone Air Pollution Through Comparative Analysis of Artificial Neural Network, Regression, and Sensitivity Models

Jam Shahzaib Khan
Department of Civil Engineering
Quaid e Awam University of Science and
Technology, Nawabshah, Pakistan
jam_shahzaib@hotmail.com

Salim Khoso
Department of Civil Engineering
Quaid e Awam University of Science and
Technology, Nawabshah, Pakistan
engr.salimkhoso@gmail.com

Zafar Iqbal
Faculty of Civil Engineering
Universiti Teknologi Malaysia,
Malaysia
zafar.thalvi@gmail.com

Samiullah Sohu
Department of Civil Engineering
Quaid e Awam University of Science and Technology,
Nawabshah, Pakistan
sohoosamiullah@gmaiol.com

Manthar Ali Keerio
Department of Civil Engineering
Quaid e Awam University of Science and Technology,
Nawabhshah, Pakistan
mantharali99@quest.edu.pk

Abstract—**Air pollution and atmospheric ozone can cause damages to human health and to the environment. This study explores the potential approach of the artificial neural network (ANN) model and compares it with a regression model for predicting ozone concentration using different parameters and functions measured by the Climate Prediction Center of US National Weather Service. In addition, this study has compared the economic viability of ANN and other measuring methods. Results showed that the ANN-based model exhibited better performance. Such model types can be beneficial to government agencies. By predicting ozone concentration government agencies can take preventive measures to avoid significant health effects, protect local populations, and help preserve a sustainable environment.**

Keywords-ozone pollution; environment; sustainability

I. INTRODUCTION

Ozone is a reactive gas, considered a subsidiary pollutant as it is not discharged directly into the atmosphere. Ozone is observed in two different regions of the atmosphere, ground level ozone ("bad ozone") in troposphere and "good ozone" in stratosphere, both with the same chemical composition of O_3. Ozone at stratosphere protects from harmful sun rays and troposphere ozone is the main component of smog [1], caused by automobile emissions, especially in urban areas. Ozone is formed by the chemical processes of oxides of nitrogen (NO_x) and volatile organic compounds (VOCs) in the presence of sunlight. The ozone concentration in urban areas is relatively high compared to rural areas, and it rises in the morning, reaches its peak in the afternoon, and decreases at night time [2]. US Environmental Protection Agency (USEPA) has set standards called air quality index (AQI) for ozone (Table I) [3].

TABLE I.   USEPA AQI STANDARDS FOR OZONE

Health concern levels	Value	Meaning
Good	0 to 50	Air quality is considered satisfactory and air pollution poses little or no risk.
Moderate	51 to 100	Air quality is acceptable, however, for some pollutants there may be moderate health concern for a very small number of people who are unusually sensitive to air pollution.
Unhealthy for sensitive groups	101 to 150	Members of sensitive groups may experience health effects. The general public is not likely to be affected.
Unhealthy	151 to 200	Everyone may begin to experience health effects. Members of sensitive groups may experience more serious health effects.
Very unhealthy	201 to 300	Health warnings of emergency conditions. The entire population is likely to be affected.
Hazardous	301 to 500	Health alert: everyone may experience serious health effects.

The ozone concentration significantly depends upon the atmospheric temperature, UV index, and emissions from different sources. USEPA has defined the overall percentage of the different sources of emission. Figures 1 and 2 show the sources of NO_x and VOC in Mississippi State [4]. Ozone affects lungs, causes asthma and lung cancer [5]. The extent of respiratory illness depends on various factors such as concentration and duration of exposure, climate characteristics, individual sensitivity, preexistent respiratory diseases, and socioeconomic status [6, 7]. This study aims to predict ozone for Jackson, Mississippi. Jackson is the largest urban area in Mississippi with a population of 173,514. Data

were collected from the Climate Prediction Center of the National Weather Service [8], and were used in an ANN to predict ozone and to find out the correlation between some variables. From the daily measured ozone data, the average was taken to observe the trend of the ozone pollution. As shown in Figure 3 a comparison to the AQI shows very unhealthy and hazardous level of health concerns [8].
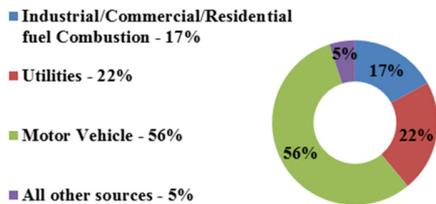
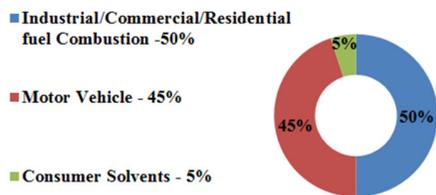

Fig. 1.   NO$_x$ sources in Mississipi.

