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

EnvironmentAsia 11(3) (2018) 221-234
DOI 10.14456/ea.2018.50
ISSN 1906-1714; ONLINE ISSN: 2586-8861

Analysis of Fine and Coarse Particle Number Count Concentrations Using Boosted Regression Tree Technique in Coastal Environment

Noor Zaitun Yahaya[1]*, Siew Moi Phang[2], Azizan Abu Samah[2], Intan Nabila Azman[1], and Zul Fadhli Ibrahim[1]

[1] School of Ocean Engineering, Universiti Malaysia Terengganu, MALAYSIA
[2] Institute of Ocean and Earth Science, University Malaya, MALAYSIA

*Corresponding Author: nzaitun@umt.edu.my

Received: April 5, 2018; Accepted: June 20, 2018

Abstract

Particle number count concentrations ([PNC]) is a new metric unit that can be used to quantify the characteristics of particles in the atmosphere. This study was conducted to explore the variability of [PNC] and the relationship between the factors that influenced this variation. The [PNC], gases (SO_2 and NO_x), and meteorological factors (wind speed, wind direction, humidity, pressure and temperature) data were gathered for a six months period from the Institute of Ocean and Earth Sciences (IOES) Station, Kelantan, Malaysia by using a particle counter (GRIMM, model EDM180), EcoTech EC9805T Series and EcoTech EC9841T Series for gases and Lasteem Model LSI for meteorological measurements. The [PNC] data were categorised into fine particle number count concentrations ($FPNC_{0.25–0.99}$ and $FPNC_{1.0\ \text{-}2.49}$) with diameters of 0.25–0.99 µm, 1.0 -2.49 µm and coarse particles number count concentrations ($CPNC_{2.5–10}$) with diameters of 2.5–10 µm. The particle number concentration were measured and reported in number count/particles at the entire size or number in every litre of air flow that pumped into the instruments (EDM180, GRIMM). The concentration of FPNC was found higher (maximum of 5,826,380 counts/L) compared to CPNC (maximum of 818 counts/litre). An artificial intelligent technique (boosted regression trees (BRT) algorithm) was constructed from multiple regression models, and the best iteration of the BRT model was performed by optimising prediction performance. The analysis revealed that the significant variation in the FPNC was largely influenced by SO_2 (46.53%), Julian day (13.71%) and wind direction (10.50%). In contrast, the CPNC was primarily influenced by wind speed (22.33%), wind direction (18.89%), Julian day (18.17%) and pressure (11.38%).

Keywords: Particle number count concentrations; Fine particles; Coarse particles; Boosted regression trees; Coastal area

1. INTRODUCTION

Aerosol can be described physically by the mass, surface and concentration of its particles that come in a range of sizes. Aerosol has an impact on the environment and on human health especially, both over the short and the long term. According to Yahaya (2013), the particle number count concentrations ([PNC]) can be used to quantify the characteristics of the particles in the atmosphere. The Air Quality Expert Group (AQEG, 2005) categorises this particulate matter by size of particle into nano, ultrafine, fine, coarse, and over 10 μm. It has been reported by the World Health Organisation that more than two million premature deaths worldwide each year are attributed to urban outdoor and indoor air pollution that contains particulate matter (WHO, 2005). Particulate matter in air pollution causes health problems in humans, especially respiratory diseases, cardiovascular disease, damages internal organs such as lungs and heart, causes cancer, increases mortality and premature death (Kampa and Castanas, 2008; Shridhar *et al.,* 2010; EPA, 2011; Janssen *et al.,* 2013). Harrison *et al.,* (1997) cited in Massey *et al.,* (2012) state that there are three sources of particulate matter that have an impact on health: (i) primary fine particles from industrial and combustion sources, predominantly road traffic; (ii) secondary aerosol, mostly ammonium sulphate and ammonium nitrate formed through photochemical reactions; and (iii) wind-blown soil and re-suspended street dust present largely in coarse fractions (2.5-10 mm).

A coastal environment is categorised by the distance from a certain point or location near the coast or shore. Hail (1970) defines a coastal zone as an area of variable width that extends seaward to the edge of the continental shelf, but which has no distinct landward demarcation. On the other hand, in a study based in Malaysia, a coastal zone is described as an area that extends inland by approximately 1 kilometre from the mean low tide level and seaward to the outermost limit of the state boundary (Mastura, 1992).

The main objective of this study was to determine the relationship between the [PNC] and a number of meteorological conditions and gases that influence the [PNC] in a Malaysian coastal area by using the boosted regression trees (BRT) technique. The interaction between the variables was observed and explored with reference to the output from the BRT. Although there was a number of research performed a simultaneous measurement of [PNC], gases and meteorological variables however, due to limited technique and approach used most model developed only include limited measurements as a predictive variables in model prediction. Therefore, a new approach which is called the boosted regression trees techniques with an artificial intelligent approach was used in this study to achieve the objectives of this study.

