Jurnal Kejuruteraan SI 5(2) 2022: 35-45
https://doi.org/10.17576/jkukm-2022-si5(2)-04

A Forecast of Surface Ozone Using Analytical Models

(Peramalan Ozon Permukaan Dengan Menggunakan Model Beranalitik)

Firdaus Mohamad Hamzah [a*,b], Ahmad Nazri Tajul Ariffin[c], Haliza Othman [a,b], Norshariani Abd Rahman[d], Mohd Khairul Amri Kamarudin [e], Mohd Saifullah Rusiman [f] & Siti Hasliza Ahmad Rusmili[c]

[a]Department of Engineering Education, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia
[b]Centre for Engineering Education Research, Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia
[c]Department of Civil Engineering Faculty of Engineering & Built Environment, Universiti Kebangsaan Malaysia, Malaysia
[d]Institute of Islam Hadhari, Universiti Kebangsaan Malaysia, Malaysia
[e]Faculty of Applied Social Sciences, Universiti Sultan Zainal Abidin, Malaysia
[f]Faculty of Applied Sciences and Technology, Universiti Tun Hussien Onn Malaysia, Malaysia

**Corresponding author: fir@ukm.edu.my*

Received 12th June 2022, Received in revised form 1st August 2022
Accepted 1st September 2022 , Available online 15th November 2022

ABSTRACT

In this study, several analytical models were tested to forecast the surface ozone concentration using Artificial Neural Network (ANN), Multiple Linear Regression (MLR) and Time Series Regression (TSR). Four study areas were selected in this study, namely Seberang Jaya in Penang, Shah Alam in Selangor, Larkin in Johor and Kota Bharu in Kelantan. The main objective of this study is to determine the appropriate analytical models MLR and ANN for surface ozone forecasting in some zones of peninsular Malaysia, to forecast surface ozone concentration with TSR model in several zones of peninsular Malaysia and to compare the performance of each model by the performance index. The performance index that will be shown in this study for the model comparison are root mean square error (RMSE), mean square error (MSE) and determination of coefficient (R^2). The ANN model showed better performance compared to the MLR and TSR models in the model comparison in each station. The station in Larkin, Johor provides high accuracy in forecasting surface ozone concentrations for each model with minimum MSE, 0.000009 ppm and RMSE, 0.0042 ppm compared to other stations. The value of R^2 is 0.33 which is highest compared to station in Seberang Jaya and Kota Bharu.

Keywords: Artificial Neural Network; Multiple Linear Regression; Time Series Regression; ozone; performance index

ABSTRAK

Di dalam kajian ini, beberapa model beranalitik diuji bagi meramal kepekatan ozon troposfera dengan menggunakan Rangkaian Saraf Tiruan (ANN), Analisis Regresi Berganda (MLR) dan juga Pemodelan regresi siri masa (TSR). Empat kawasan kajian telah dipilih di dalam kajian ini iaitu Seberang Jaya di Pulau Pinang, Shah Alam di Selangor, Larkin di Johor dan Kota Bharu di Kelantan. Objektif utama dalam kajian ini adalah menentukan model beranalitik iaitu MLR dan ANN yang sesuai untuk peramalan ozon permukaan di beberapa zon semenanjung Malaysia, meramalkan kepekatan ozon permukaan dengan model regresi siri masa di beberapa zon semenanjung Malaysia dan membandingkan prestasi setiap model melalui indeks prestasi. Indeks prestasi yang akan dilihat di dalam kajian ini dalam perbandingan model adalah punca kuasa dua bagi purata ralat (RMSE), purata ralat kuasa dua (MSE) dan pekali penentuan (R^2). Model ANN menunjukkan prestasi yang lebih baik berbanding dengan model MLR dan TSR dalam perbandingan model yang dilakukan. Stesen di Larkin, Johor memberikan ketepatan dalam meramal kepekatan ozon permukaan yang sangat tinggi bagi setiap model dengan nilai MSE iaitu 0.000009 ppm dan RMSE 0.0042 ppm adalah yang minimum berbanding stesen lain. Nilai R^2 pula adalah 0.33 adalah tertinggi berbanding stesen di Seberang Jaya dan Kota Bharu.

Kata kunci: Rangkaian saraf tiruan; Analisis regresi berganda; Regresi siri masa; ozon; indeks prestasi

INTRODUCTION

The formation of the atmosphere is the result from the combination of gases in the air and one of them is ozone. The diatomic molecule of oxygen (O^2) used by humans to breathe on their daily life and makes up almost 20% of the atmosphere (Prasad et al. 2016). There is only 1/3,000,000 oxygen tri-atom molecule content of ozone

(O^3) in atmospheric gases. Ozone is the product of various chemical reactions. However, the main mechanism of production and displacement in the atmosphere is the absorption of ultraviolet (UV) radiation energy from the sun (Wilmouth et al. 2018). The ozone layer is a major compiler located in the stratosphere with the maximum ozone mixing ratio layer occurring at 35 kilometers above sea level (Potdar et al. 2018). According to the Malaysian Meteorological Department, the main function of ozone is to filter the harmful ultraviolet (UV) rays produced by the sun's rays from reaching the earth's surface. If UV radiation is not filtered, it able to penetrate surfaces that protect flora and fauna, and in turn result in severe genetic damage. This will increase exposure to diseases for animals and disease outbreaks to agricultural and forest ecosystems. Increasing UV radiation in aquatic systems also affects nutrient compounds and consequently the maintenance of the photosynthetic process of aquatic life. The absorption of harmful UV radiation by stratospheric ozone is critical to all life on the earth.

