EnvironmentAsia
The international journal published by the Thai Society of Higher Education Institutes on Environment

Available online at www.tshe.org/ea/index.html
EnvironmentAsia 10(1) (2017) 118-129
DOI

Analysis of Daytime and Nighttime Ground Level Ozone Concentrations Using Boosted Regression Tree Technique

Noor Zaitun Yahaya [a], Nurul Adyani Ghazali [a], Sabri Ahmad [b], Mohammad Akmal Mohammad Asri [c], Zul Fahdli Ibrahim [c] and Nor Azam Ramli [d]

[a] *School of Ocean Engineering, Universiti Malaysia Terengganu, 21030, Kuala Terengganu, Malaysia*
[b] *School of Mathematic and Informatic, Universiti Malaysia Terengganu, 21030, Kuala Terengganu, Malaysia*
[c] *Universiti Malaysia Terengganu, Terengganu, Malaysia*
[d] *School of Civil Engineering, Universiti Sains Malaysia, Penang, Malaysia*

Abstract

This paper investigated the use of boosted regression trees (BRTs) to draw an inference about daytime and nighttime ozone formation in a coastal environment. Hourly ground-level ozone data for a full calendar year in 2010 were obtained from the Kemaman (CA 002) air quality monitoring station. A BRT model was developed using hourly ozone data as a response variable and nitric oxide (NO), Nitrogen Dioxide (NO_2) and Nitrogen Dioxide (NO_x) and meteorological parameters as explanatory variables. The ozone BRT algorithm model was constructed from multiple regression models, and the 'best iteration' of BRT model was performed by optimizing prediction performance. Sensitivity testing of the BRT model was conducted to determine the best parameters and good explanatory variables. Using the number of trees between 2,500-3,500, learning rate of 0.01, and interaction depth of 5 were found to be the best setting for developing the ozone boosting model. The performance of the O_3 boosting models were assessed, and the fraction of predictions within two factor (FAC2), coefficient of determination (R^2) and the index of agreement (IOA) of the model developed for day and nighttime are 0.93, 0.69 and 0.73 for daytime and 0.79, 0.55 and 0.69 for nighttime respectively. Results showed that the model developed was within the acceptable range and could be used to understand ozone formation and identify potential sources of ozone for estimating O_3 concentrations during daytime and nighttime. Results indicated that the wind speed, wind direction, relative humidity, and temperature were the most dominant variables in terms of influencing ozone formation. Finally, empirical evidence of the production of a high ozone level by wind blowing from coastal areas towards the interior region, especially from industrial areas, was obtained.

Keywords: stochastic; algorithm; ozone; R software; interactions

1. Introduction

Ozone plays a major role in oxidation processes and radiation transfer in the atmosphere. Exposure to ambient ground-level ozone affects the human respiratory system, especially the functioning of the lung. The exposure to high concentrations of ground-level ozone has been shown to reduce physical performance because the increased ventilation rate during physical exercise increases the effects of ground-level ozone exposure. Moreover, ground-level ozone may adversely affect the respiratory airways, rendering it more responsive to other inhaled toxic substances and bacteria. Repeated long-term exposure to ozone has the potential to affect human health and cause irreversible damage to the lungs. The effect of ozone in various countries around the world is becoming crucial especially in countries whose economies are dependent on agriculture, because ozone and its precursors have the potential to adversely affect crop yields. In particular, high ground-level-ozone exposure affects slow-growing crops and long-lived trees. According to Heck *et al*. (1988), ozone can cause considerable yield losses in sensitive crop species by contaminating water and nutrient supplies.

In Malaysia, the Malaysia Ambient Air Quality Guidelines for ozone is 0.10 ppm for hourly and 0.06 ppm for eight hour average. It was reported that ground-level ozone was found to be one of the major air pollutants and the annual average daily maximum one-hour O_3 were increased by 0.5 % compared to 2010 (Department of Environment Malaysia, 2011). In addition to PM_{10}, it has been reported that urban areas recorded higher levels of ozone due to higher traffic volume and a conductive atmospheric condition resulting in it's formation (Department of Environment Malaysia, 2011). These conditions resulted and recorded an unhealthy atmospheric condition at various locations in the Klang Valley and in Negeri Sembilan, Perak, Kedah, and Johor (Department of Environment Malaysia, 2011).

O_3 results from complex chemical reactions when the primary pollutants interact under the action of sunlight. The complex photochemical formation of this secondary pollutant is regulated by both natural and anthropogenic emissions and also by the meteorological conditions (Abdul-Wahab and Al-Alawi, 2002; Sadanaga *et al*., 2003). The formation of ozone in the upper atmosphere can be explained as a chemical process involving radiant energy (hυ) from the sun. Certain wavelengths in the ultraviolet range are able to break oxygen molecule (O_2) into monoatomic (reactive) oxygen atoms, O. When an O_2 receives a photon (hυ), it dissociates into O. These atoms can combine with an O_2 to form ozone (Finlayson-Pitts and Pitts, 2000; Seinfeld and Pandis, 2006). At high altitudes (above 20 km), O are produced by photolysis of O_2 by absorption of deep ultraviolet radiation. At lower altitudes, where radiation is no longer than 280 nm, the only source of O is the NO_2 photolysis. A reaction that converts NO into NO_2 without consuming an ozone molecule could make ozone accumulate (Teixeira *et al*., 2009). This reaction occurs in the presence of hydrocarbons. In particular, peroxy radicals (RO_2) produced during the oxidation of hydrocarbon molecules react with NO to form NO_2, leading to an increased ozone production. This is where VOCs participate in O_3 formation. VOCs play a central role in processes by which 'free radicals' convert NO into NO_2 without destroying O_3. Reaction (1) indicates how NO can be converted to NO_2 without losing O_3 in the process.

$$RO_2 + NO \longrightarrow NO_2 + RO \quad (1)$$

* R is hydrocarbons (a class of VOCs)

$$NO_2 + h\nu \longrightarrow NO + O \quad (2)$$

$$O + O_2 + M \longrightarrow O_3 + M \quad (3)$$

Net Process:

$$RO_2 + O_2 + h\nu \longrightarrow RO + O_3 \quad (4)$$

In Malaysia, Ghazali *et al.*, (2010) reported that O_3 exhibited strong day-today variations. The report also indicated that the presence of sunlight, O_3 was performed by photochemical reactions, which involved its precursors and UV radiations. That lead to high level of O_3 concentration were observed during noontime coinsided with maximal UV radiation recorded (Ghazali *et al*., 2010).