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), who then added a stochastic element to the boosting algorithm by taking a random sample from the training dataset without replacing it with observation data in the iteration (Friedman, 2002). The advantages of techniques based on the decision tree are explained in De'ath (2000) and can be summarised as follows: flexibility to handle a wide range of response types, rank statistics that result in invariance of the tree to any monotonic transformations of the explanatory variables, ease and robustness of construction, ease of interpretation of complex results involving interactions, and the ability to handle missing values in both the response and explanatory variables. Furthermore, extreme outliers do not have an effect on the prediction results. Therefore, the BRT technique offers advantages over other methods such as linear regression and multiple regression in the context of air pollution modelling (Carslaw and Taylor, 2009). The most recent BRT technique developed by Friedman (2002) has been used in several proposed modelling and forecasting methods in the fields of ecology (Leathwick *et al.,* 2006; De'ath, 2007; Elith *et al.,* 2008) and atmospheric environment (Carslaw and Taylor, 2009; Yahaya, 2013; Munir *et al.,* 2014).

In this study we focused on the most dominant particle number which are FPNC in range 0.25 – 1.0 μm hereafter call FPNC and CPNC with 2.5 – 10 μm hereafter call CPNC.

2. MATERIALS AND METHODS

A 10-minutes data of the [PNC], SO_2 (ppb), NO_X (ppb), and selected meteorological factors (humidity, temperature, pressure, wind speed and wind direction) for a 6-month period (6 January to 5 July 2015) were measured and gathered from the Institute of Ocean and Earth Sciences (IOES) Station Bachok, Malaysia. The IOES Station is located on the east coast of Peninsular Malaysia (N 6.0086; E 102.4259) and is shown in Figure 1. The location was chosen because the monitoring tower (approximate 20 meters height) is mounted approximately 100 metres from the water's edge of the South China Sea. The distance of the IOES Station from Kota Bharu, the capital city of Kelantan, is approximately 30 kilometres and mounted at approximate 20 meters high monitoring tower. These IOES station is categorised as a rural area where agriculture is the main activity of the residents and fishing activities also take place along the coastal area. [PNC] were monitored using a particle counter (Model EDM180, GRIMM, Germany) which provided 1-minute concentrations of 31 channels of the [PNC] (Particle Diameters = 0.265–34 μm) and reported in units of counts/litre. The EDM180 uses a patented laser with a 90° scattering angle to detect the particles. It also has a measuring chamber that the sample of air enters from the top, which ensures that only one particle is measured at a time. The SO_2 (ppb) and NO_X (ppb) gases were measured by an EcoTech EC9805T Series and EcoTech EC9841T Series, Australia, respectively. The particles, gases and meteorological data were then collected in a paperless recorder (Brainchild Data Logger Model VR18). The weather stations Model LSI Lasteem from the United Kingdom were used to measure the humidity (%), temperature (oC), pressure (pascal), wind speed (m/s) and wind direction (degree from the North).

From the previous research on the aerosol studies, besides the natural and anthropogenic sources, the meteorological factors also give effects to the size distribution of the particles. These factors also important in the new particle formation events, suggested that there are natural processes leading to significant production of particles. (Vakeva *et al.,* 2000)

Total of ten-minute 25,958 data ([PNC], SO_2, NO_X and selected meteorological conditions) were compiled in a Microsoft Excel spreadsheet in .csv format. The R programming language, which was developed by the Development Core Group R (2008), was used to analyse the data statistical and graphically. R is 'GNU S'.

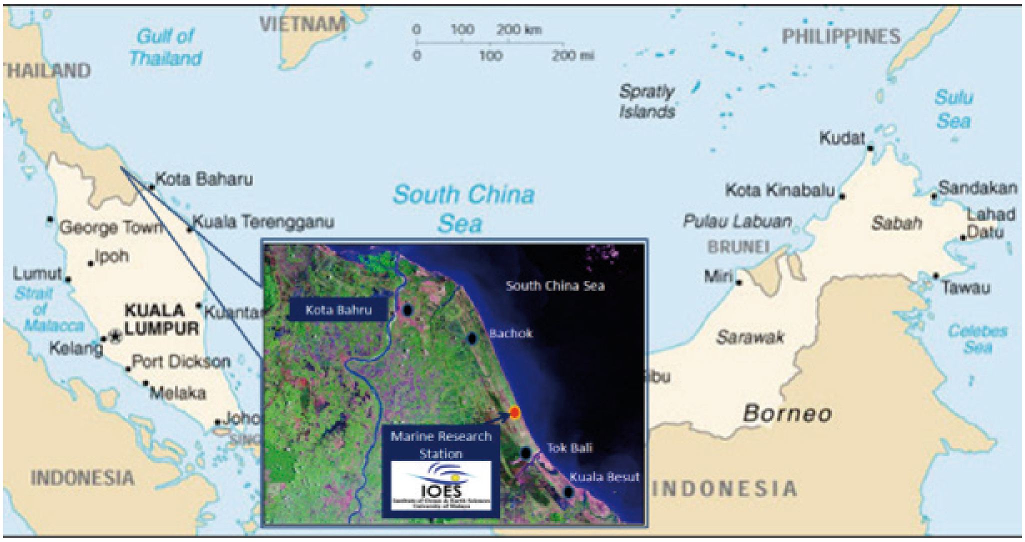

Figure 1. IOES monitoring station at Bachok, Kelantan, Malaysia

R software provides a wide variety of statistical and graphical techniques including linear and nonlinear modelling, statistical tests, time series analysis, classification, and clustering among others (Hornik, 2014). In this study, the "openair" packages (Carslaw 2013) and the generalised boosting machine (gbm) package (Ridgeway, 2010) and its packages were used to analyse the data and develop the BRT algorithm.