Ge et al. (2020) through his study describes emissions from industrial facilities, vehicles exhaust, gasoline vapor, and chemical solvents are examples of major emission sources for pollutants. Yerramilli et al. (2012) and Jud et al. (2020) describe that surface ozone formed from the oxidation process of volatile organic compounds (VOCs) in the presence of nitrogen oxides (NOx) and the intensity of sunlight. (Gómez-Losada et al. 2018) in his study on surface ozone forecasting in urban areas stated that surface ozone concentrations can change rapidly in a matter of hours and days. Analysis and prediction of air quality parameters is an important issue that requires immediate and serious attention, especially in atmospheric and environmental research (Tan et al. 2016). Most large cities in the world has exceeded dangerous levels of ozone concentration and posed a major threat to human health (Ma et al. 2020; Ariyajunya et al. 2021). Given the deleterious effects of increased ozone level on individual health, agricultural production, and city air quality, the current study emphasises the importance of taking targeted steps to reduce the discharge of anthropogenic precursors of surface ozone scales.(Kim et al. 2021; Sharma et al. 2021).

The main factors for ozone elevation levels are temperature increases, photolysis reactions nitrogen oxide, variations in boundary layers as a function of temperature during day and night (Pires et al. 2012). Surface ozone concentrations are seen to be high in summer due to higher photochemical activity. In winter, surface ozone concentrations show lower values (Sharma et al. 2016). Although ozone is formed mainly in urban areas, it is proven that higher concentrations of these pollutants are measured in rural areas, resulting in a reduction in crop yields (Tai et al. 2017).

Other than that, air pollution caused by anthropological activities and natural disasters has been a major challenge to environmental problems over the past few decades (Aghamohammadi et al. 2018). The effects of climate change include warmer air temperatures, rainfall variability, snowfall, increased evaporation, and rising sea levels. Such changes can trigger a variety of dangers including heat, drought and flooding (Liu et al. 2020). Air quality and ozone in the future will change in response to changes in anthropogenic emissions as well as climate change. Concentrations of air pollution and ozone are sensitively linked to climate change because the life cycle of air pollution is influenced by many meteorological factors such as temperature, wind speed, sunlight (Nguyen et al. 2020). These impacts are shown to be in a longer period and will affect human health, infrastructure, forests, agriculture, water supply and marine systems (Nordin et al. 2020).

Air pollution modelling is a very complex task. Several models have been applied for this purpose, which include deterministics and statistics. In addition to numerical models, air quality forecasters often use statistical models derived from local data. These statistical models range from simple regression models to complex artificial neural networks (Pires et al. 2011). In this study, three analytical models were used namely linear and nonlinear models. For the linear model, multiple linear regression modeling (MLR) and time series regression (TSR) was used and for the nonlinear model, artificial neuron network (ANN) modeling was used. The use of analitical models for surface ozone forecasting is gaining ground for the study of ozone. The use of this model is recommended for applications where the dependencies between variables are either unknown or highly complex. The variables used in this study were such as month, temperature, humidity and ozone. The stated variables are also the main focus in this study for predicting surface ozone.

Time series regression (TSR) is used for the study because it has its own set of advantages, such as the ability to forecast ozone concentration using only data from ozone time series and no data from other sources (Fadhilah Abd. Razak et al. 2009). ANN is used as a method for surface ozone prediction in which both functional activation and the number of hidden neurons are solved using genetic algorithms that also optimize threshold values that help distinguish between ozone behaviors (Pires et al. 2012). The epoch value is an important value in obtaining the best architecture and accurate predictions in the ANN model. In the study of Pires et al. (2011) he has used a maximum epoch value of 500. The study of Sousa et al. (2007) used a maximum of 10000 epochs while Taylan (2017) study used a maximum of 3000 epochs in his study. MLR is a popular method used in forecasting (Muhamad, Ul-Saufie & Deni 2015). The results of this study indicate that ANN obtained good results compared to MLR when combined with the correction method. ANN also achieved better results in predicting daily ozone values (Hoshyaripour et al. 2016). On the other hand, machine learning (ML) techniques have emerged and proven to be more effective for ozone prediction (Gong et al. 2016). In a study (Sharma et al. 2016) on surface ozone concentration and its behavior, there were differences in ozone concentration readings during the day and night.

Therefore, the main purpose of this study is to to determine the appropriate statistical model multiple linear regression (MLR) and artificial neural network (ANN) for surface ozone forecasting in some zones of peninsular Malaysia, to predict surface ozone concentration with time series regression (TSR) model in several zones of peninsular Malaysia and to compare the performance of each model by looking at performance index.

METHODOLOGY

STUDY AREA AND DATA SET

The map in Figure 1 shows the location of the station for this study located in peninsular Malaysia. There are four air quality monitoring station which is located in the north, west, south and east peninsular Malaysia. The data used for this study is from year 2010 to 2018. There are two variables used which are the dependent variable and also the independent variable. The dependent variable used in this study was ozone. The independent variables used in this study were month, temperature and humidity.

Table 1 show the specific location of the air quality monitaring station. In the north zone, the station is at Sek. Keb. Seberang Jaya II, Perai, Pulau Pinang. For west zone, the station is located at Sek. Keb. TTDI Jaya, Shah Alam, Selangor. Meanwhile in south zone, the air quality monitoring station is at Larkin, Johor. For the last station in the east, it is located at SMK Tanjung Chat, Kota Bahru, Kelantan. All stations selected are located in the capital of each state and areas with high population and population density. In addition, the selected area is a developed area and has many industrial areas around the air quality monitoring station.

TABLE 1. Location of air quality monitoring station

No.	Station ID	Station	Latitude	Longitude	Zone	Period of data
1.	CA0009	Sek. Keb. Seberang Jaya II, Perai	N5°32’51.9”	E100°40’36.2”	North	2010-2018
2.	CA0025	Sek. Keb. TTDI Jaya, Shah Alam	N3°10’23.2”	E101°33’25.8”	West	2010-2018
3.	CA0019	Larkin, Johor	N1°29’32.1”	E103°14’25.3”	South	2010-2018
4.	CA0022	SMK Tanjung Chat, Kota Bahru	N6°8’49.5”	E102°14’51.4”	East	2010-2018

FIGURE 1. Map of the air quality monitoring station in peninsular Malaysia

ARTIFICIAL NEURAL NETWORK

Artificial neural network (ANN) is an information processing system based on human brain behavior. ANN are also known to be part of a computing system designed to simulate the way the human brain analyzes and processes information (Hafeez et al. 2020). This is the basis of artificial intelligence (AI) by solving problems that may be impossible or difficult by humans (Taylan 2017). The structure of ANN consists of elemental information processing units, called neurons. Neurons will be arranged into several layers and interconnected with each other through synaptic weights. This weighting factor known as "synaptic weight is appropriate in each relationship between neurons. Synaptic weight represents the interaction concentration between each pair of neurons and the activation function calculates the potential of each neuron. The input synaptic weight is a number, which when multiplied by the input gives a weighted input. these weights are then added together and if they exceed a predetermined threshold value, the neurons are activated and give a response result..