Awang *et al*. (2015) study the concentration trends of O_3 during daytime and nighttime at several stations in Malaysia. It was reported that higher ozone level was observed at nighttime at a Kemaman station comparing to other stations across Malaysia. Studies conducted by Hsieh-Hung *et al*. (2008) in Taiwan indicated that sea-land breeze may transport air pollutants from coastal polluted areas to inland and spread O_3 to an offshore distance of 20-30 km resulting in relatively high O_3 concentrations at an offshore site. Field measurement results showed that the highest O_3 concentrations during daytime were observed at inland sites and that the O_3 concentrations were relatively higher at offshore sites during nighttime. The study also concluded that the accumulation of O_3 in the near-ocean region due to sea-land breeze influenced the spatial and temporal distribution of O_3 in the coastal region of Southern Taiwan. The influence of sea breeze from sea to land during daytime and land breeze from land to sea during night time can transported sea salt aerosol from sea to land at day time and bring back new chemical-based of sea salt aerosol to ocean at night. These chemical mixture of ozone precursor such as NO_x and sea salt aerosol can produce nitryl chloride at night which subsequently produces ozone the next sunrise (Wallace and Hobbs, 2006)

A wide range of methods have been used to study ozone formation in the atmospheric environment. Several researchers studying O_3 have focused on trend analysis and ozone prediction (Huang and Smith, 1999). More recently, Gardner and Dorling (2000) conducted a study in Norwich, UK, in which a regression tree was used for predicting surface ozone concentrations. The latest ozone study that used the CART method was one involving the daily forecasting of the extent by which the ozone levels would exceed the standards set by Italian law (Bruno *et al*., 2004). Bruno *et al*. (2004) used the CART based on the episode selection method for the air quality modelling of ozone and for performing integrated control strategy analysis. Studies by Robeson and Steyn (1990), Chaloulakou *et al*. (1999), Hubbard and Cobourn (1998), Gardner and Dorling (2000) and Wang *et al*. (2003) used a neural network to predict the daily maximal O_3 in selected areas and Barrero *et al*. (2006) used linear and nonlinear regression models to predict O_3 concentrations. In addition, Abdul-Wahab *et al*. (2005) and Hassanzadeh *et al*. (2008) used a combination of multiple linear regression (MLR) and principal component analysis to predict O_3 concentrations in Kuwait and Iran, respectively. More recently, study by Ghazali *et al*. (2010) employed statistical regression techniques to predict ground-level ozone in Malaysia. Although multiple regression based ozone prediction models have been developed, the demand for more accurate models continues to exist (Ghazali *et al*., 2009; 2010). More recently, Munir *et al*. (2015) analysed ground level of ozone in the arid climate of Makkah and applying k-means algorithm to the one calendar year of ozone data. It was reported that ozone

levels was found maximum in September when the temperature levels are relatively low and not in June and July when temperature levels are too high. This similar with report by Steiner *et al*. (2010) which indicated that chemical and biophysical feedback have an influenced to the formation of ozone in the extreme temperature.

Recent research has demonstrated that a Boosted Regression Trees (BRT) model with stochasticity can be used to obtain the best model that can deal with a high level of complexity and large datasets and yield substantial outcomes (Yahaya *et al*., 2011a; Yahaya *et al*., 2011b; Yahaya, 2013). Boosting is a general method that can be used to 'boost' the model accuracy of any given learning algorithm and was first developed by Friedman (2001) and then adding the stochastic element in the algorithm in 2002 (Friedman, 2002). The difference in the stochastic BRT algorithm is that at each iteration, a subsample of the training data is randomly extracted from the full training data set without being replaced. Regression tree models have been used in a few fields such as ecology (De'ath *et al*., 2007; Leathwick *et al*., 2006; Elith *et al*., 2008), social sciences (Kriegler, 2007), and remote sensing (Lawrence *et al*., 2004). BRT were applied by Carslaw and Taylor (2009) in the context of air pollution (NO_x). Their study examined the use of BRT to understand the source characteristics of NO_x at a location of high source complexity at London's Heathrow airport. More recently, extensive work was performed by Yahaya, *et al*., (2011b) and Yahaya, (2013) in City of Leeds, UK, and these studies appear to be the first to have examined and used BRTs for analysing air pollution data, traffic data, and meteorological conditions intensively.

The BRT approach laid the foundation for a new generation of boosting algorithms, called stochastic boosted regression trees (SBRT) which controlled through a 'bag fraction' that specifies the proportion of data to be selected at each step. The default bag fraction is 0.5, implying that in each iteration, 50% of the data are randomly drawn from the full training set without being replaced. Five tuning parameters should be controlled (in addition to the distribution): the training sample size relative to the training population (*bagfraction*), number of iterations (*nr*), learning rate (*lr*), maximal tree depth (*interaction depth*), and number of observations in each terminal node for determining the optimal number of iterations that should be tested for different predictors. This research collated a high volume of simultaneous measurements of ozone and meteorological parameters and applied new methods of ozone analysis involving BRTs by using R software (R Development core Team, 2008) and R packages.

2. Materials and Methods

2.1 Site description and data used

Hourly ground-level ozone data of O_3, Nitrogen Oxides (NO_x) in the form of Nitric Oxide (NO) and Nitrogen Dioxide (NO_2) and meteorological variables (wind speed and direction, humidity and temperature) data for a full calendar year (from 1 January 2010 to 31 December 2010) were obtained from the Kemaman (CA002) air quality Terengganu, which is located at the east coast of Peninsular Malaysia. The district of Kemaman in Terengganu state is located to the south-east of peninsular Malaysia (latitude: 4.283°N; longitude: 103.433°E) and is categorised as a residential station by the Department of Environment, Malaysia and is shown in Fig. 1. This location was chosen because it is located approximately 5.7 km eastwards from a coastal environment and 2.5 km from industrial areas which potentially influences the ozone levels.