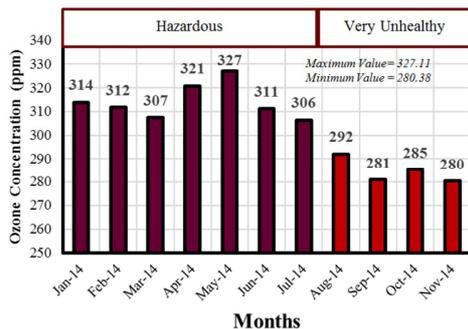Fig. 2.   VOC sources in Mississipi.

Fig. 3.   Average of daily ozone of Jackson, MS-2014

II.   ANN MODEL DEVELOPMENT

An ANN model was developed based on the available data. Initially, ANN architecture was developed and used data from 2010 to 2014. These data were given numbers to identify the intensity of the UV index based on the number of each day of each year. However, some data was missing. Due to the large number of data sets available, it has been decided to remove the missing data sets in order to maintain accuracy. The model architecture was developed in order to work as a computational model using Neuronets [9]. ANNs are often used as a black-box model and show any continuous function to discretional accuracy, where the number of nodes is large [10]. In the model development, dependent and independent categories were found and formed in the required ANN format. Data was also classified as training, testing, and validation sets. Correlation and regression were carried out to determine the correspondence between the variables involved.

In the first phase, the ANN was set to run on training and testing data sets to obtain the required hidden nodes and training iterations for the optimal model. For the next phase, the best network obtained from the previous phase, was verified on the database validation sets. The best performing network obtained from the previous phase has been then retrained on all available patterns in the database in order to account for all information embodied in the database. This retraining provides reliable prediction and better accuracy. Research studies carried out using this approach had shown that the train-all stage is recommended to obtain a better performing model [11].

III.   ANN MODEL ARCHITECTURE

Data considered in this study was taken from the Climate Prediction Center of National Weather Service. Additional data of air operation was taken from Federal Aviation Authority, USA (FAA) [12]. The data was then organized for utilization by the ANN. In this case, the database was set to include seven dependent variables and one independent variable as listed below:

1. Independent Variables:

- Day
- Clear sky UVI
- Cloudy sky UVI
- Cloud transmission "Percent probability clear"
- Solar zenith angle "Percent probability scattered"
- Aerosol transmission "Percent probability broken"
- Air operation data of Jackson, MS

2. Dependent Variable:

- Total ozone concentration

Measured ozone was based on 5 inputs: clear sky UVI, cloudy sky UVI, cloud transmission, solar zenith angle, and aerosol transmission. In this study, air operation (flights operated at Jackson international airport) was also used as an input to identify the impact of air operation on the ozone pollution. The ANN model was developed and selected based on statistical accuracy measures such as maximum values of coefficient of determination (R^2), averaged-squared-error (ASE) and minimum values of mean absolute relative error (MARE). Collectively, 1059 data sets were used for the ANN modeling, namely 479 for training, 232 for testing, 232 for validation, and 116 were kept for true validation. Data was used for training and testing in order to obtain the optimal network. In this case, optimum network was achieved using 2 hidden nodes at 20,000 iterations. Accuracy statistics found on the optimum network were MARE-tr (MARE on training data)=0.329%, MARE-ts (MARE on testing data)=0.318%, R^2-tr=0.997, R^2-ts=0.997, ASE-tr=0.000034 and ASE-ts=0.000032.