The MLP consists of an input layer in which there are artificial neurons corresponding to the input data. After the input layer there is also one or more hidden layers with one or more artificial neurons. Each artificial neuron on each hidden layer is connected and exchanges information with all other neurons from the previous layer and the next layer. Lastly is the output layer, where there are "target" artificial neurons following. The activation function used for hidden layer is hiperbolic tangent while output layer use identity or linear activation function. Information will always flow from the input neuron to the output neuron until the end of an exercise. MLP will be tested with three different layers of hidden number of neurons. The three layer MLP will be implemented using SPSS software. The cross validation process will be done during the training phase to avoid excessive training as well as errors. Equation (1) is an equation to obtain the output value in this model. *W* is the weight, *x* is the input value and *b* is the bias.

$$y = W_2(W_1x + b_1) + b_2 \tag{1}$$

Multiple Linear Regression

MLR or also known as multiple regression describes or predicts one variable from another variable when the two have a linear relationship (Samprit Chatterjee & Hadi 2006). MLR is a statistical technique that uses several independent variables namely month, temperature and humidity to predict the outcome of the dependent variable namely ozone. The goal of MLR is to model the linear relationship between the independent variable and the dependent variable by assembling linear equations with the observed data. In this study, there is one equation used to analyze the data for MLR. Equation (2) used is like the equation where for the observation i = n, yi is the dependent variable (ozone), xi is the independent variable (time, temperature and humidity), β_0 is the y -axis intercept and β_p is the coefficient of the slope.

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} \tag{2}$$

TIME SERIES REGRESSION

Surface ozone forecasting will be carried out at each station using time series regression modeling method. Time series regression is a statistical method for predicting future responses based on the response history of the variable to be predicted. This time series regression can predict the ozone concentration from the average of the data obtained. The time series regression in this study will look at the trend of ozone concentration for each study station . Next, from this average trend of ozone concentration, a model equation will be constructed with linear lines for each study station. The equation to be used for this model is as in equation (3) where Where β_0 is the intercept of the y -axis and β_1 is the coefficient of the slope. While *t* is the time from the year this surface ozone concentration forecast will be implemented

$$\hat{y} = \widehat{B_0} + \widehat{B_1}t \tag{3}$$

MODELS COMPARISON

The model validation process will be performed after optimal output is obtained from training and testing. validation needs to be done to know the accuracy of the developed model as a whole. in this study, the coefficient of determination (R^2) will be used to evaluate the accuracy of the ANN model and the MLR model while mean square error (MSE) and root mean square error (RMSE) will be used to determine which models is more accurate model in predicting surface ozone concentrations. R^2 is the guide used to measure accuracy of the developed model. In this way, accuracy is expressed by values 0 to 1 where value one indicates the output of the developed model is very accurate with the value obtained through experiments conducted in the laboratory. Calculation of R^2 is as shown in Equation (4)

$$R^2 = \frac{\sum_{i=1}^{n} (x-y)^2}{\sum_{i=1}^{n} (x-z)^2} \tag{4}$$

MSE is the average squared error between actual data and forecast data. Equation (5) shows the formula MSE calculations to be used.

$$MSE = \frac{1}{n}\sum_{i=1}^{n} (x-y)^2 \tag{5}$$

RMSE is the square root of the square errors for all the data set. The formula to be used for the RMSE calculation is as in Equation(6).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (x - y)^2} \qquad (6)$$

Where n is the total of data set, x is the average actua data and y is the prediction from models.

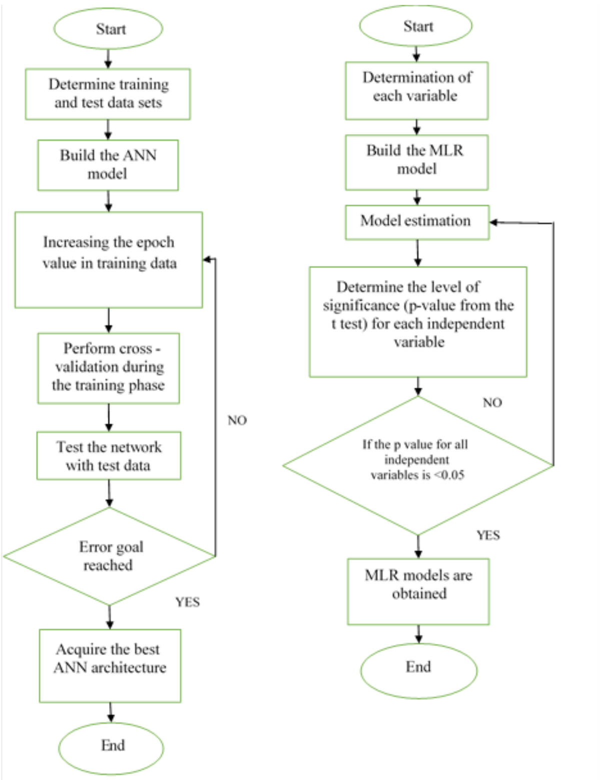


FIGURE 2 Flowchart for ANN and MLR

RESULTS AND DISCUSSION

Monthly average data for each zone from 2010 to 2018 were used to construct this model. A total of 108 data averaged for each variable were analyzed using SPSS software. These data are data from January to December for the year 2010 to 2018.

