



Research Paper

Ozone trend and concentration in Doha City: Time series models versus neural network

Accepted 8th April, 2018

ABSTRACT

This study aimed to investigate the concentration of the Ozone layer in Doha City, compare between different air pollutants and test its relationship with the main meteorological factors. A comprehensive time series analysis using artificial neural network technique was conducted, and appropriate models were determined for future forecast. The bivariate correlation, as well as regression analysis, indicated that there were no significant relationships between the Ozone and other pollutants. On the other hand, the Ozone concentration was significantly related with all meteorological factors. It is concluded that the Ozone concentration is within Qatar standards for air pollution. However, there was a linear trend with a slight increase that needed to be controlled. ANN outperformed time series models such as non-seasonal data ARIMA and Holt's trend models.

Adil Yousif

Department of Math, Stat, and Physics, Qatar
University, Doha, Qatar.

E-mail: adilyousif202@gmail.com.

Key words: Ozone, Qatar, neural network, ARIMA, air quality.

INTRODUCTION

Qatar covers 11,586 km² of land with a population of about 2.5 million (2017). According to world atlas site, the Climate of Qatar during summer through early fall (June - September) are very hot, with temperatures commonly exceeding 120°F (48°C). During the fall months (September - October), weather continues to remain on the hot side, with daily highs averaging in the mid to upper 80's (26°C). From winter through spring (November - May), temperatures can moderate to as low as 62°F (16°C) (<http://www.worldatlas.com/as/qa/where-is-qatar.html>).

The weather question site stated that, "the ozone layer is a deep layer in the stratosphere, encircling the Earth that has large amounts of ozone in it. The layer shields the entire Earth from much of the harmful ultraviolet radiation that comes from the sun. Interestingly, it is also this ultraviolet radiation that forms the ozone in the first place".

According to Fahey et al. (2000), "Ozone is a gas that is naturally present in our atmosphere. Each ozone molecule contains three atoms of oxygen and is denoted chemically as O₃. Ozone is found primarily in two regions of the atmosphere. About 10% of atmospheric ozone is in the troposphere, the region closest to Earth (from the surface

to about 10-16 kilometers (6-10 miles). The remaining ozone (90%) resides in the stratosphere, primarily between the top of the troposphere and about 50 km (31 miles) altitude. The large amount of ozone in the stratosphere is often referred to as the "ozone layer".

Teather (2013) claimed that "Qatar faces a growing risk of health-related problems due to poor air quality originating from both natural and anthropogenic sources. The Environment Statistics Annual Report (2013), recently issued by the Ministry of Development, Planning and Statistics, breaks down Qatar's air quality levels from 2008 until its latest figures in 2012. This is followed by the data released by the World Health Organization earlier this year, which suggested Doha's air is among the most polluted in the world" (Teather, 2013).

The State of Qatar plays a major role in getting rid of the ozone hole by ratified international conventions "it is implementing the Vienna Convention on the Protection of the Ozone Layer and the Montreal Protocol on Substances that Deplete the Ozone Layer" (<http://www.worldatlas.com/webimage/countrys/asia/qatar/qawweather.htm>).

In a study about Doha pollution, David stated that "Reducing the emission of pollutants can reduce bad ozone in the air surrounding humans, plants, and animals. Major sources of pollutants include large cities where fossil fuel consumption and industrial activities are greatest. Many programs across the globe have already been successful in reducing the emission of pollutants that cause excess ozone production near Earth's surface" (David, 2016).

According to Al Ashraf (2008), "Ozone is one of the most important layers among them because it saves us from dangerous Ultra-Violet rays which cause different types of skin diseases and cancer to the human. The ozone layer was discovered in 1917, while the depletion of ozone layer was discovered in 1970. Lanouar et al. (2016), in a study on Doha air quality, added that" air pollution has many adverse impacts on people's health.

Objectives of the study

The aim of this study was to examine the relationship between the Ozone and the other pollutants measured by the ministry of environment in Qatar, test its relationship with the meteorological factors and build a forecasting model to predict the daily average ozone level.