3. RESULT AND DISCUSSION

3.1 Statistical Analysis Results

The statistical analyses were performed on the 10-minute average values of the [PNC], SO_2 and NO_X and the meteorological factors, as summarised in Table 2. In this study we focused on the most dominant particle number which are FPNC in range and CPNC. The results show that higher level of FPNC mean concentration of 281,513 counts/litre (maximum: 5,826,380 ± 420,666 counts/L) and the mean of 37.46 counts/L (maximum: 818 ± 50 counts/L) were recorded for CPNC were recorded. This phenomena may link to sea spray that emitted fine particles due to coagulation and chemical reaction which is not included in this study. Statistically, the mean of both FPNC and CPNC was higher compared to the median, which indicates that possible extreme events occurred over the period under study which in this case maybe from the monsoon and season variability factors.

3.2 Model Development Process

The BRT technique was applied in this study to analyse the relationship between fine and course particles, meteorological factors and the gases. The models were fitted in R 3.0.2 software by using the gbm package, version 1.6-3.1 (R Development Core Team, 2008; Ridgeway, 2010). Three methods can be used to estimate the optimal number of iterations through the fitted gbm: the independent test set (test) method, out-of-bag estimation (OOB), and cross-validation (cv) (Yahaya *et al.*, 2011;

Yahaya, 2013; Yahaya *et al.,* 2016). There are three important parameter settings for the BRT algorithm: number of trees (nt), learning rate (lr) and interaction depth or tree complexity (tc). A boosting algorithm sample for a [PNC] BRT model with a nt value of 10,000 was simulated for the total [PNC] sample. The error distribution for this study assumed a Gaussian error distribution with similar assumptions to those applied in Carslaw and Taylor (2009). A set of data that uses training datasets were computed to determine the best iteration and the minimum error by using the optimum BRT parameters with stochastic approaches.

Ridgeway (2010) determines the optimal number of iterations required by using the independent test set method that involves the use of a single holdout base dataset. A boosting algorithm sample for the particle model with nt = 10,000 was simulated for the datasets obtained from Bachok, Kelantan with lr = 0.01, tc = 5, and cv.fold = 10, which are the values suggested by Ridgeway (2010), Carslaw and Taylor (2009) and Yahaya (2013) for analysing an air pollution dataset. A 10-fold cv was used in the gbm in order to obtain an estimate of the optimal number of boosting iterations and plotting performance measures (Ridgeway, 2010).

3.3 Model Performance

The model was developed by using the training dataset and the model fitting performance was then evaluated by using the error bias analyses. The model performance was then evaluated by using the testing datasets. First, the randomness of the model was set by using the set.seed function. The dataset was split into 70% for training and 30% for testing by using train.fraction function. The best combination will give the lowest value in the root mean square error (RMSE). The model evaluation stat

factor of two (FAC2), correlation coefficient (R), and index of agreement (IOA) – were computed to compare the predictive performance of the models.

When the BRT algorithm was set using parameters lr = 0.005, 0.05, tc = 5, 5, nt = 9,999, 9,997, and number of data (n) = 23,597, 23,597 for FPNC and CPNC, respectively, it achieved the minimum predictive error and this iteration of the algorithm was found to best fit the data. These settings were selected based on the lowest RMSE value of 192,920.3 and 17.73, respectively.

Table 1. Summary of Statistical Data for Variables Measured at IOES Station

Variables	Mean	Median	Maximum	SD
FPNC (count/litre)	281,513	141,361	5,826,380	420,666
CPNC (count/litre)	37.46	21	818	50
SO2 (ppb)	11.06	0.72	215.94	35.22
NOX (ppb)	1.46	1.1	7.16	1.03
Humidity (%)	73.89	77.27	100	27.38
Temperature (˚C)	26.54	26.25	32.87	2.22
Pressure (hPa)	1008	1008	1016	2.47
Wind speed (m/s)	3.164	2.44	12.23	2.08

Pearson's correlation coefficient, the R value and the coefficient of determination (R^2) between the observed and modelled data indicated that the performance of the model was good. A correlation coefficient value approaching 1 shows a perfect model between the two variables. The FPNC and CPNC R (R^2) values were 0.90 (0.81) and 0.94 (0.87), respectively, which indicates that both the observed and modelled number counts for both the fine and coarse particles were in good correlation with each other. This result implies that more than 90% of the variations in both of these two types of [PNC] were explained by the explanatory variables. The FAC2 values for both FPNC and CPNC were 0.85 and 0.81, respectively, which are in the recommended range of 0.5 to 2. Thus, the fraction of predictions within a factor of two of the observed values is the ratio between model-predicted and observed variables. Hence, the developed models produced results that were within the acceptable value range, meaning that it is suitable for use as a predictive model.