Table 1 shown data used for the model development. Statistical data analysis was carried out by using a comprehensive statistical software R programming language (R Development Core Team, 2008) and its packages openair (Carslaw, 2013) and generalised boosted regression models (*gbm*) (Ridgeway, 2010) package. The model variables were also used to

Table 1. The variables used in the ozone boosting algorithm setting for daylight and nighttime

	Variables	Descriptions and units
Model	**Response** Ozone	Ozone concentration, O_3 (ppm)
	Predictors 1.Gases	Nitric oxides (NO) (ppm) Nitrogen dioxides (NO_2) (ppm) Nitrogen Oxides (NO_x) (ppm)
	2.Meteorological parameters	Temperature (^oC) Relative Humidity (% rh) Wind speed (km/hr) Wind direction (from the North)

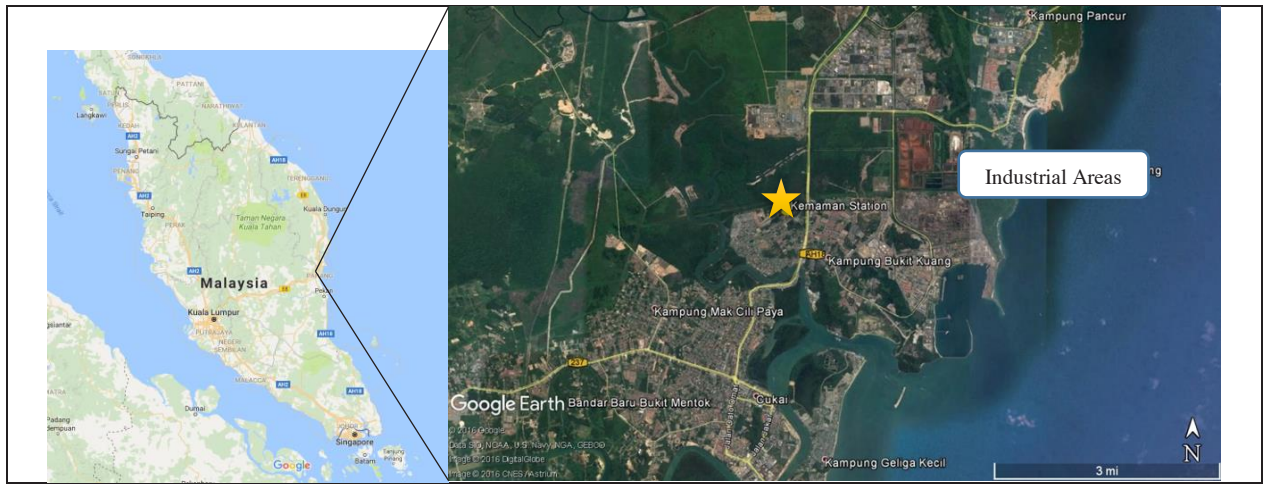

Figure 1. Map of CA002 air quality and meteorological monitoring station at Telok Kalong primary school, Kemaman, Terengganu, Malaysia

segregate patterns observed during daytime and nighttime. Malaysia situated closed to the equator therefore the daytime period is always almost twelve hours in length which normally refers to daytime is referring to time of the day between 07.00 am till 07.00 pm and nighttime will started from 07.00 pm till 07.00 am in the morning (Mohammed *et al*., 2013). The hours of daytime for each day in 2010 data were defined on the basis of Malaysia (GMT +8) sunrise and sunset times, and the remaining hours were considered as nighttime as stated by Clapp and Jenkin (2001). The calculation of day and night time durations were obtained from the 'astronomical applications' facility on the US Naval Observatory website (http://aa.usno.navy.mil/AA/). The daytime and nighttime ozone concentrations generally differed probably due to more human activities such as traveling, and agriculture activities during daytime than nighttime that influenced the level of pollutants emitted. High concentrations of O_3 were expected during midday because of photochemical reactions, which involved ozone precursors and UV radiation, unlike nighttime (Awang *et al*., 2015).

2.2 Development processes

The models were fitted in the R 3.0.2 software by using the *GBM* package version 1.6-3.1 (R Development Core Team, 2008; Ridgeway, 2010) and a combination of other packages, namely *sp, rJava, raster, dismo, survival, reshape and lattice, parallel, gplot,* and *ggplots.* There are three methods for estimating the optimal number of iterations through the fitted *GBM*; the independent test set (*test*) method, *OOB*, and CV. The specification of the three-model fitting parameters, namely *nt, lr,* and *tc,* were used for developing the BRT model. The cross-validation fold (CV) was also crucial for selecting the optimal settings (Yahaya *et al*., 2011b; Yahaya, 2013). A boosting algorithm sample for an ozone BRT model with an number of tree, *nt* value of 10000 was simulated for the daytime and nighttime datasets of Kemaman with learning rate, *lr* = 0.01, interaction depth, *tc* = 5, and *cv.fold* = 10, and the results were obtained using the algorithm setting presented in Fig. 2. The CV-fold of 10 was used in the *GBM* for obtaining an estimate of the optimal number

```
gbm2 <- gbm(o3 ~ no2 + no + nox + temp + humidity + ws + wd,
            data = night,
            distribution="gaussian",
            n.trees= 10000,
            shrinkage=0.01,
            interaction.depth=5,
            bag.fraction = 0.5,
            train.fraction = 0.5,
            cv.folds = 10,
            keep.data = TRUE,
            verbose = TRUE,
            n.minobsinnode = 10)
```

Figure 2. Best modelled BRT algorithm setting using the ozone data

of boosting iterations and plotting performance measures (Ridgeway, 2010; Yahaya, 2013). Ridgeway (2010) determining the optimal number of iterations by using the independent test set method involves the use of a single holdout base dataset. A similar shrinkage value (*lr*) of 0.001 was suggested by Ridgeway (2010) and used by Carslaw and Taylor (2009), and Yahaya (2013) for analysing an air pollution dataset. In this dataset, the optimal number of boosting iterations obtained was 2888 and 3498, 0.01 learning rate and interaction depth of 5 were found best fitted for the daytime and nighttime data respectively.

3. Results and Discussion

3.1 Statistical analysis results

Statistical analyses were performed for the hourly average values of the nitrogen gases and meteorological variables. The results show that the mean concentration of hourly average for O_3 recorded during daytime is 0.027 ppm (maximum: 0.081 ppm) and that obtained for nighttime is 0.0111 ppm (maximum: 0.052 ppm). Ozone concentrations during daytime were greater than those during nighttime by approximately 58.88%, suggesting that more ozone is formed during daytime because of the presence of solar radiation, high temperature, and other gases. However, the daytime level is still below the Malaysian Ambient Air Quality Guidelines for ozone which is 0.10 ppm for hourly and 0.06 ppm for eight hour average. The levels of NO_2, NO_x, and NO were also observed to be higher during daytime by more than 32%, and these gases potentially contributed to ozone formation. It was found that the mean data for humidity is 74.3% for daytime and 83.8% for nighttime, indicating a 13% increase at nighttime. The percentile difference between daytime and nighttime mean for the ambient temperature and wind speed is also high at 7.4% and 55%, respectively.