Data for this study

The ministry of Environment in Doha adopted three ways to measure air pollutants; from simple physical and chemical measurements, to sophisticated electronic methods, but they all involve measuring the ambient pollutant concentrations at a given place over a given period of time. Passive sampling is the simplest and cheapest way to screen air quality, and gives a general indication of average pollution concentrations over a time ranging from a week to several months. Passive samplers are so called because they do not involve the pumping of any air; instead, a physical process, such as diffusion, controls the flow. The most commonly used passive samplers are diffusion tubes, and mostly measure nitrogen dioxide and benzene. The tubes (which are 71 mm long with an internal diameter of 11 mm) contain two stainless steel gauze stone end. These contain an absorbent to trap the pollutant to be measured. Thus, in the case of nitrogen dioxide, the absorbent used is triethalamine, which converts the nitrogen dioxide to nitrate, and this is trapped in the steel gauze to be analyzed later in the laboratory. The other end of the tube is left open to the atmosphere, facing down to earth to prevent any rain or dust entering the tube. To ensure that the tubes do not collect pollutant after leaving their site, they are sealed before their journey to the laboratory. The low cost of the tubes enables sampling at a number of points in an area of interest and this can be useful in highlighting 'hotspots' of high concentrations

where more detailed studied may be required. However, recent studies comparing nitrogen dioxide concentrations measured with diffusion tubes and an automatic nitrogen dioxide analyzer found that the diffusion tubes tend to overestimate the levels by about 10%. Other variables which could contribute to this inaccuracy include: errors in the analysis laboratory; the tube location; the tube quality; meteorology. Active sampling methods (which are semi-automatic) use physical or chemical methods to collect pollutant samples. The common method involves pumping a known volume of air through a collector, such as a filter or chemical solution, for a known period of time. The collector is later removed and taken to a laboratory for analysis. Samples can be collected each day, providing measurements for short periods of time at a relatively low cost as compared with automatic methods. Active sampling can also be used for longer periods of time, such as a month. Active sampling methods are used to measure: SO₂, using bubbler samplers; NO₂; PM₁₀, using gravimetric filter samplers; and lead. Using (continuous) automatic sampling methods, pollutants are continually collected and analyzed every hour (or more frequently). The sample is analyzed on-line and in real time and the data are stored within the analyzer, or a separate logger and may be downloaded remotely by a modem. The accuracy produced enables pollution episodes to be analyzed in detail and related to traffic flows, meteorology and other variables. To ensure that the data produced is accurate and reliable, a high standard of maintenance, calibration, operation and quality assurance/control procedures are required, making automatic sampling methods the most expensive way routinely employed. Automatic techniques can be used to measure: ozone; oxides of nitrogen; sulphur dioxide; carbon monoxide; and particulates matter.

The data of this study were obtained using automatic sampling methods. Measurements have been collected at three locations in Doha: "Aspire", "Qatar University" (QU), and "Corniche". The longitude and latitude of these areas are shown on a map (Figure 1). The available data, is recorded hourly and daily for the period from 1st January, 2010 to 30th September, 2016.

Variables in the study

Measurements are available for thirteen pollutants and five types of meteorological data. For time units, we have the time of day and the day of the week (DD/MM/YY/HH). The full list of the variables were divided in two categories: Meteorology: air pressure (AP), temperature (TEMP), wind speed (WS), wind direction (WD), relative humidity (RH); and Pollutant; ozone (O₃), nitric oxide (NO), nitrogen dioxide (NO₂), nitrogen oxides (NO_x), Sulphur dioxide (SO₂), carbon monoxide (CO), total hydrocarbon (THC), methane (CH₄), hydrogen Sulphide (H₂S), non-methane hydrocarbon (NMHC) and particulate matter (PM_{1.0}, PM_{2.5}, PM₁₀).



Figure 1: Map showing the longitude and latitude of three locations in Doha.