Conditional quantiles are a very useful way of considering model performance against observations for continuous measurements (Wilks, 2005). The conditional quantile plot splits the data into evenly spaced bins. For each predicted value bin, the corresponding values of the observations are identified and the median, 25/75th and 10/90 percentile (quantile) are calculated for that bin. Next, the data are plotted to show how these values vary across all bins. Figure 2(a) and (b) show the conditional quantile plot for the model and observations for the FPNC and CPNC, respectively. The blue diagonal line indicates the result required for a perfect model, while the red line shows the median value for the predictions. From the figure, the plot for the CPNC has an almost a perfect median line (red line) compared to that for the FPNC. The shading in the plot shows the predicted quantile interval. There is still some spread, especially for the FPNC, because even for a perfect model a specific quantile interval will contain a range of values. However, for the number of bins used in this plot the spread is very narrow. The histogram shows the counts of predicted values for both types of [PNC] had been forecast to be more frequent, especially on the right tail of the histogram. Hence, both developed models gave an acceptable range of values for the dataset. The scatter plots in Figures 2(c) and (d) illustrate the differences between the observed and modelled data for the FPNC and CPNC, respectively. The 1:1 diagonal is solid and the 1:0.5 and 1:2 lines are dashed, which indicates how close the datasets are to 1:1 relationships and the points that are within FAC2, as stated by Carslaw (2013).

3.4 Boosted Regression Trees Result

There are three main outputs that can be obtained from the BRT analysis output which are the partial dependence plot, the relative influences of the variables, and the interactions between the variables (Yahaya, 2013). The BRT modelling process can be used to examine the relationships between the independent variables (gases and meteorological variables) and the dependent variable (PNC). Visualisation of the fitted functions in a BRT model is achieved by using partial dependence functions and is an effective way to show the response after accounting for the effects of all the other variables in the model (Friedman, 2001). In this case, models were pulled from the gbm best iteration output and then plotted using partial dependence plots.

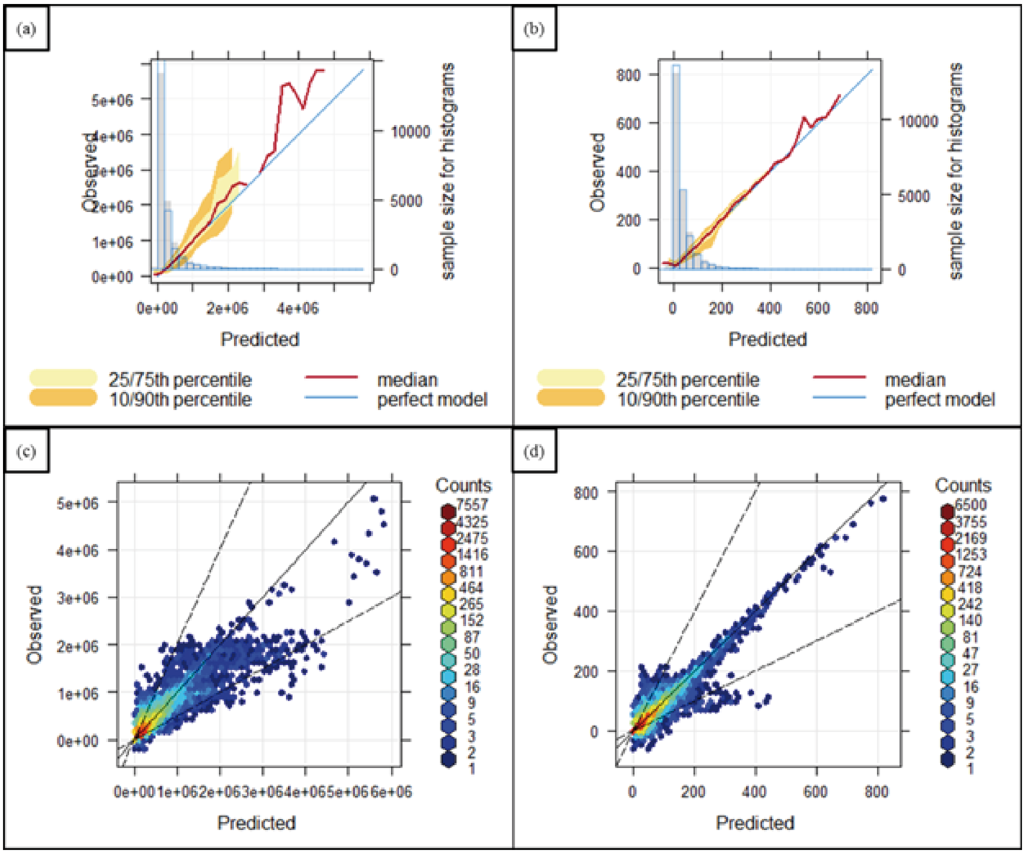

Figure 2. (a) conditional quantile plot for FPNC; (b) conditional quantile plot for CPNC; (c) scatter plot for FPNC;(d) scatter plot for CPNC for IOES Station datasets

3.5 *Partial Dependence Plot*

The partial dependence plot for the gases and meteorological variables demonstrates the relationships between these variables and the fitted model of the [PNC] for the IOES Station, and these relationships were also examined and compared between variables. Figure 3 and Figure 4 show partial dependence plots derived from gbm output for the FPNC and CPNC at the IOES Station, respectively.