TABLE 2. Descriptive statistics of the monthly average of surface ozone concentrations

Station	Minimum (ppm)	Maximum (ppm)	Average (ppm)
Seberang Jaya, Pulau Pinang	0.0001	0.119	0.01415
Shah Alam, Selangor	0.0001	0.148	0.02227
Larkin, Johor	0.0001	0.147	0.01620
Kota Bharu, Kelantan	0.0001	0.082	0.01537

Trends ozone concentration vary by region and zone. Table 2 shows the descriptive analysis for each station that was selected in this study. The average monthly reading of ozone concentration at Seberang Jaya Station is 0.01415 ppm and both minimum and maximum readings are 0.0001 ppm and 0.027 ppm. For the Station in Shah Alam, the minimum reading recorded is 0.0001 ppm. The maximum reading is 0.148 ppm and the average ozone concentration is 0.02227 ppm. The station in Larkin, Johor showed a minimum reading of 0.0001 ppm, a maximum reading of 0.147 ppm and an average of 0.0162 ppm. And for the last station which is the Station in Kota Bharu, the minimum reading is 0.0001 ppm, the maximum reading is 0.082 ppm while the monthly average ozone concentration is 0.1537. All minimum, and maximum readings are the result of analysis from data obtained between 2010 to 2018, which is for 9 years. The different reading for ozone concentration in each station is due to dfferent in temperature and humidity which when temperature is high and humidity is low, the ozone concentration will be high.

For the ANN model, a total of three variables which is ozone concentration, temperature and humidity were conducted for the input layer where the ratio for training was 70% and theratio for testing was 30%. This ratio makes the data for January 2010 to February 2016 which has 74 data is data for training. Next the data for March 2016 to December 2018 which has 34 data is as test data. The 7:3 ratio was chosen for each analysis for each station because this ratio is the best ratio for obtaining the best model and forecast. The maximum epochs value applied in this model training is set at 5000 because this value is able to provide accuracy on the prediction with a minimal value of error after several epoch value conversion attempts are performed.

The best result obtained for this model is with 3 hidden neuron in station Seberang Jaya, Shah Alam and Kota Bharu. For station in Larkin, the best model corresponded to the one with 4 hidden neuron. The result of the output value obtained will then be compared with the average of the actual data until the minimum error is obtained by looking at the squared average error found in the table. The maximum epoch set is only if the number of errors achieved is minimal.

The MLR model is determined by a combination of statistically significant regression parameters (Pires et al. 2011) that achieve a minimum number of squared errors. MLR model for each study station that was conducted with the average of the data from 2010 to 2018. Through this model equation shown, the forecast from the MLR model can be calculated from January 2010 to December 2018.

Significant values or p-values should be taken into account with p-values being less than 0.05. Significant values for month, temperature and humidity less than 0.05 indicate these variables influence changes in surface ozone concentration patterns. In this case, the variable for humidity in Seberang Jaya, Shah Alam and also Kota

Bharu has shown a p-value in excess of 0.05. as well as the temperature variable in Kota Bharu which showed a p-value exceeding the value of 0.05, then the null hypothesis was accepted indicating that humidity in Seberang Jaya, Shah Alam and Kota Bharu and for temperature in Kota Bharu did not affect the surface ozone concentration from 2010 to 2018. Once the significant level of p-value is taken into account for modeling, then the MLR model for each station can be generated to make a prediction of the surface ozone concentration.

Apart from looking at p-values that only refer to meteorological parameters such as temperature and humidity, an increase in ozone concentration readings can also influenced by the parameters of other pollutants in the area such as sulfur dioxide, carbon monoxide and also other pollutants. For example, a station in Seberang Jaya showed that humidity did not affect surface ozone concentration readings in this area. This is because high humidity levels are also capable of producing high ozone concentration readings. The presence of other pollutants can affect the surface ozone concentration other than the temperature and humidity which originally if the temperature is high and humidity low, then the ozone concentration readings will increase. The MLR model for Seberang Jaya, Penang (7), Shah Alam, Selangor (8), Larkin, Johor (9) and Kota Bharu, Kelantan (10) station is represented by:

$$O_3 = -0.05452 + 0.000011_{(0.000002)} \text{ (Month)} + 0.0024_{(0.0005)} \text{ x (Temperature)} \quad (7)$$

$$O_3 = 0.0085 - 0.000037_{(0.000015)} \text{ x (Month)} + 0.00052_{(0.00013)} \text{ x (Temperature)} \quad (8)$$

$$O_3 = 0.00845 + 0.00002_{(0.000011)} \text{ x (Month)} + 0.0007_{(0.0003)} \text{ x (Temperature)} - 0.000177_{(0.0006)} \text{ x (Humidity)} \quad (9)$$

$$O_3 = 0.012791 + 0.00004_{(0.000014)} \text{ x (Month)} \quad (10)$$

For this MLR model, the ozone concentration in Seberang Jaya station had a negative correlation with the constant value while it had a positive correlation with month and humidity. For station in Shah Alam, the correlation had negative value for month while it had positive correlation with temperature. Next is station in Larkin, Johor which had a negative correlation for humidity and positive correlation for the month and temperature. For the last station which located in Kota Bharu, Kelantan had a positive correlation for month.

The next model is the time series regression (TSR) model used to predict surface ozone concentrations in the following year. The variable used for this TS model is the average ozone concentration from 2010 to 2018 and also the month which has a total of 108 data. Through the similarity of this model, the forecast for the next year can be known through this TSR model.

Same with the MLR model, the TSR model also needs to take into account a significant value or p-value of less than 0.05. However, the significant values that need to be seen in this model are only for the month variable. In this case, all study stations showed that the p-value for the moon was less than 0.05 and this indicates that the month variable influences in the pattern of surface ozone concentration for each station.