METHODOLOGY

Two main methods were used for building a predictive model for the concentration of the ozone in the State of Qatar. Two time series models were found to appropriately fit the data, namely: Holt's Trend corrected models, and none-seasonal Box and Jenkins Autoregressive Moving Average (ARIMA) model. A none-linear technique, which is an Artificial Neural Network (ANN), such as the multiple perceptron layers, was used. The adequacy of the models was checked using MSE and error forecasting of the observed values followed by the project data. Regression analysis, as well as correlation coefficient was used to investigate the relationship between the ozone and other pollutant factors and meteorological covariates. Analysis of variance was also used to test if there is a significant difference in the concentration of the ozone, between the three stations.

Time series techniques

Since the time series plots indicate that the parameters are not constant, then, Holt's Trend Exponential Smoothing Technique is more relevant for forecasting, as presented in Equations (1) and (2):

$$L_t = \alpha \frac{y_t}{s_{t-s}} + (1 - \alpha)(L_{t-1} + b_{t-1}) \quad (1)$$

$$b_t = (L_t + L_{t-1}) + (1 - \beta)_{-1} \quad (2)$$

Equation (3) presents None-seasonal Autoregressive Moving Average (ARIMA) Technique, which was also used as a main linear stochastic model for forecasting:

$$Z_t = \delta + \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (3)$$

Artificial neural network

Artificial Neural Network can offer a valid approximation of a vast class of nonlinear process (Babikir and Mwambi, 2014). Researchers also added that, "the most significant advantage of the ANN models over the classes of linear models is that ANNs are universal approximators that can approximate a large class of functions with high degree of accuracy". In a recent research, it was stated that ANN models either control or at least resist the limitations of traditional forecasting models, including misspecification, biased outliers, and assumption of linearity (Hill et al., 1996). Narasimhan et al. (2000) stated that the advantage of neural systems is that they are trained using historical data and easily can incorporate complex nonlinear codependent variables (such as temperature and relative humidity).

Multilayer Perceptron (MLP) procedure yields a predictive model for one or more dependent (target) variables based on the values of the predictor variables. MLP is a widely used approach in regression-type problems. An MLP network consists of three layers: input layer, hidden layer, and output layer as shown in Figure 2. Neuron sums up the values of inputs parameters according to their assigned weights, and adds a bias. By applying the transfer function, the values of the outputs would be determined. The number of neurons in the input layer is determined based on the number of input parameters (Moghaddam et al., 2016). MLP network is a function of one or more predictors (also referred to as independent variables or inputs) that minimizes the prediction error of one or more target variables (also referred to as outputs):

$$E = \frac{1}{2} \sum_{i=1}^p \|y_t - t_i\|^2 \quad (4)$$

Where y is an output vector and t is a target.

The network model used in this study, has a single hidden layer and contains n nodes. The model has feed-forward functionality. Linear neuron activation function of the network model is presented in Equation (5):

$$y_{t+h} = \alpha_0 + \sum_{j=1}^n w_j g(\alpha_0 + \sum_{i=1}^p w_i y_{t+h}) + \sum_{i=1}^p \beta_i y_{t-i} + \epsilon_{t+h} \quad (5)$$

The transfer function used for this model is hyperbolic tangent function that have a range between -1 to 1 and therefore, the variables entered in MLP must be normalized using the following function:

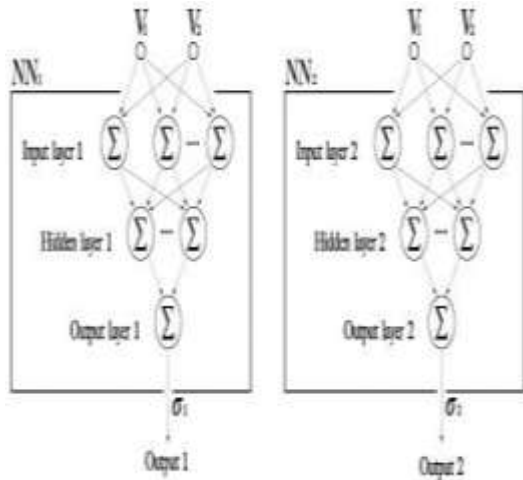


Figure 2: The multilayer perceptron (MLP) procedure.

$$y_{norm} = 1 - 2 \left(\frac{y_{max} - y}{y_{max} - y_{min}} \right) \quad (6)$$

Where y is the observed value.