From the BRT partial dependence plot analysis, each variables namely time of the day, wind speed and temperature have a negative relationship with the FPNC level, which shows that an increment in the wind speed has led to decrease the number count of these particles. This finding is in agreement with Yahaya (2013), who conducted a [PNC] study in the city of Leeds, UK. However, the concentrations of SO_2 and NO_X showed different trends, where the FPNC increased with the increment in each gases, especially during night-time (land breeze). In the area under study, these gases are normally produced by either diesel or petrol engines or industrial or construction activities.

The inconsistent fitted partial dependence plot were obtained for the time system (Julian day) and humidity factors. This indicated that these factors are not that important for forecasting the particle number in this case. These fluctuating results also indicate that there are probably other factors at play that were not identified or addressed by this study. The fitted model also showed that t FPNC increased dramatically to the maximum concentration (approximately 25,958 particles/litre) and then remained stable when the wind blew from 180° to 260° (from the mainland). However, the FPNC concentration decreased when the wind blew from 260° to 320°, that is, from the sea transition edge. Thus it can be concluded that most of the fine particles were carried by the wind from the land, probably directly from the agricultural and fishing activities nearby.

It was found that, the relationships between the CPNC and the temperature, pressure and SO_2 variables were all positive. In other words, the CPNC increased steadily with the increase in these two meteorological factors and with increased SO_2. Thus a similar pattern was indicated for both fine and coarse particles in relation to SO_2. However, in contrast to the result for FPNC, CPNC decreased with decreasing wind speed and time of the day. The Julian day showed a fluctuating relationship with the CPNC, while there was an inconsistent relationship between NO_X and the CPNC.

The fitted model also showed that the particle concentrations increased when the wind blew from approximate 75° to 100° (from the sea) and that the CPNC increased dramatically and then remained stable when the wind blew from 100° to 150° (sea transition edge) at a high concentration (70 particles/litre). Then the concentration decreased when the wind blew from 150° to 200° (from the land). This indicates that most of the coarse particles were carried by the wind from the 'sea', probably directly from the crystallisation processes acting on the sea salt spray.

3.6 *Relative Influence Variables*

The variable influence or relative importance of the decision tree ensembles was based on the decision tree influences as described by Breiman *et al.* (1983) and was then proposed by Freidman (2001). The decision tree influences was then implemented in the gbm package. The relative influence represents to what extent the response (dependent) variable is influenced by the predictor (independent) variables. The relative influence of each variable obtained from a BRT analysis is scaled such that the total is expressed as a percentage. Figure 3 shows the plots and values of the variables that influenced the [PNC]s obtained from this analysis. Figure 3(a) shows that SO_2 influenced almost half (46.53%) of the FPNC value, with the Julian day having the second highest influence of, 13.71% followed by the wind direction at 10.5%. The other parameters, namely, pressure, NO_X, temperature, humidity, time (hour) and wind speed, accounted for percentage values of less than 10% each, and therefore exhibited much lower influence on the FPNC. As for the CPNC value, Figure 3(b) shows that wind speed, wind direction and Julian day influenced this type of particle the most, at 22.33%, 18.89% and 18.17%, respectively. The other variables, SO_2, temperature, humidity, NO_X, and time (hour) had less than a 10% influence each on the CPNC.