Figures 4-7 show the comparison of train, test, validate and train-all modeling stages respectively. These graphs

show the best prediction models obtained for all modeling stages.
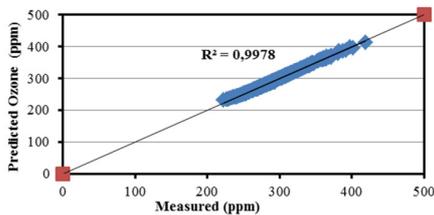

Fig. 4.   Training predicted results versus measured values for ozone concentration.
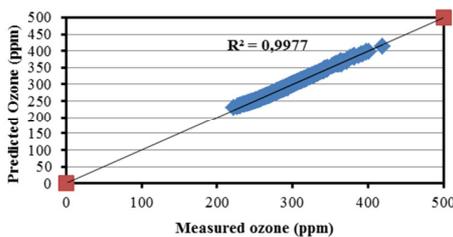

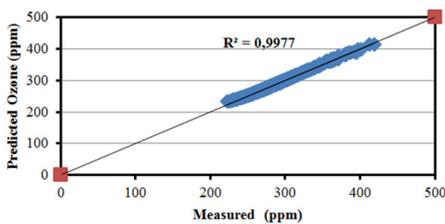Fig. 5.   Testing predicted results versus measured values.

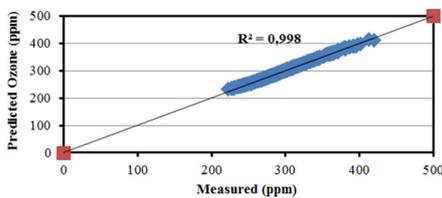Fig. 6.   Validation predicted results versus measured values.

Fig. 7.   Train-all predicted results versus measured values.

IV.   REGRESSION MODEL

Regression model has been developed using Microsoft excel data analysis toolkit. While developing the required linear regression prediction model, all 1059 data sets were used. Same input and output variables with the ANN model were used. Linear regression approach was used, and the following regression equation was developed:

$$Q = (-11125.86) + (-0.03)Day + (-27.77)Clear\ Sky\ UVI \\ + (-2.07)Cloudy\ Sky\ UVI + (0.19)Cloud\ Transmission \\ + (-5.50)Solar\ Zenith\ Angle + (123.87)Aerosol\ Transmission \\ + (0.01)Air\ Operation\ Data \quad (1)$$

From the linear regression model, it can be observed that aerosol transmission has a very high impact in the output. On the other hand, the clear sky UVI is placing moderate impact on output, while the air operation data of Jackson does not have any significant impact. The developed linear regression

model produced a statistical accuracy measure R^2 of 0.79 with standard error of 13.29. In comparison, an R^2 value as 0.998 was obtained via the developed ANN model. This translates to about 25% increase in prediction accuracy when compared to the regression model. The regression model comparison graph is shown in Figure 8. As can be noted, a wider cloud is shown in Figure 8 in contrast to the very thin cloud depicted in Figures 4-7.
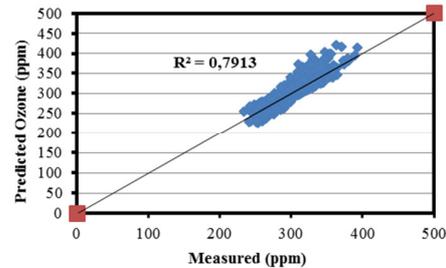

Fig. 8.   Regression model predicted results versus measured values

V.   SENSITIVITY ANALYSIS

Sensitivity analysis is a systematic study of the behavior of the model that reacts over ranges in the variation of inputs and parameters [1]. It can be used to understand whether the predicted model behavior is constant under all parameters or it is changing with respect to change of independent variables. The basic objective of research was to identify the factors affecting ozone concentration and to develop an appropriate model for predicting ozone levels of urban areas of Jackson, Mississippi based on the UVI index data and air operation data. From the correlation matrix, it was observed that some of the independent input variables were highly correlated, namely day, clear sky UVI, cloudy sky UVI, cloud transmission, solar zenith angle. Therefore, sensitivity analysis was not carried out on these variables. Aerosol transmission and air operation were the only two independent input variables found to have the least correlation values. Therefore, sensitivity analysis was carried for these two variables by keeping all other variables constant. Accordingly, ANN output vs. aerosol transmission and air operation sensitivity plots were obtained as shown in Figures 9 and 10, respectively. It can be observed that both variables do not impact on ozone concentration levels.
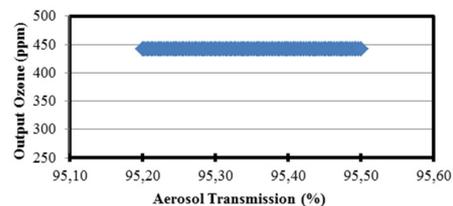