RESULTS AND DISCUSSION

At the beginning, the data were examined for the missing values and other errors and data-cleaning process was performed. ANOVA was used to check the variability among the three stations and the no significant difference found on the daily readings. Based on the ANOVA results and due to the small area between the stations, the daily averages for the three stations were computed and considered as the daily ozone level for the city of Doha.

From the correlation matrix, the O_3 shows no significant correlation with 12 pollutants and a significant relationship with one pollutant; NO_2 with p-value (0.037). Whereas all metrological factors, such as wind speed, wind direction, temperature, and humidity were significantly related to O_3 with p-value (<0.0001). The descriptive statistics showed maximum daily average level of the O_3 as $59.86 \mu g/m^3$, which is within a good standard (0 – $90 \mu g/m^3$). Qatar University station reported highest levels in most of the reading as compared with other station with maximum daily average of $61.67 \mu g/m^3$.

The time series plot in Figure 3 shows slight increase in the maximum ozone level and the parameters (level and growth rate) are not fixed. The autocorrelation function shows the first difference yields in a stationary set of data since it cuts off more rapidly.

For the Holt's trend exponential smoothing model, the smoothing constants that yield minimum mean square error for the level and trend were 0.2 and 0.1, respectively. For the Box Jenkins model, the autocorrelation function

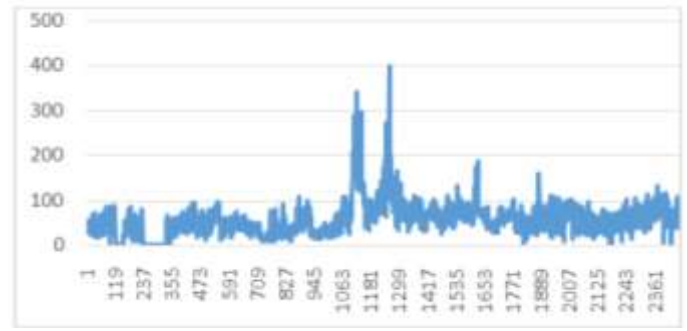


Figure 3: Time Series plot for daily maximum observations.

cuts off after lag three while the partial autocorrelation function dies down quickly for a one difference transformed data, which suggested ARIMA(0,1,3) model with the following equation:

$$y_t = \delta + y_{t-1} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \theta_3 a_{t-3}$$

From the sample data the estimators are:

$$\hat{\delta} = 0.01311, \quad \hat{\theta}_1 = 0.3511, \\ \hat{\theta}_2 = 0.3264, \quad \hat{\theta}_3 = 0.1061$$

With maximum LBQ p-value = 0.014, up to lag 36 and MSE (13.856).

The MLP procedure was based on the O_3 as the dependent variables, the independent variables were minimum temperature, maximum temperature, minimum wind speed, maximum wind speed, humidity as covariate, and wind direction as the factor. The hidden layer activation function was hyperbolic tangent, where the output layer activation function was the identity function. From the model summary, the relative change in the training error criterion was 0.0001, with mean square error was 7.29. The normalized importance of the factors placed the maximum wind as the most important factor, followed by the maximum temperature. The first half of the data (three years) were used for ANN training and the observations of the ozone concentration after the date of the project data were used for model validation.

Since the wind speed is placed as a most significant factor for the level of O_3 , a calendar and wind rose were obtained for ten months in 2016 (Figure 4).

Conclusion

From the Analysis of variance (ANOVA), no significant difference was observed between the Ozone levels in the three stations (Qatar University, Aspire Zone and Al-Cornish station). As a result, the daily average for the three stations was considered as the ozone level for the entire Doha City.