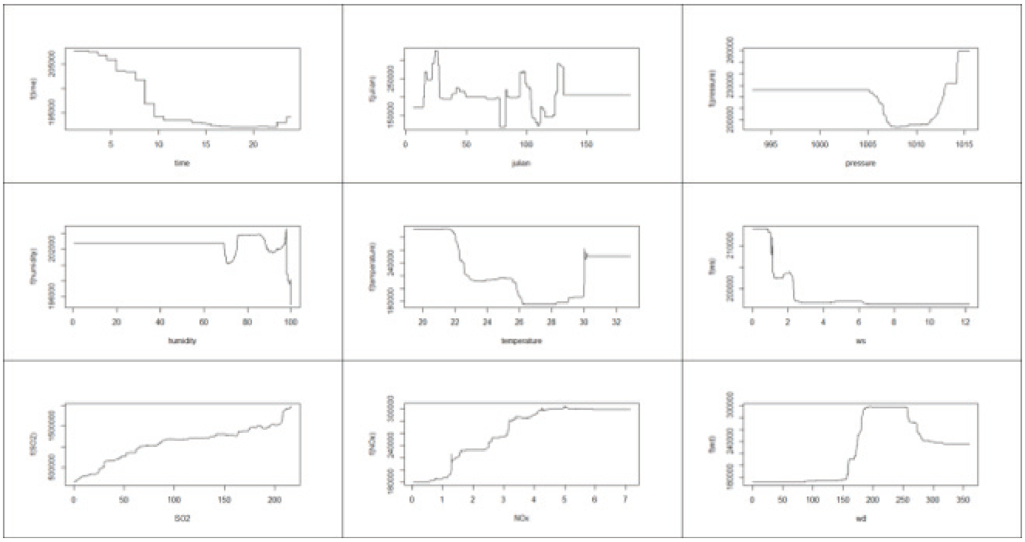

Figure 3. Partial dependence plot for FPNC and dependent variables

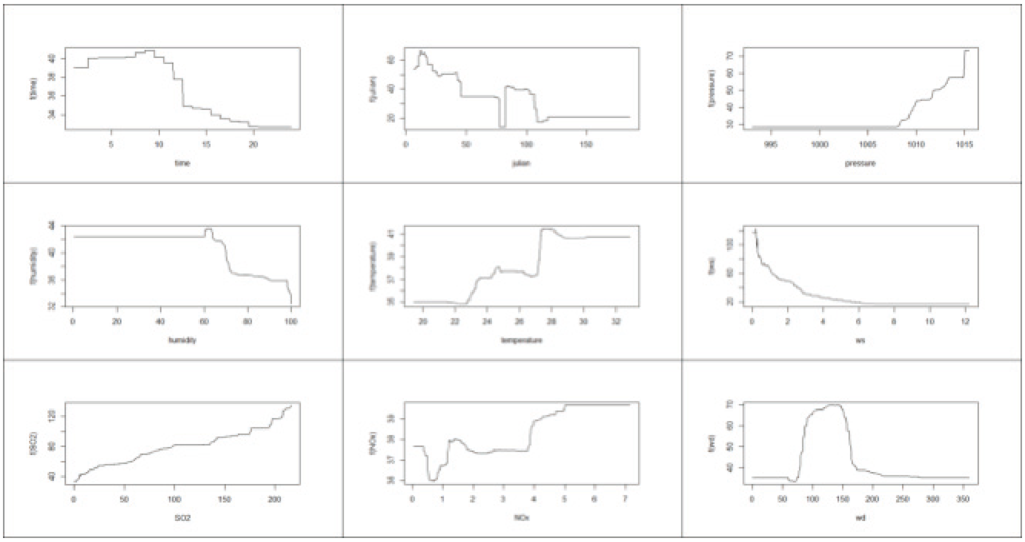

Figure 4. Partial dependence plot for CPNC

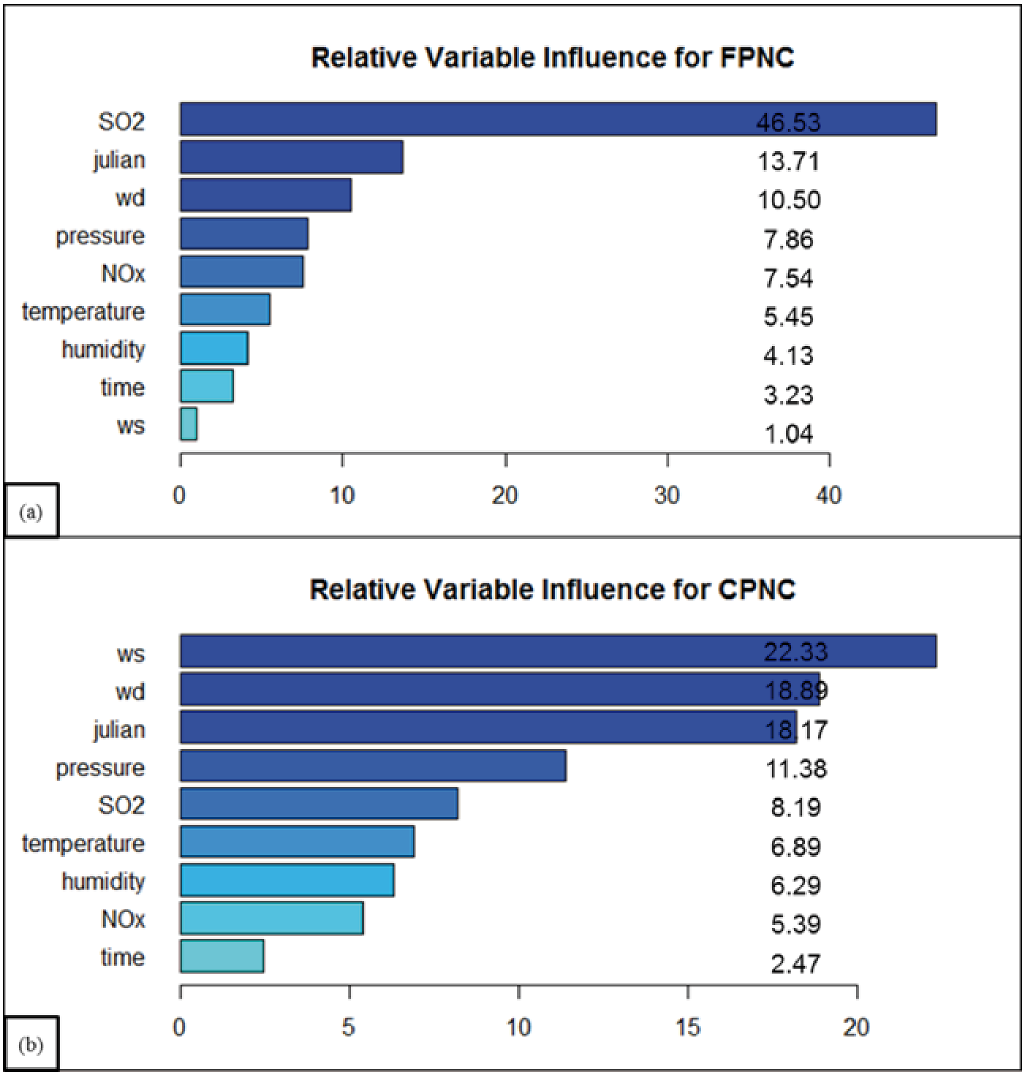

Figure 5. (a) relative variable influence for FPNC; (b) relative variable influence for CPNC