The TSR model for each station is as shown below. This model only uses time series independent variables because the model does not depend onmeteorological factors such as temperature and humidity. This method has its own advantages because the ozone prediction process is performed only by using data from the ozone time series only. The TSR model for station Seberang Jaya, Penang (11), Shah Alam, Selangor (12), Larkin, Johor (13) and Kota Bharu, Kelantan (14) is represented by:

$$O_3 = 0.016873 - 0.000011_{(0.0000021)} \text{ x (Month)} \quad (11)$$

$$O_3 = 0.023681 - 0.000026_{(0.000012)} \text{ x (Month)} \quad (12)$$

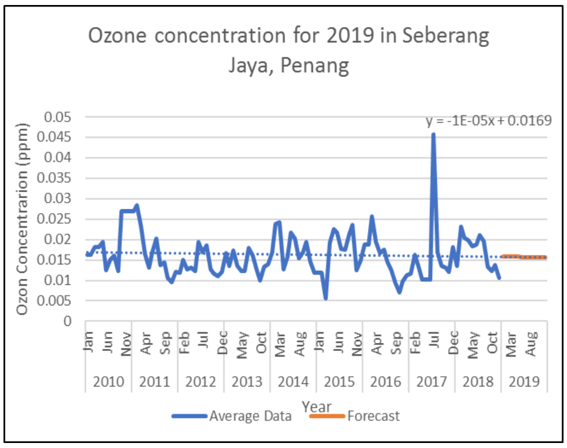
$$O_3 = 0.012926 + 0.000023_{(0.000010)} \text{ x (Month)} \quad (13)$$

$$O_3 = 0.012791 + 0.000040_{(0.000014)} \text{ x (Month)} \quad (14)$$

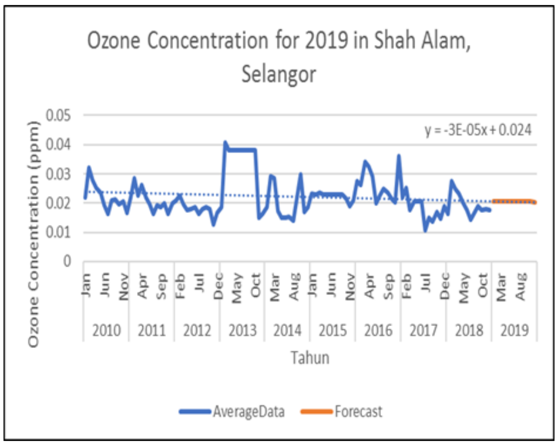
For the TSR model, the ozone concentration in Seberang Jaya, Penang and Shah Alam, Selangor had negative correlation for the month. This shows that the ozone concentration in Seberang Jaya and Shah Alam station is decreasing through year 2010 to 2018. Besides, the ozone concentration in Larkin, Johor and Kota Bharu, Kelantan had a positive correlation with the month which show that the ozone concentration in Larkin and Kota Bharu is increasing through year 2010 to 2019.

Surface ozone forecasting is implemented with a time series regression model in SPSS software. Surface ozone forecasting is done for the year of 2019. This forecast is implemented to find out the surface ozone reading in the following year. The variable taken into account to predict the ozone concentration in the following year is the average ozone concentration and also month. Using these variables, modeling for predicting ozone concentration can be implemented by looking at the linearly trend line in the reading of ozone concentration against the month.

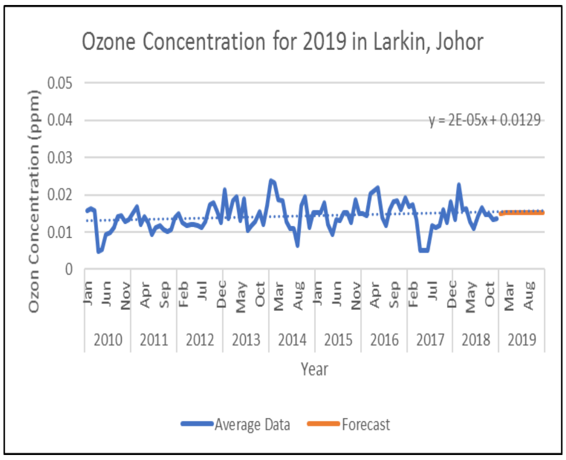
Figure 3 shows the surface ozone forecasting performed in the four study areas. For areas in Seberang Jaya and Shah Alam, the forecast surface ozone concentration is seen to decrease for 2019. This is because the linear line obtained from the average each year is negative which shows a decrease in surface ozone concentration readings from 2010 to 2018. While for the area Larkin and Kota Bharu, the forecast of surface ozone concentration is seen to increase in line with the linear line obtained which is a positive value that shows an increasing trend in surface ozone concentration readings from 2010 to 2018.



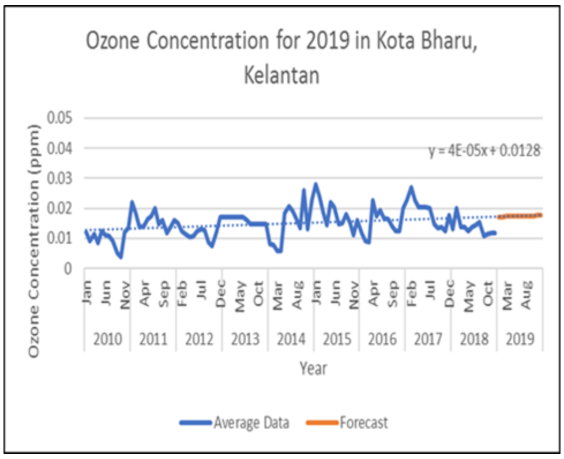
a. Seberang Jaya, Penang



b. Shah Alam, Selangor



c. Larkin, Johor



d. Kota Bharu, Kelantan

FIGURE 3. Forecast ozone concentration using time series regression model

TABLE 3. Performance index of models for each station

	ANN	MLR	TSR
MSE	0.000019	0.000022	0.000025
RMSE	0.0044	0.0046	0.005
R^2	0.17	0.04	0.028

a. Seberang Jaya, Penang

	ANN	MLR	TSR
MSE	0.000018	0.00004	0.000043
RMSE	0.0042	0.0063	0.0066
R^2	0.6	0.10	0.02

b. Shah Alam, Selangor

	ANN	MLR	TSR
MSE	0.000009	0.000014	0.000015
RMSE	0.0031	0.0037	0.0039
R^2	0.33	0.08	0.03

c. Larkin, Johor

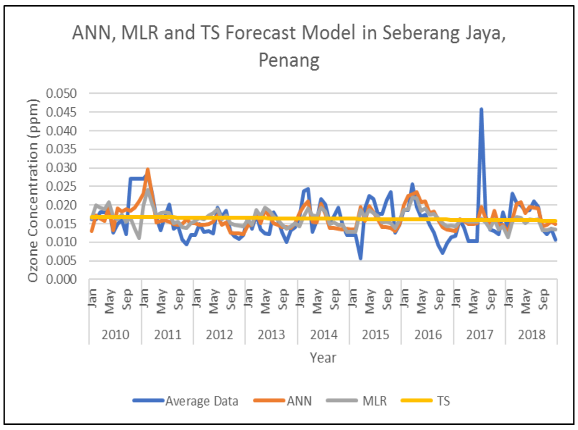
	ANN	MLR	TSR
MSE	0.000035	0.000039	0.000043
RMSE	0.0059	0.0063	0.0066
R^2	0.12	0.02	0.04