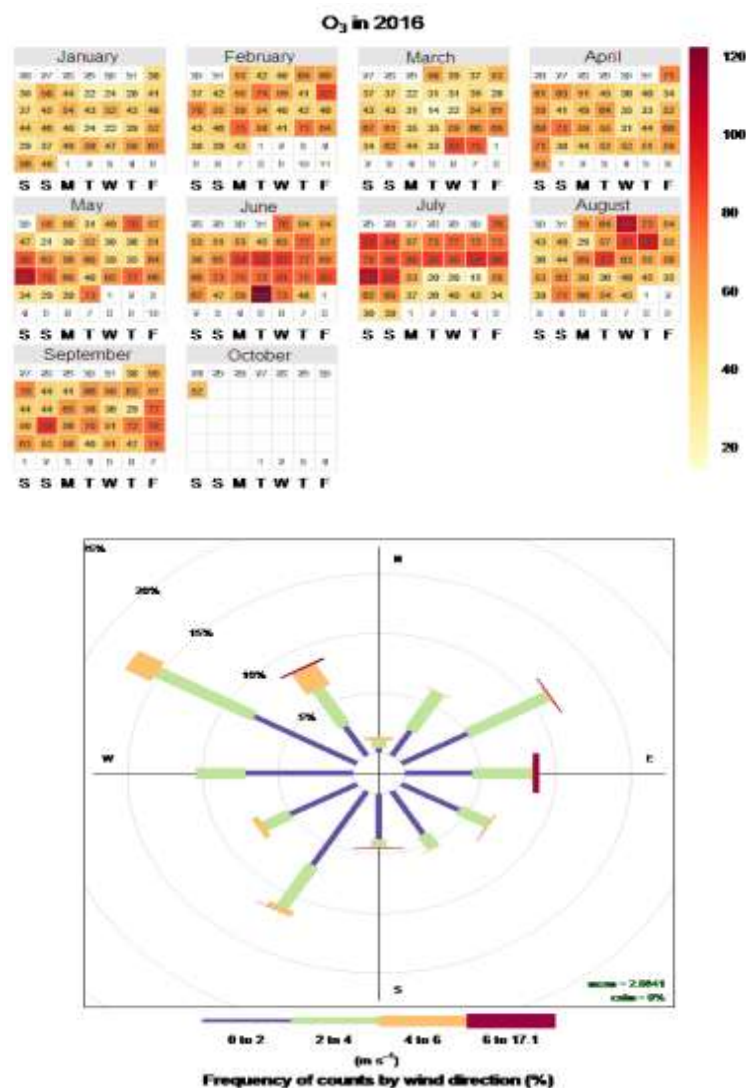


Figure 4: A calendar and wind rose obtained for ten months in 2016.

From the regression analysis of the dependent variable Ozone (O₃), it was shown that NO₂ was a significant factor of O₃, which was included in the model, but PM₁₀, NO, SO₂, CO and THC were not significant. However, from the bivariate correlation between each pollutant variables, the correlation between O₃ and (PM₁₀, NO, NO₂, SO₂, CO, THC) was not significant. Also, the correlation between NO and SO₂ was significant but, between NO and (O₃, PM₁₀, NO₂, C, THC) was not significant. The correlation between NO₂ and (SO₂, TH) was significant, between NO₂ and (O₃, PM₁₀, NO, CO) was not significant. Moreover, the correlation between SO₂ and (NO, NO₂, THC) was significant, but between SO₂ and (O₃, PM₁₀, CO) was not significant. Furthermore, the correlation between CO and (O₃, PM₁₀, NO, NO₂, SO₂, THC) was not significant, between THC and (NO₂, SO₂) was significant but, between THC and (O₃, PM₁₀, NO, CO) was not significant.

From the time series analysis, it was found that the Ozone concentration in Doha had an increasing trend with very low rate, which require caution from authorized personnel.