The wind speed and direction data were monitored by a 10-m meteorological mast monitored by the Department of Environment Malaysia. The distribution of the prevailing wind direction and wind speed data points in each cell is displayed on a colour frequency scale for the12-month sampling period and presented as a polar frequency plot in Fig. 3 separated by day and nighttime data. The plot is critical in that it provides understanding of the pattern of winds that drives many dispersion and dilution processes between daytime and nighttime. The direction from where the wind blew during daytime is also clearly shown which most of the wind blew between 0° to 180° from the North (coinciding with the sea breeze direction) and in contrast the wind blew mostly from approximate 300° from the north during nighttime which co-inside blew the land breeze direction (from the mainland to the station). The diurnal plots for observed O_3, NO_x, NO and NO_2, and meteorological variables that were plotted for entire sampling time are shown in Fig. 4(a) and 4(b). The concentration of O_3 reached a peak and slowly down in the afternoon to become steady during nighttime. High level of nitrogen oxide concentrations coinsided following similar common trends with the traffic flows (Yahaya, 2013) and the speed of the wind in which contributed to dilution and mixture of pollutants to produce O_3.

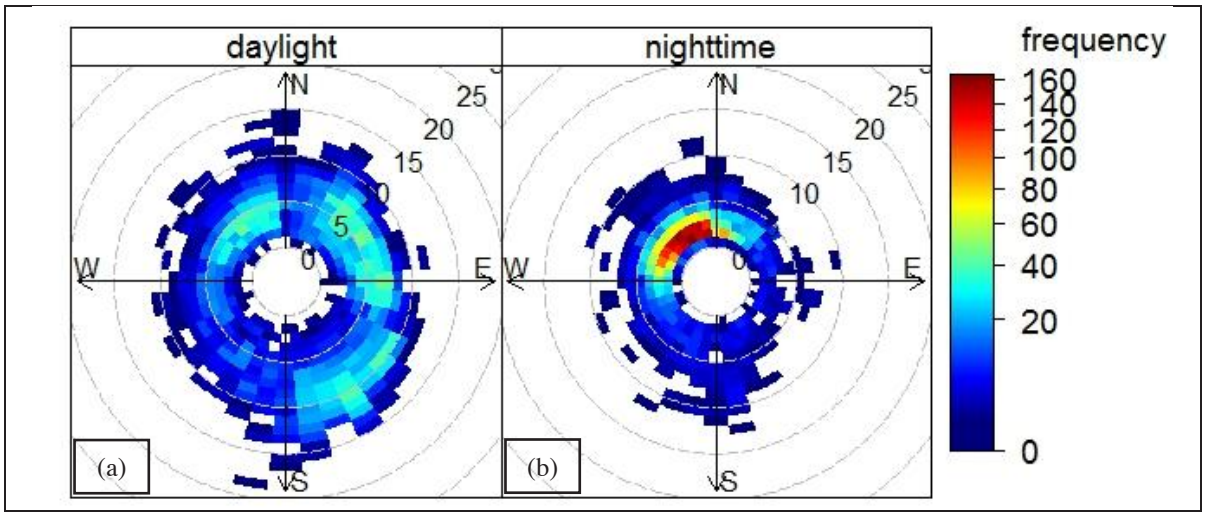

Figure 3. Wind rose polar frequency plot for the hourly average wind speed and direction over the entire sampling duration at the CA002 for (a) daytime and (b) for nighttime distributions.

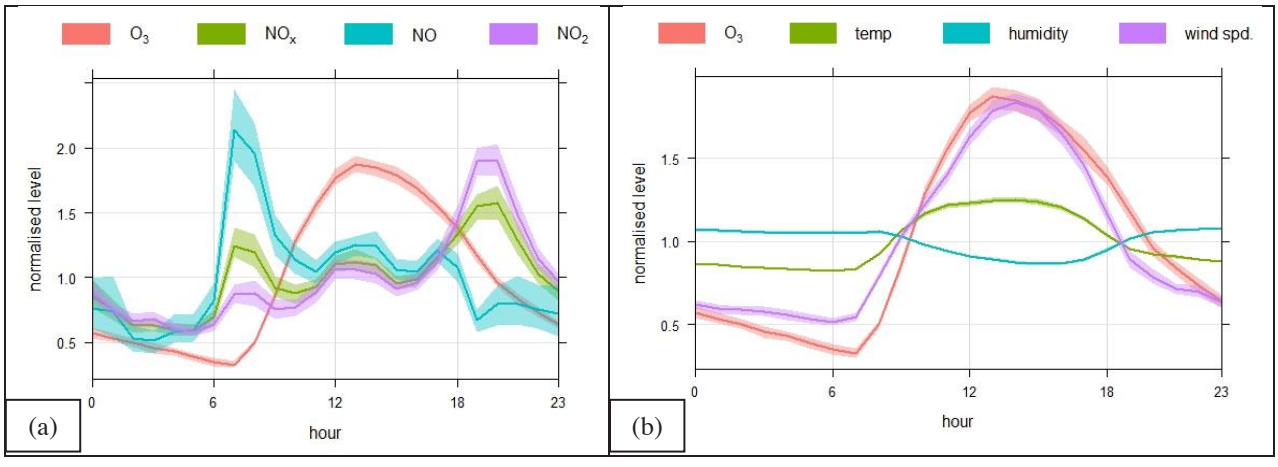

Figure 4. (a) Diurnal profile of ozone and other pollutant, (b) diurnal profile of ozone and meteorological parameters obtained from one year data.

3.2 Model performance

The model fitting performance was evaluated by using error bias analyses and performed using the model fitted and the observed ozone dataset to be an example of an operational evaluation (Derwent *et al*., 2010; Willmott *et al*., 2012). The model was evaluated by comparing the modelled and observed variables through a statistical analysis and graphic representation such as time series to detect any over fitting and overly optimistic predictive performance (Yahaya, 2013). Performance indicators such as the fraction of predictions (FAC2), normalized mean bias (NMB), coefficient of efficiency (COE), and R^2 value are mostly used for comparing and evaluating simulation models. The conditional quartile plot and scalar accuracy measures of the observed and predicted/ modelled pairs represented the results graphically (Yahaya, 2013).

Fig. 5(a) and 5(b) illustrates the conditional quantile plot of the observed and modelled ozone data for Kemaman stations for daytime and nighttime respectively. This diagram illustrates the conditional quantile distributions of the observations given in terms of the selected quantiles in comparison to 1:1 diagonal blue line which represents a perfect model (Wilks, 2006). Comparing daytime and night time distributions, the nighttime modelled data distribution and counts of the predicted ozone data. It also can be seen that more extreme ozone values being forecast more frequently, especially on the right tail of the histogram during nighttime. Thus, the developed model yielded an acceptable range of values for these datasets. Fig. 6(a) and 6(b) illustrates the scatter plot of the differences between obserbved and modelled data for Kemaman stations for daytime and nighttime respectively. The 1:1 line is solid and the 1:0.5 and 1:2 lines are dashed which help show close a datasets are to a 1:1 relationships and also show the points that are within a factor of two (FAC2) as stated by Carslaw (2013). The relationships between observed and modelled are also indicated in the figures.