3.7 Interaction between Variables

The BRT technique can also identify the interactions between the variables and these can be represented graphically by considering all possible pair-wise variable combinations in turn by using the plot.gbm command in the gbm package. For the purpose of this study, the interactions between the most important variables, namely, SO_2 and Julian day for the FPNC, and wind speed and wind direction for the CPNC were examined. The fitted predictor data were pulled from the particle boosting model further investigated by plotting contour lines by using the Akima and plotrix package. Akima 2D interactions illustrate the interaction between two selected parameters. In this case, the top two variables that influenced the two types of [PNC] were selected for further analysis.

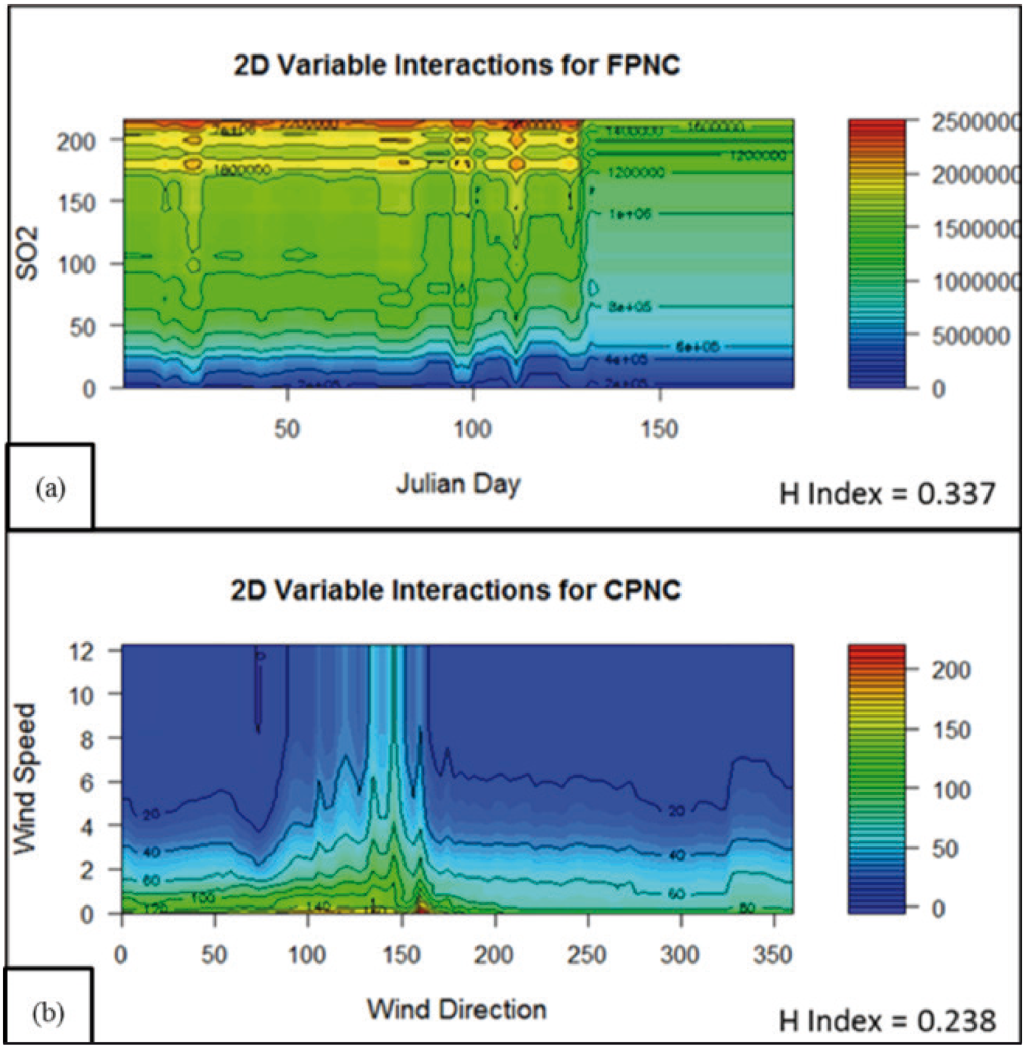

Figure 6. (a) Akima 2D interaction between variables for FPNC (SO_2 and Julian Day); (b) Akima 2D interaction between variables for CPNC (wind speed and wind direction)

The relative strength interaction effect index (H-Index) illustrates the degree to which the predictors interact in determining the response of the variables. This can be done by computing the strength of the interactions between two selected variables by applying the Friedman's H statistics method using a nonlinear model (Friedman and Popescu, 2008). The H-values range from 0 to 1. The value of zero means that there are no interactions between the variables, while the closer the value is to 1, the stronger the interaction between the variables.