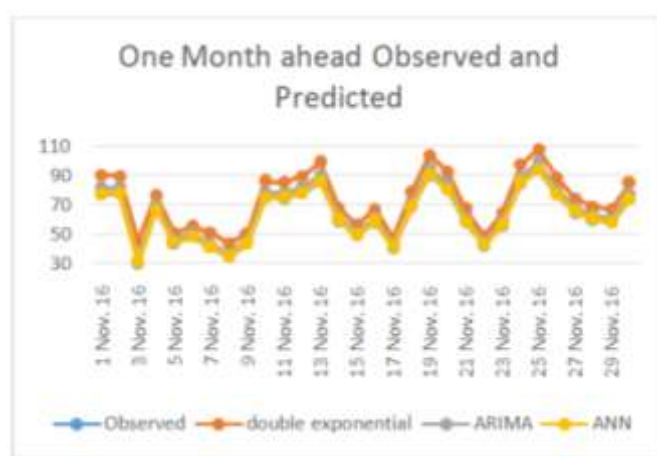
As regard the prediction models, the ANN outperform the time series models with a maximum error equal (0.005), followed by ARIMA model with a maximum forecasting error equal to (0.011). Whereas the maximum error for exponential smoothing model was (0.017) and the Quadratic trend model yielded maximum error of (0.282). Table 1 shows the root mean square error (RMSE) for all models used in this study.

In ARIMA a univariate time series was used were in ANN procedure other covariates were incorporated which may affect the comparison.

The ANN model and the two best time series models (exponential smoothing and ARIMA models) were then tested versus the observed values of the daily average

Table 1: The root mean square error (RMSE) for all models used in this study.

Model	Quadratic trend model	Double exponential model	ARIMA model	ANN model
RMSE	22.519	16.036	13.856	7.29

**Figure 5:** The predicted and observed values for thirty days.

throughout the month of November 2016. Figure 5 shows the predicted and observed values for thirty days.

ACKNOWLEDGEMENT

This study is part of a National Priority Research Program (NPRP) project funded by Qatar foundation, without which it would not be possible to run the research.

REFERENCES

- Al Ashraf A (2008). The depletion of Ozone Layer -what can we do? White Paper. College of Business and Economics, Qatar University. 1(3): 1- 7.
- Babikir A, Mwambi H (2014). A factor—Artificial neural network model for time series forecasting: The case of South Africa. In Neural Networks (IJCNN), 2014 International Joint Conference. IEEE. pp. 838-844
- David C (2016). Package, openair. Tools for the Analysis of Air Pollution Data.
- Fahey DW, Gaol RS, Del Negro LA, Keim ER, Kawa SR, Salawitch RJ, Wennberg PO (2000). Ozone destruction and production rates between spring and autumn in the Arctic stratosphere. Geophys. Res. Lett. 27(17): 2605-2608.
- Hill T, O'Connor M, Remus W (1996). Neural network models for time series forecasts. Manage. Sci. 42(7): 1082-1092.
- <http://www.worldatlas.com/as/qa/where-is-qatar.html>.
- <http://www.worldatlas.com/webimage/countrys/asia/qatar/qaweather.htm>.
- Lanouar C, Yousef Al-Malk A, Al Karbi K (2016). Air Pollution in Qatar: Causes and Challenges; White Paper. College of Business and Economics, Qatar University. 1(3): 1- 7.
- Moghaddam AH, Moghaddam MH, Esfandiyari M (2016). Stock market index prediction using artificial neural network. J. Econ. Financ. Admin. Sci. 21(41): 89-93.
- Narasimhan R, Keller J, Subramaniam G (2000). Ozone modeling using neural networks. J. Appl. Meteorol. 39: 292-296.
- Teather F (2013). Examining the links between air quality, climate change and respiratory health in Qatar, Avicenna. 1(9).

Cite this article as:

Yousif A (2018). Ozone trend and concentration in Doha City: Time series models versus neural network. Acad. J. Environ. Sci. 6(4): 107-112.

Submit your manuscript at:

<http://www.academiapublishing.org/ajes>