Fig. 9.   Sensitivity analysis of aerosol transmission on ozone concentration.

VI.   RESULTS DISCUSSION

Results obtained from ANN and regression models clearly depict the performance and accuracy of both models. The comparison of the prediction accuracy measures (Table

II) shows that the ANN model excels. Sensitivity analysis indicates that there is no impact found on the ozone levels from prevailing values of aerosol transmission and/or air operation. However, ANN model implicates that the coefficient of determination has apparently the same value. Thus, train, test, validate and train-all stages of ANN also speculate that data is smooth. ASE and MARE show that there is very little error between observed and predicted values.
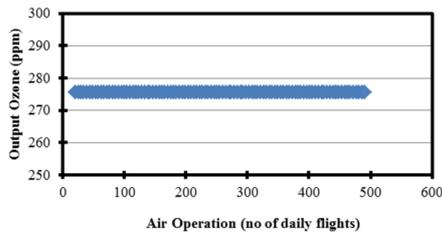

Fig. 10.   Sensitivity analysis of air operation

TABLE II.   ANN RESULTS

Training data	Testing data	Validation data	Train-all data
MARE-tr= 0.329%	MARE-ts= 0.318%	MARE-val= 0.304%	MARE-trall= 0.287%
ASE-tr= 0.000034	ASE-ts= 0.000032	ASE-val= 0.000032	ASE-trall= 0.000029
R^2-tr= 0.9978	R^2-ts=0.9976	R^2-val=0.9977	R^2-trall=0.998

A Microsoft Excel interface (based on developed ANN prediction models) was developed to project the ozone concentration levels from a given set of input variables to facilitate the prediction process (Figure 11).

Fig. 11.   Excel interface application screen shot.

It is worth noting that two ANN models were implemented in the interface where model 1 is based on training, testing and validation stages only, and model 2 is based on the train-all data. The prediction accuracy of model 2 is expected to be higher than that obtained from model 1. However, since both models yield very high prediction accuracy, the difference in this case will be minimal. Both models are included in the interface in order to enable the user to assess the predicted values from two different perspectives. It is worth noting that both models have similar architecture of 2 hidden nodes and similar number of connection weights. However, the values of the connection

weights are different because training data sets for each model are different.

VII.   ECONOMICAL ANALYSIS

There are many ways of measuring ozone in the atmosphere, such as aircrafts, high altitude balloons, satellites, and on ground instruments [13]. Table III shows the relative equipment, cost, expertise and accuracy for atmospheric ozone measurements.

TABLE III.   COMPARISON OF EQUIPMENT FOR OZONE MEASURING [14]

Equipment	Cost (approx.) US $	Training and skills needed	Review
Spectrophotometer			
Dobson spectrophotometer	100,000	Training: Dobson/Brewer for global observations: 2 weeks. Dobson training is available on a space limited basis at SOO-HK. Brewer maintenance requires high technical abilities.	New instruments are not currently available commercially.
Brewer spectrophotometer	100,000		Calibrations are required every 1-2 years and are available commercially.
M-124 Filter ozonometer	15,000		New instruments are not currently available commercially.
Ozonesondes			
One balloon flight package	600 to 800	To prepare and conduct an ozone sounding, one technician with some experience in meteorological instrumentation and training in simple chemical laboratory handling is required. To perform an ozone sounding, pre and post flight operations take about 8 hours. Training for performing an ozone sounding will take from 2 to 4 weeks at a sounding site.	NA
Ground station equipment costs	30,000		NA
Surface Ozone			
UV ozone monitor with data system	15,000	Personnel: Station operator: half a day per week. Training: Operator for station instrument: one week	Various annual costs including inlet filters and instrument repairs when required: about 300/year
Network standard	15,000		

Table III implies that the various methods used to measure ozone air pollution cost much more than the ANN calculating technique. ANNs have been widely used in various sectors and have played a vital role in modernizing computer assisted techniques by handling large data sets and predicting with quite small error.