d. Kota Bharu, Kelantan

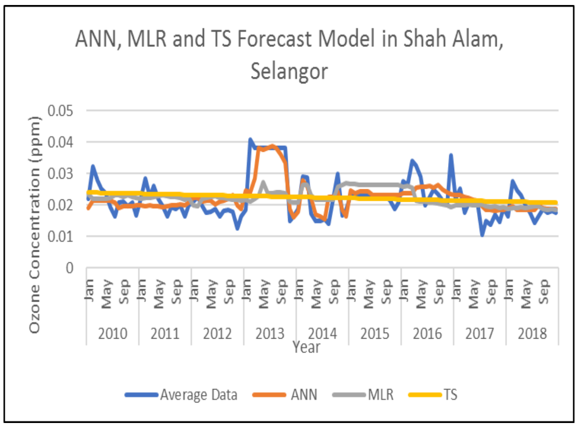
Comparisons between models need to be made to obtain the best model for surface ozone forecasting. This comparison is performed by taking into account the error between the average of the actual data and the predicted values from each model. The errors obtained from the results minus the average of the data as well as the predictions from the model will be applied with the performance index to enable the evaluation of the model to be done. The performance index conducted in this study is the determinant coefficient, (R^2) which is used to evaluate the accuracy on the model. Meanwhile, mean square error (MSE) and root mean square error (RMSE) will be used to determine a more accurate model in predicting surface ozone. Models with small MSE and RMSE values are more accurate models. Table 4 shows the performance indices for each model.

From Table 3, it is found that the ANN model shows small MSE and RMSE values of 0.000019 ppm and 0.0044 ppm at Seberang Jaya station. This proves that the ANN model has accuracy in forecasting at Seberang Jaya station. Similarly, the value of the determinant coefficient which shows the highest value among the three models is 0.17. this indicates that the ANN model is the best model in surface ozone forecasting at Seberang Jaya station, Penang.

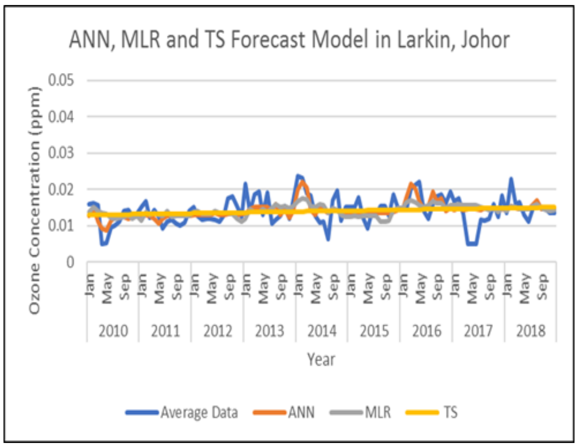
For the Shah Alam station, the ANN model also showed small MSE and RMSE values compared to other models with values of 0.000018 ppm and 0.0042 ppm. As for the determinant coefficient, this model shows a value of 0.61 which makes it the most accurate model for surface ozone forecasting in Shah Alam, Selangor. Comparisons between models at Larkin station, Johor also show that the ANN model is the best with model accuracy having small errors. The MSE and RMSE values for this station are 0.000009 ppm and 0.0031 ppm. As for the determinant coefficient, the ANN model also showed a good accuracy reading with a value of 0.33. For the last station, the station in Kota Bharu, Kelantan, the ANN model showed the best MSE and RMSE values with values of 0.000031 ppm and 0.0056 ppm. The value of the determinant coefficient for the ANN model also showed the best reading compared to the other models with a value of 0.12 and RMSE values with values of 0.000031 ppm and 0.0056 ppm. The value of the determinant coefficient for the ANN model also showed the best reading compared to the other models with a value of 0.12.



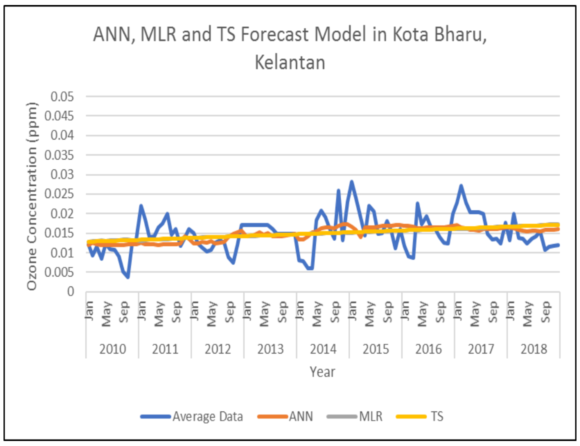
a. Seberang Jaya, Penang



b. Shah Alam Selangor



c. Larkin Johor



d. Kota Bharu, Kelantan

FIGURE 4. Comparison of forecast model in each station.

Figure 4 shows a comparison of model prediction values between the ANN, MLR and TSR models. The error for the ANN model is minimal in all study areas, namely in Seberang Jaya, Shah Alam, Larkin and Kota Bharu. This indicates that the ANN model is the best model because it predicts the surface ozone concentration almost equal to the average value of the actual ozone concentration.

Based on the results obtained in this study, studies involving surface ozone forecasting are very important and serious for the early detection of surface ozone concentration patterns in the future. This can provide an overview to researchers and local authorities to provide attention and planning to curb the increase in surface ozone concentration in one of the areas so that the health of local people is maintained. This is because inhaling and inhaling ozone in the air can cause harmful reactions in the respiratory system (Hamid et al. 2017).