The correlation coefficient (R) and coefficient of determination (R^2) between the observations and the fitted model obtained from the aforementioned analysis show the performance of the model fitted. A correlation coefficient of ±1 indicates a perfect model between two variables. The daylight and nighttime R (R^2) values between the fitted model and the observed datasets were found to be *0.83* (*0.69*) and *0.74* (*0.55*), respectively, which suggests that the method is statistically valid. However, nighttime predicted ozone are not good which indicated only 74% of ozone explained by explanatory variables compared to daytime. Factors such as the influenced of the wind that blew from the sea and carried mixed of ozone precursors such as nitryl chloride which produced ozone during daytime as stated by Wallace and Hobbs (2006) and the high wind speed compared to nighttime that potentially contributed to the mixture of precursors of ozone during daytime.

This implies that more than 70% of the variation of ozone concentrations was explained by the explanatory variables while there other variables such as solar intensity and other gases have an influenced to ozone concentrations which is not included in this analysis. FAC2 and NMB values are within acceptable ranges if they are between 0.5 and 2.0. In this case, the value of FAC2 for Kemaman is 0.93 and 0.79 for daytime and nighttime, respectively. Furthermore, the recommended values of the NMB also ranged between acceptable range of -0.112 and -0.154 for daytime and nighttime, respectively. In this case the developed model was within an acceptable value range and performance well against the observations over the entire observation period.

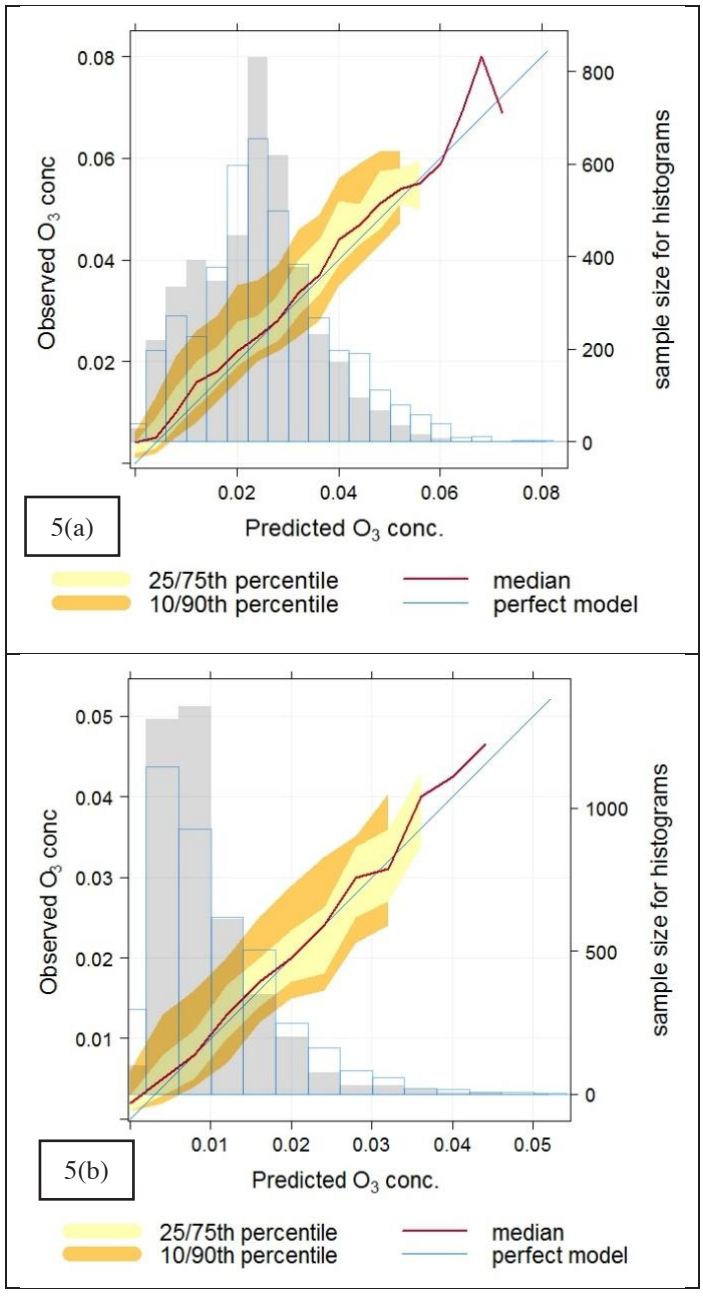

Figure 5. Conditional Quantile plot fors for ozone (a) daytime and (b) for nighttime modelled for Kemaman station datasets

3.3 Boosted regression trees result

One of the steps in the BRT interpretation involves obtaining a partial dependence plot for showing the relationship between the predictor and the response variables. Fig. 7 shows the fitted explanatory variables involved in this model to the concentration of response variable (O_3) for daytime data. It was found that positive relationships (relatively increased) were obtained between the O_3 concentration and NO_2 and NO_x concentrations, temperature, and wind speed for the daytime data. NO is a primary contaminant, whereas O_3 and a large percentage of NO_2 are secondary contaminants, formed through a set of complex reaction. NO is converted to NO_2 via a reaction with O_3. During daytime, NO_2 is converted back to NO as a result of photolysis, leads to regeneration of O_3. A reaction that converts NO into NO_2 without destroying ozone occurs in the presence of hydrocarbons. Previous study conducted by Ismail *et al*. (2016) was found that a weak linear relationship between O_3 and NO_x, suggesting that Kemaman area is not NO_x sensitive but possibly VOC-sensitive area. This explain why NO_2 and O_3 have positive relationship. This result indicates that the increment of NO_2 in the air will potential produces more O_3 with the existing of high