VIII.   CONCLUSION AND RECOMMENDATIONS

In this study, ANN approach was used to predict, compare and verify ozone concentration at Jackson, Mississippi. ANN-based models were observed to perform

better than linear regression model. This research also illustrated the economic viability of all models for calculating and measuring ozone in the atmosphere. The developed models can be beneficial (via the use of a user friendly interface) to government agencies and other stakeholders. By predicting ozone concentration, government agencies can take preventive measures. This study can also be extended further by including additional input variables such as VMT (vehicle miles travel) and emissions from the industry within the Jackson region that might impact the ozone concentration levels.

REFERENCES

[1] S. Cordiner, R. Baciocchi, M. Attina, "A Sensitivity Analysis of Ozone Formation to Ambient Air Composition by Means of Photochemical Models", Water, Air Soil Pollution: Focus, Vol. 2, No. 5-6, pp. 573–85, 2002

[2] B. J. Bloomer, J. W. Stehr, C. A. Piety, R. J. Salawitch, R. R. Dickerson, "Observed Relationships of Ozone air Pollution with Temperature and Emissions", Geophysical Research Letters, Vol. 36, No. 9, 2009

[3] US Environmental Protection Agency, Air Quality Index Basics, available at: https://airnow.gov/index.cfm?action=aqibasics.aqi (accessed April 12, 2017)

[4] US EPA Office of Air Quality Planning and Standards, National Emissions Inventory 2016, available at: https://www3.epa.gov/cgi-bin/broker?_service=data&_debug=0&_program=dataprog.state_1.sas&pol=NOX&stfips=28 (accessed April 12, 2017)

[5] R. B. Devlin, W. F. McDonnell, R. Mann, S. Becker, D. E. House, D. Schreinemachers, H. S. Koren, "Exposure of Humans to Ambient Levels of Ozone for 6.6 Hours Causes Cellular and Biochemical Changes in the Lung", American Journal of Respiratory Cell and Molecular Biology, Vol. 4, No. 1, pp. 72–81, 1991

[6] M. Kampa, E. Castanas, "Human health effects of air pollution", Environmental Pollution, Vol. 151, No. 2, pp. 362-367, 2007

[7] Environmental Protection Agency (EPA), Ozone Pollution 2016, available at https://www.epa.gov/ozone-pollution (accessed April 12, 2017)

[8] NWS, Climate Prediction Center, UV Index: Annual Time Series, available at: http://www.cpc.ncep.noaa.gov/products/stratosphere/uv_index/uv_annual.shtml (accessed April 12, 2017)

[9] Y. Najjar, X. Zhang, "Characterizing the 3D Stress-Strain Behavior of Sandy Soils: A Neuro-Mechanistic Approach", in: Numerical Methods in Geotechnical Engineering, pp. 43–57, CRC Press, 2000

[10] M. Manngard, J. Kronqvist, J. M. Boling, "Structural learning in artificial neural networks using sparse optimization", Neurocomputing, Vol. 272, pp. 660-667, 2018

[11] H. Yasarer, Y. Najjar, "Development of a mix-design based Rapid Chloride Permeability assessment model using neuronets", 2011 International Joint Conference on Neural Networks, San Jose, USA, July 31- August 5, 201

[12] FAA. Air Traffic Activity System (ATADS): Airport Operations 2014, available at: https://aspm.faa.gov/opsnet/sys/Airport.asp (accessed April 12, 2017)

[13] D. W. Fahey, M. I. Hegglin, "How is ozone measured in the atmosphere?", in: Twenty Questions and Answers About the Ozone Layer: 2010 Update, World Meteorological Organization, 2010

[14] World Meteorological Organization, Global Atmosphere Watch, Global Atmosphere Watch Measurements Guide, GAW Report No. 143, WMO, 2001