There is no denying that areas that are developing and also have high population densities will lead to various pollutions including ozone pollution. This situation can be seen in this study where the selected study area is an area located in each state capital that is very fast with industrial areas and has a relatively high population density.

Ozone pollution has worsened in recent years, with an increasing trend in its concentration (Ma et al. 2020). This is due to the rapid development and also coming from industrial areas, motor vehicle smoke, greenhouse gases and various other sources of pollution. This pollution itself is also capable of damaging the environment when it is released into the earth's atmosphere (Sharma et al. 2016). Indirectly, temperature increases also occur as a result of the release of anthropogenic sources and photochemical reactions involving these pollutants (Wai et al. 2020).

Malaysia has had rapid economic development and urbanisation, and transportation facilities have led to increased fossil fuel usage in recent decades, resulting in increased air pollution, particularly in industrial districts and cities. As a result, the need of focusing in study on a regression analysis of selected atmospheric factors and surface ozone concentration (Tan et al. 2016).

High ozone levels and air pollution are dangerous for natural ecosystems such as plants, animals and even humans. Therefore, there is a regulatory need to control pollutant emissions to reduce surface ozone concentrations (Ahamad et al. 2020)

Although ozone pollution can be said to occur in urban areas, rural areas or areas with less pollution near urban areas will also be affected directly or indirectly. One of the causes of this phenomenon is the transport of ozone by wind to areas with less industry and pollution where the rate of ozone depletion is less than in urban areas (Rafael et al. 2019).

Ozone is a major component in the troposphere and plays an important role in air quality, atmospheric oxidation capacity and climate change. The results indicate that increased ozone pollution is capable of having a high impact on chronic obstructive pulmonary disease (COPD) (Huang et al. 2018).

Analysis and forecasting for air quality parameters including surface ozone concentration is an important issue that requires immediate and serious attention, especially in atmospheric and environmental research (Tan et al. 2016). This is because it is one of the most important factors that affect the quality of every life.

Therefore, the use of models in the study to predict surface ozone concentrations and other air quality parameters is very important. This is because any accurately predicted increase in ozone limit values allows environmental authorities to implement short-term and long-term pollution control measures as well as ozone concentration reduction strategies to protect the population (Corani et al. 2016).

The ANN model is a model that has good accuracy in terms of predicting surface ozone concentrations. This cannot be denied by referring to the results from this study as well as some results from other researchers such as (Hoshyaripour et al. 2016) who stated that ANN modeling seems to be an appropriate tool to predict the average ozone concentration in a particular location with errors in much lower calculations. Minimal error results allow accurate forecasting results to be obtained.

For MLR and TS models, these two models are linear models that are widely used in the study of surface ozone concentration. In statistical tools, regression analysis is commonly used to analyze data (Muhamad, Ul-Saufie & Deni 2015) which is used to show the relationship between dependent variables and independent variables (Samprit Chatterjee & Hadi 2006). MLR modeling is easier than ANN to model. The MLR and TS models also allow to see whether the forecast pattern of ozone concentration is either decreasing or increasing for an area.

CONCLUSION

In a conclusion, the study that has been carried out is very important and should be taken into consideration from time to time. This is because understanding the concentration of surface ozone is very important to ensure public health is maintained. Through this study, it can provide a prediction in the change of surface ozone concentration either increasing or decreasing in each study area. In this study, all three objectives have been successfully achieved. The use of ANN and MLR analytical models can be used in surface ozone forecasting surface ozone in peninsular Malaysia. In addition, the forecasting of surface ozone using the TSR model shows the trend of forecasting different surface ozone concentrations by region. Next is for model comparison, the ANN model gives better predictive performance than the MLR and TSR models. Overall, the ANN model showed good performance in the prediction of surface ozone concentrations. The station in Larkin, Johor provides accuracy in predicting very high surface ozone concentrations for each model with MSE values of 0.000009 ppm and RMSE 0.0042 ppm being the minimum compared to other stations. The value of R^2 is 0.33 is the highest compared to stations in Seberang Jaya and Kota Bharu.

ACKNOWLEDGEMENT

The authors would like to thank Universiti Kebangsaan Malaysia for their financial support under the grant, GUP-2020-013.

REFERENCES

Aghamohammadi, N. & Isahak, M. 2018. *Climate Change and Air Pollution in Malaysia*. Springer Climate: 241–254.

Ahamad, F., Griffiths, P.T., Latif, M.T., Juneng, L. & Xiang, C.J. 2020. Ozone trends from two decades of ground level observation in Malaysia. *Atmosphere* 11(7): 1–16.

Ariyajunya, B., Chen, Y., Chen, V.C.P., Kim, S.B. & Rosenberger, J. 2021. Addressing state space multicollinearity in solving an ozone pollution dynamic control problem. *European Journal of Operational Research* 289(2): 683–695.

Corani, G. & Scanagatta, M. 2016. Air pollution prediction via multi-label classification. *Environmental Modelling and Software* 80(September): 259–264.

Fadhilah Abd. Razak, Mahendran Shitan, Dan, A. H. H. & Abidin, I. Z. 2009. Load forecasting using time series models. *Jurnal Kejuruteraan* 21(53-62.

Ge, S., Wang, S., Zhang, J., Xu, Q. & Ho, T. 2020. Ozone pollution control strategies examined by Empirical Kinetics Modeling Approach over the Beaumont-Port Arthur region in Texas of USA. Atmospheric Pollution Research(September): 1–10.

Gómez-Losada, Á. 2018. Forecasting ozone threshold exceedances in urban background areas using supervised classification and easy-access information. *Atmospheric Pollution Research* 9(6): 1052–1061.

Gong, B. & Ordieres-Meré, J. 2016. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. *Environmental Modelling and Software* 84: 290–303.