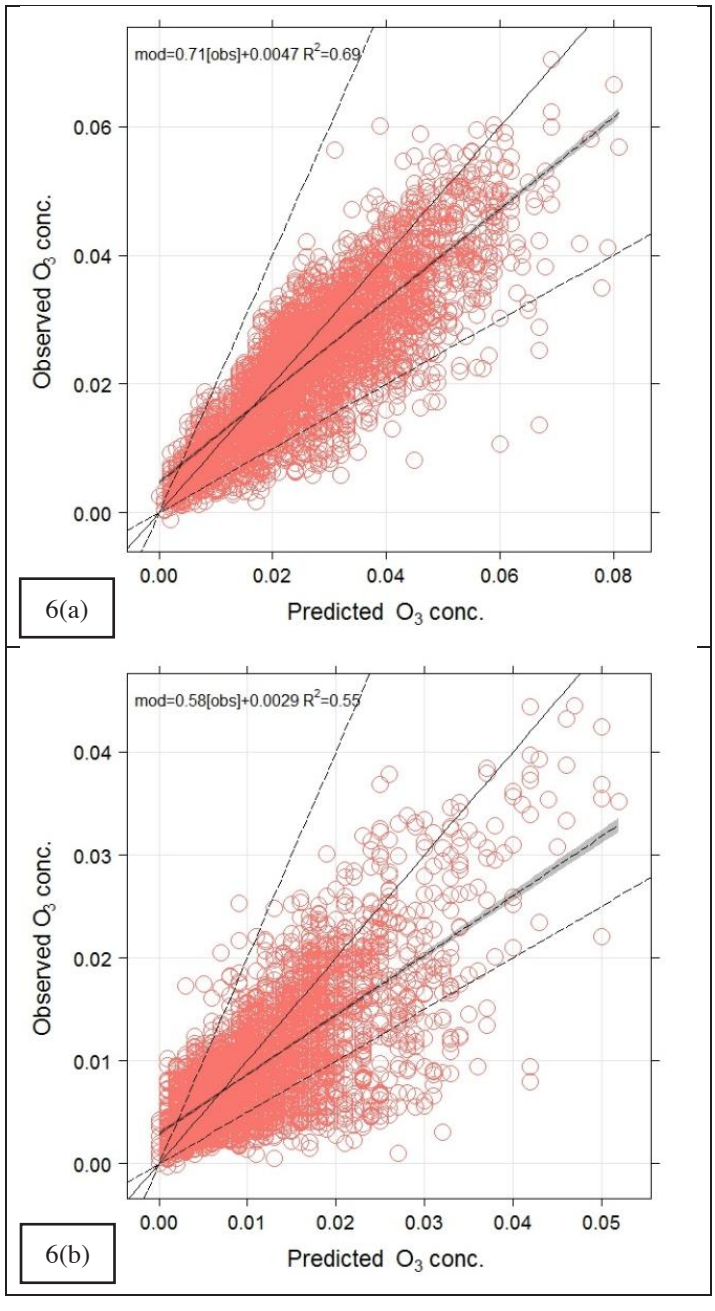

Figure 6. Scatter plots for observed and modelled for (a) nighttime and (b) for nighttime data for Kemaman datasets

temperature and the high wind would trigger the turbulence of air in which trigger the formation of O_3 as explained by Tu *et al*. (2007). High wind speeds tend to increase the dilution of the variables (e.g., nitrogen gases) thereby influencing ozone formation. The results also clearly indicates that industrial activities emitted most of NO_2 has an influenced to produce more O_3. The relationships between ozone formation and the meteorological variables were in agreement with those expected from theories of the physical processes associated with ozone formation. For example, O_3 formation increased both with the wind speed and temperature. Thus, the higher the wind speed and temperature, the higher the ozone concentrations were indicated. However, the O_3 concentration decreased steadily with a decrease in humidity variables

Fig. 8 shows partial dependence plots for nighttime at Kemaman stations. Overall, except for the humidity (negative relationships), similar trends were observed which shows a positive relationship with the O_3 level or a trend opposite to that obtained during daytime. However, parameters such as the nitrogen oxide concentrations, temperature and wind speed have a positive relationships with the ozone concentration at night but the concentration level was lower compared to daytime which caused by the absent of solar intensity during nighttime.

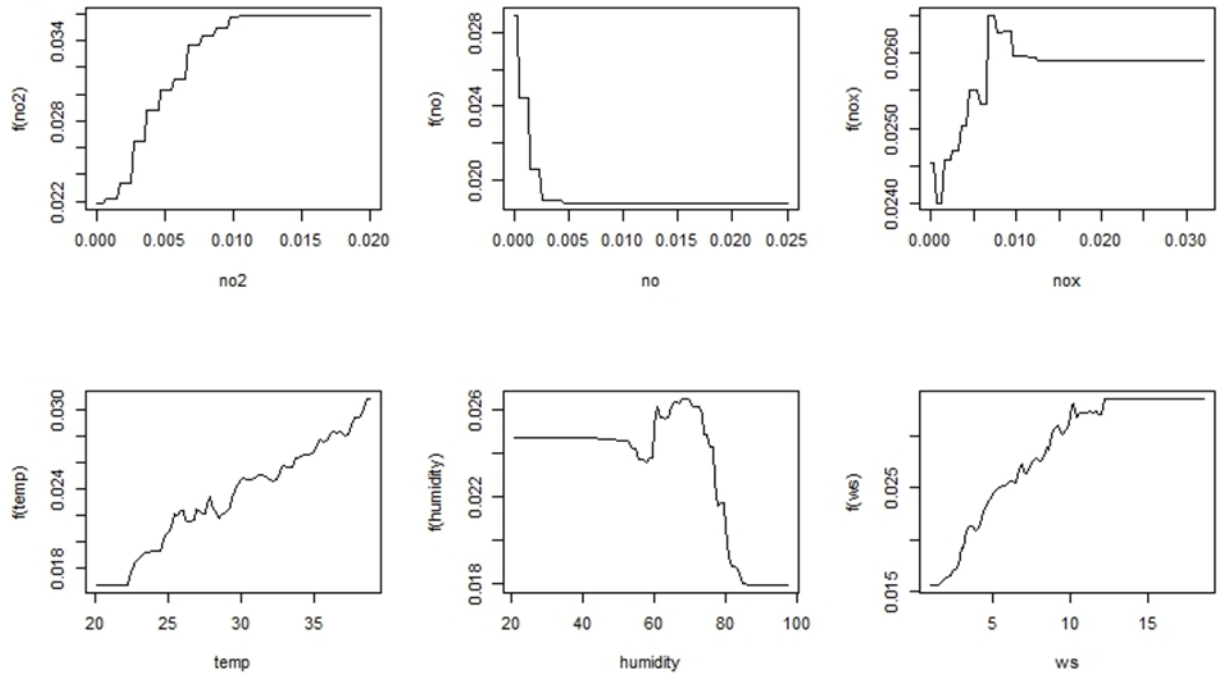

Figure 7. Partial dependence plots showing the variation of the hourly daytime ozone concentration at Kemaman station with each variable used in the BRT model.

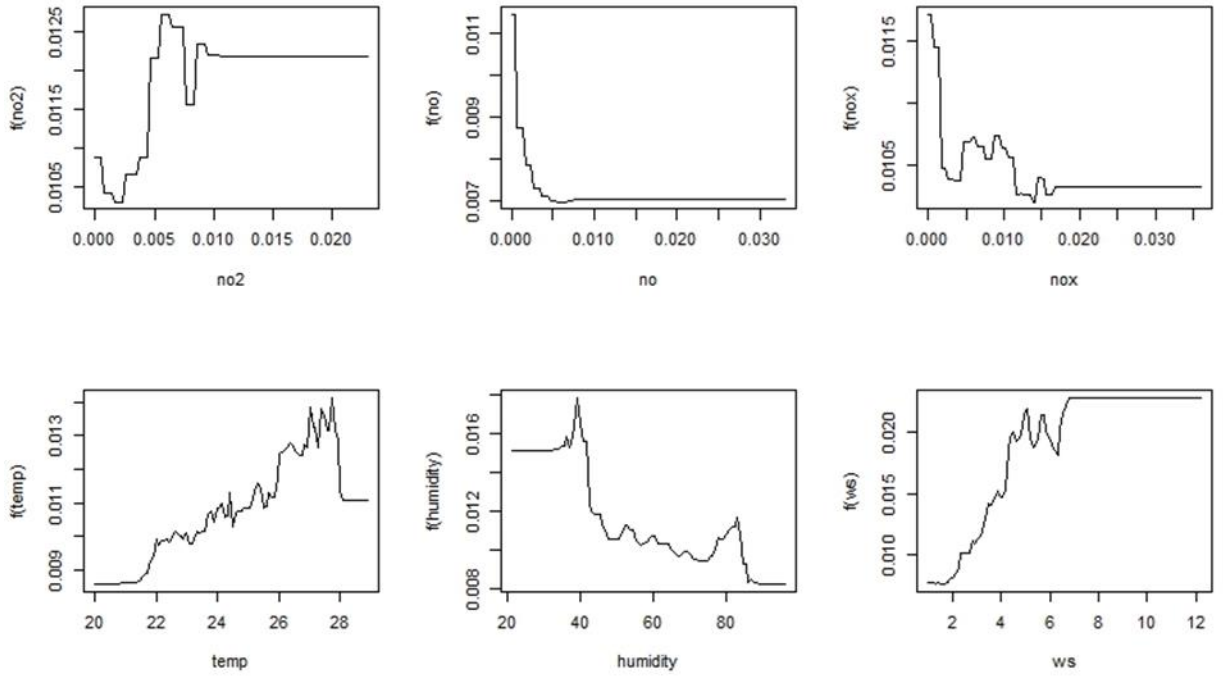

Figure 8. Partial dependence plots showing the variation of the hourly nighttime ozone concentration at Kemaman station with each variable used in the BRT model.

3.4 Relative importance of the predictors

The analysis for Kemaman strongly indicated that seven crucial variables affected ozone concentration. The variables NO, NO_2, NO_x, temperature, humidity, wind speed, and wind direction have been commonly used in other air pollution studies and are known to influence ozone and air pollutant levels. The relative influences of the variables that contributed the most to ozone formation or destruction were scaled and expressed as a percentage. The higher relative importance (percentage value) of the predictor shows the extent to which factors influence the hourly ozone concentrations.

The four meteorological variables with the most influence on ozone formation/destruction and concentrations at Kemaman station during daytime are the wind speed: 26.85%; temperature: 18.68%; relative humidity: 18.52%; wind direction: 15.74%; and variables associated with gases show the least relative importance: NO_2, NO, and NO_x showed relative importance values of 9.44%, 8.1% and 2.60%, respectively. The relative importance (percentage value) of the predictors contributing the most to ozone formation at Kemaman station during nighttime was as follows: wind speed: 38.97%; wind direction: 22.07%; temperature; 13.38; and humidity:13.16%. The lowest percentage of relative influence was 4.89% for NO, implying that this compound had the weakest influence on ozone formation compared with the other parameters. Both industrialization and vehicular traffic congestion in the this study were observed to contribute substantially to a high level of ozone pollutants which contributed substantially to a high level of ozone pollutants

4. Conclusions

The used of a boosting model as a statistical tool for predicting ozone formation is a new approach to the analysis of ozone data. The BRT model has many advantages compared with other techniques, such as its capability to select relevant variables during the algorithm setting, fit accurate functions, and automatically identify the best iteration from multi model developed. In addition, the prediction techniques used to determine the relative importance of the different variables in influencing the ozone level could indicate the variables contributing the most to ozone pollution. The developed model is expected to be useful for decision-makers because it can enable them to consider precursors parameters when identifying measures to mitigate the current air pollution. Compared with other techniques, the BRT method is helpful for dealing with complex ozone data and moving from conventional analytical models to onces which include multi-level methods that can account for variable intersections, a clear understanding of day and nighttime and the understanding of ozone concentration in station closed to coastal environment can be understood.

References

Abdul-Wahab SA, Bakheit CS, Al-Alawi SM. Principal component and multiple regression analysis in modelling of ground-level ozone and factors affecting its concentrations. Environmental Modelling and Software 2005; 20(10): 1263-71.

Abdul-Wahab SA, Al-Alawi SM. Assessment and prediction of tropospheric ozone concentration levels using artificial neural networks. Environmental Modelling and Software 2002; 17(3): 219-28.

Awang NR, Ramli NA, Yahaya AS, Elbayoumi M. High nighttime ground-level ozone concentrations in Kemaman: NO and NO_2 concentrations attributions. Aerosol and Air Quality Research 2015; 15(4): 1357-66.

Barrero MA, Grimalt JO, Canton L. Prediction of daily ozone concentration maxima in the urban atmosphere. Chemometrics and Intelligent Laboratory Systems. 2006.

Bruno F, Cocchi D, Trivisano C. Forecasting daily high ozone concentrations by classification trees. Environmetrics 2004; 15(2): 141-53.

Carslaw DC. The openair manual: open-source tools for analysing air pollution data. King's College of London. 2013.

Carslaw DC, Taylor PJ. Analysis of air pollution data at a mixed source location using boosted regression trees. Atmospheric Environment 2009; 43(22-23): 3563-70.

Chaloulakou A, Assimacopoulos D, Lekkas T. Forecasting daily maximum ozone concentrations in the Athens Basin. Environmental Monitoring and Assessment 1999; 56(1): 97-112.

Clapp LJ, Jenkin ME. Analysis of the relationship between ambient levels of O_3, NO_2 and NO as a function of NO_x in the UK. Atmospheric Environment 2001; 35(36): 6391-405.

Department of Environment Malaysia. Malaysian environmental quality report. Ministry of Natural Resources and Environment, 2011; 17-18.

Derwent D, Fraser A, Abbott J, Jenkin M, Willis P, Murrellst. Evaluting the performance of air quality models. Report to the Department for Environment, Food and Rural Affairs, the Scottish Executive, Welsh Assembly Government and the Department of the Environment Northern Ireland, 2010.

De'ath G. Boosted trees for ecological modelling and prediction. Ecology 2007; 88(1): 243-51.

Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. Journal of Animal Ecology 2008; 77(4): 802-13.

Finlayson-Pitts BJ, Pitts JNJ. Chemistry of the upper and lower atmosphere: theory, experiments, and applications. Academic Press, San Diego, CA, USA. 2000.

Friedman JH. Stochastic gradient boosting. Computational Statistics and Data Analysis 2002; 38(4): 367-78.

Friedman JH. Greedy function approximation: a gradient boosting machine. The Annals of Statistics 2001; 29(5): 1189-232.

Gardner MW, Dorling SR. Statistical surface ozone models: an improved methodology to account for non-linear behaviour. Atmospheric Environment 2000; 34(1): 21-34.

Ghazali NA, Ramli NA, Yahaya AS. A Study to investigate and model the transformation of nitrogen dioxide into ozone using time series plot. European Journal of Scientific Research 2009; 37(2): 192-205.

Ghazali NA, Ramli NA, Yahaya AS, Yusof NFF, Sansuddin N, Al Madhoun WA. Transformation of nitrogen dioxide into ozone and prediction of ozone concentrations using multiple linear regression techniques. Environmental Monitoring and Assessment 2010; 165(1): 475-89

Hassanzadeh S, Hosseinibalam F, Omidvari M. Statistical methods and regression analysis of stratospheric ozone and meteorological variables in Isfahan. Physica A: Statistical Mechanics and its Applications 2008; 387(10): 2317-27.

Heck WW, Taylor OC, Tingey DT. Assessment of crop loss from air pollutants. Elsevier Applied Science, London. 1998; 552.

Hsieh-Hung T, Tsung-Hung T, Chung-Shin Y, Chung-Hsuang H, Chitsan L. Effects of sea-land breezes on the spatial and temporal distribution of gaseous air pollutants at the coastal region of Southern Taiwan. Journal of Environmental Engineering and Management 2008; 18(6): 387-96.

Huang LS, Smith RL. Meteorologically-dependent trends in urban ozone. Florida State University (FSU) Technical Report M-916, National Institute of Statistic Science, Research Triangle Park, NC 27709. 1999.

Hubbard MC, Cobourn WG. Development of a regression model to forecast ground-level ozone concentration in Louisville, KY. Atmospheric Environment 1998; 32 (14-15): 2637-47.

Ismail M, Abdullah S, Yuen FS, Ghazali NA. A ten-year investigation on ozone and it precursors at Kemaman, Terengganu, Malaysia. EnvironmentAsia 2016; 9(1): 1-8.

Kriegler B. Cost-sensitivive stochastic gradient boosting within a quantitative regression framework. Ph.D. Thesis. University of California, LA, 2007.

Lawrence R, Bunn A, Powell S, Zambon M. Classification of remotely sensed imagery using stochastic gradient boosting as a refinement of classification tree analysis. Remote Sensing of Environment 2004; 90(3): 331-36.

Leathwick JR, Elith J, Francis MP, Hastie T, Taylor P. Variation in demersal fish species richness in the oceans surrounding New Zealand: an analysis using boosted regression trees. Marine Ecology Progress Series 2006; 321: 267-81.

Mohammed NI, Ramli NA, Yahya AS. Ozone phytotoxicity evaluation and prediction of crops production in tropical regions. Atmospheric Environment 2013; 68: 343-49

Munir S, Habeebullah TM, Mohammed AMF, Morsy EA, Awad AHAH, Seroji AR, Hassan IA. An analysis into the temporal variations of ground level ozone in the arid climate of Makkah applying k-means algorithms. EnvironmentAsia 2015; 8(1): 53-60.

R Development Core Team. A language and environment for statistical computing. *In*: Computing, (*Ed*: R. F. F. S.). Vienna, Austria. 2008.

Ridgeway G. GBM: generalized boosted regression models. R packages version 1.6-3.1. 2010.

Robeson SM, Steyn DG. Evaluation and comparison of statistical forecast models for daily maximum ozone concentrations. Atmospheric Environment. Part B. Urban Atmosphere 1990; 24(2): 303-12.

Sadanaga Y, Matsumoto J, Kajii Y. Photochemical reactions in the urban air: recent understandings of radical chemistry. Journal of Photochemistry and Photobiology C: Photochemistry Reviews 2003; 4(1): 85-104.

Seinfeld JH, Pandis SN. Atmospheric chemistry and physics from air pollution to climate change. 2nd ed. New Jersey: John Wiley & Sons Inc., 2006.

Steiner AL, Davis AJ, Sillman S, Owen RC, Michalak AM, Fiore AM. Observed suppression of ozone formation at extremely high temperatures due to chemical and biophysical feedbacks. Proceedings of the National Academy of Science of the United States of America 2010; 107(46): 19685-90.

Teixeira EC, Ramos de Santana E, Wiegand F, Fachel J. Measurement of surface ozone and its precursors in an urban area in South Brazil. Atmospheric Environment 2009; 43(13): 2213-20.

Tu J, Xia ZG, Wang H, Li W. Temporal variations in surface ozone and its precursors and meteorological effects at an urban site in China. Atmospheric Research 2007; 85(3-4): 310-37.

Wallace JM, Hobbs PV. Atmospheric science: an introduction survey. 2nd ed. United States of America: Academic Press Publication. 2006.

Wang W, Lu W, Wang X, Leung AYT. Prediction of maximum daily ozone level using combined neural network and statistical characteristics. Environment International 2003; 29(5): 555-62.

Wilks DS. Statistical methods in the atmospheric sciences. 2nd ed. Elsevier,United State. 2006.

Willmott CJ, Robeson SM, Matsuura K. A refined index of model performance. International Journal of Climatology 2012; 32(13): 2088-94.

Yahaya NZ, Tate JE, Tight MR. Studying particle number noncentrations (PNC) in an urban street canyon: using boosted regression trees, BRT. Proceedings in The International Conference on Humanities, Social Sciences and Science Technology, Manchester University UK. 2011a.

Yahaya NZ, Tate JE, Tight MR. Analyzing roadside particle number concentrations using boosted regression trees (BRT). Proceedings in The European Aerosol Conference, Manchester University, 2011b.

Yahaya NZ. Temporal and spatial variation of ultra-fine particles in the urban environment. Ph.D. Thesis of the Institute for Transport Studies, University of Leeds, United Kingdom, 2013

Received 7 November 2016
Accepted 4 January 2017

Correspondence to
Dr.Noor Zaitun Yahaya
School of Ocean Engineering,
Universiti Malaysia Terengganu,
Kuala Terengganu,
Malaysia
Tel: +60-9668-3972
Fax: +60-6683-441
E-mail: nzaitun@umt.edu.my