Hafeez, A., Ammar Taqvi, S.A., Fazal, T., Javed, F., Khan, Z., Amjad, U.S., Bokhari, A., Shehzad, N., Rashid, N., Rehman, S. & Rehman, F. 2020. Optimization on cleaner intensification of ozone production using Artificial Neural Network and Response Surface Methodology: Parametric and comparative study. *Journal of Cleaner Production* 252: 119833.

Hoshyaripour, G., Brasseur, G., Andrade, M.F., Gavidia-Calderón, M., Bouarar, I. & Ynoue, R.Y. 2016. Prediction of ground-level ozone concentration in São Paulo, Brazil: Deterministic versus statistic models. *Atmospheric Environment* 145: 365–375.

Huang, J., Li, G., Xu, G., Qian, X., Zhao, Y., Pan, X., Huang, J., Cen, Z., Liu, Q., He, T. & Guo, X. 2018. The burden of ozone pollution on years of life lost from chronic obstructive pulmonary disease in a city of Yangtze River Delta, China. *Environmental Pollution* 242: 1266–1273.

Jud, M.A., Juneng, L., Tangang, F.T., Latif, M.T., Chung, J.X. & Ahamad, F. 2020. Madden Julian oscillation modulation for surface ozone in Peninsular Malaysia. *Atmospheric Environment* 233(April): 117577.

Kim, H. C., Lee, D., Ngan, F., Kim, B.-U., Kim, S., Bae, C. & Yoon, J.-H. 2021. Synoptic Weather and Surface Ozone Concentration in South Korea. *Atmospheric Environment* 244(117985.

Liu, Z., Herman, J.D., Huang, G., Kadir, T. & Dahlke, H.E. 2020. Identifying climate change impacts on surface water supply in the southern Central Valley, California. *Science of the Total Environment*(xxxx): 143429.

Ma, Y., Ma, B., Jiao, H., Zhang, Y., Xin, J. & Yu, Z. 2020. An analysis of the effects of weather and air pollution on tropospheric ozone using a generalized additive model in Western China: Lanzhou, Gansu. *Atmospheric Environment* 224(2): 117342.

Muhamad, M., Ul-Saufie, A.Z. & Deni, S.M. 2015. Three days ahead prediction of daily 12 hour ozone (O3) concentrations for urban area in malaysia. *Journal of Environmental Science and Technology* 8(3): 102–112.

Nguyen, G.T.H., Shimadera, H., Uranishi, K., Matsuo, T. & Kondo, A. 2020. Numerical assessment of PM2.5 and O3 air quality in Continental Southeast Asia: Impacts of future projected anthropogenic emission change and its impacts in combination with potential future climate change impacts. *Atmospheric Environment* 226(March): 117398.

Nordin, N.H., Kaida, N., Othman, N.A., Akhir, F.N.M. & Hara, H. 2020. Reducing Food Waste: Strategies for Household Waste Management to Minimize the Impact of Climate Change and Contribute to Malaysia's Sustainable Development. *IOP Conference Series: Earth and Environmental Science* 479(1).

Pires, J.C.M., Alvim-Ferraz, M.C.M. & Martins, F.G. 2012. Surface ozone behaviour at rural sites in *Portugal. Atmospheric Research* 104–105: 164–171.

Potdar, S.S., Nade, D.P., Pawar, R.P., Jeni Victor, N., Nikte, S.S., Chavan, G.A., Taori, A. & Siingh, D. 2018. Statistical analysis of total column ozone during three recent solar cycles over India. *Journal of Atmospheric and Solar-Terrestrial Physics* 181(October): 44–54.

Rafael, C.C., Javier, G.G., Ariza-Villaverde, A.B., Gutiérrez de Ravé, E. & Jiménez-Hornero, F.J. 2019. Can complex networks describe the urban and rural tropospheric O3 dynamics? *Chemosphere* 230: 59–66.

Samprit Chatterjee, & Hadi, A. S. 2006. Regression analysis by example. Wiley-Interscience.https://www.wiley.com/en.us/Regression+Analysis+by+Example%2C+4th+Edition-p-9780470055458

Sharma, S., Sharma, P., Khare, M. & Kwatra, S. 2016. Statistical behavior of ozone in urban environment. Sustainable Environment Research 26(3): 142–148.

Sharma, R., Sachdeva, K. & Sharma, A. R. 2021. Variations of Surface Ozone Levels in Urban Area of India: A Focus on Night-Time Residual Concentrations. *EQA-International Journal of Environmental Quality* 42 (22-35).

Sousa, S.I.V., Martins, F.G., Alvim-Ferraz, M.C.M. & Pereira, M.C. 2007. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environmental Modelling and Software* 22(1): 97–103.

Tai, A.P.K. & Val Martin, M. 2017. Impacts of ozone air pollution and temperature extremes on crop yields: Spatial variability, adaptation and implications for future food security. *Atmospheric Environment* 169: 11–21.

Tan, K.C., San Lim, H. & Jafri, M.Z.M. 2016. Prediction of column ozone concentrations using multiple regression analysis and principal component analysis techniques: A case study in peninsular Malaysia. *Atmospheric Pollution Research* 7(3): 533–546.

Taylan, O. 2017. Modelling and analysis of ozone concentration by artificial intelligent techniques for estimating air quality. *Atmospheric Environment* 150: 356–365.

Wai, K.M., Tan, T.Z., Morakinyo, T.E., Chan, T.C. & Lai, A. 2020. Reduced effectiveness of tree planting on micro-climate cooling due to ozone pollution—A modeling study. *Sustainable Cities and Society* 52(July 2019): 101803.

Wilmouth, D.M., Salawitch, R.J. & Canty, T.P. 2018. *Stratospheric Ozone Depletion and Recovery*. Green Chemistry: An Inclusive Approach, hlm.

Yerramilli, A., Srinivas Challa, V., Rao Dodla, V.B., Myles, L.T., Pendergrass, W.R., Vogel, C.A., Tuluri, F., Baham, J.M., Hughes, R., Patrick, C., Young, J. & Swanier, S. 2012. Simulation of surface ozone pollution in the Central Gulf Coast region during summer synoptic condition using WRF/Chem air quality model. *Atmospheric Pollution Research* 3(1): 55–71.