Figure 6(a) shows the Akima 2D interactions for FPNC. The two variables that had the most influence on the FPNC, SO_2 and Julian day, show a good relation with the H-Index value 0.337 along with the FPNC. The high FPNC indicated by the red colour shows that the value of SO_2

exceeded 200 ppb during the early hours of the Julian day. In contrast, where the H-Index for the wind speed and wind direction was found of 0.238 in value as shown in Figure 6(b). From the plot, the high CPNC was brought about by the slow speed of the wind (between 0 and 2 m/s) which came in from the ocean to the west.

CONCLUSION

This paper presented the results of an analysis of the use of the boosted regression tree model as a statistical tool to predict the number count of fine and coarse particles (FPNC and CPNC) in a coastal area on the east coast of Malaysia during the period of 5 January to 6 July 2015. The used of BRT as a tools for forecasting particles is an advance approach in which the model development processes promised the best model fitting that can suit most data type. Moving from traditional way to an artificial intelligent approach provides an advance analysis and gives a better way to understand the pattern, model fitting and interactions between variables. The understanding of these fundamental sciences of particles especially at the coastal environment is important that can be used for prediction purposes. The aim of the study was to investigate the influence of two gases and selected meteorological variables on these two types of [PNC] by using the boosting technique. Based on the result of the evaluation of the performance of the predictive models developed for FPNC and CPNC, both of the models achieved a good fit between the observed and predicted values, where more than 90% of the variation in these two types of [PNC] was explained by the explanatory variables. Further work can be done by including another station at different location in Peninsular Malaysia such as at the West Coast, South Malaysia and also at the mid land of Malaysia. Comparison between different types of artificial intelligent analysis may also interesting to give a better understanding and explore more options to analyse big data which currently adopted in current world.

ACKNOWLEDGEMENT

This work was supported by the Ministry of Higher Education Malaysia for funding and supporting the postgraduate student study. The authors also would like to extend sincere thanks you and gratitude to the Institute of Ocean Earth Science of University Malaya, Bachok Kelantan, Malaysia for providing us with data and facilities and University Malaysia Terengganu for supporting this research.

REFERENCES

Air Quality Expect Group (AQEG). Particulate Matter in the United Kingdom. Department of Environment, Food and Rural Affairs, UK. 2005.

Breiman L, Friedman JH, Olshen RA, Stone CJ. Classification and Regression Trees. Belmont, CA: Wadsworth Publishing; 1983.

Carslaw DC, Taylor PJ. Analysis of air pollution data at a mixed source location using boosted regression trees. Atmospheric Environment 2009; 43: 3563-3570.

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

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. Journal of Atmospheric Environment 2001; 35: 6391-6405.

De' ath G., Fabricius KE. Classification and Regression Trees: A Powerful Yet Simple Technique for Ecological Data Analysis. Ecology 2000; 81: 3178-3192.

De'ath, G. Boosted trees for ecological modelling and prediction. Ecology 2007; 88: 243-251.

Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. Journal of Animal Ecology 2008; 77: 802-813.

Environmental Protection Agency (EPA). Air quality criteria for particulate matter. Research Triangle Park: US Environmental Protection Agency; 2004.

Freund Y. Boosting a weak learning algorithm by majority. Inform Computation 1995; 121: 256-285.

Friedman JH. Greedy function approximation: A gradient boosting machine. The Annals of Statistics 2001; 29: 1189-1232.

Friedman JH. Stochastic gradient boosting. Computational Statistics and Data Analysis 2002; 38: 367-378.

Friedman JH, Popescu BE. Predictive learning via rule ensemble. Annals of Applied Statistics 2008; 2: 916.

Hail JR. Applied Geomorphology. Oxford: Elsevier; 1970.

Harrison RM, Deacon AR, Jones MR, Appleby RS. Sources and processes affecting concentrations of PM_{10} and $PM_{2.5}$ particulate matter in Birmingham (U.K.). Atmospheric Environment 1997; 31: 4103-4117.

Kurt H. {R} {FAQ}, Retrieved on 2014, from http://CRAN.Rproject.org/doc/FAQ/R-FAQ.html. 2014.

Janssen NAH, Fischer P, Marra M, Ameling C, Cassee FR. Short-term effects of $PM_{2.5}$, PM_{10} and $PM_{2.5\text{-}10}$ on daily mortality in the Netherlands. Science of the Total Environment 2013; 20-26: 463-464.

Kampa M, Castanas E. Human health effects of air pollution. Environmental Pollution 2008; 151: 362-367.

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: 331-336.

Leathwick JR, Elith J, Franchis 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: 267-281.

Munir S, Habeebullah TM, Mohammed AMF, Morsy EA, Awad AH, 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 2014; 8 (1): 53 – 60.

R Development Core Team. R: A Language and Environment for Statistical Computing. In: R Foundation for Statistical Computing, Vienna, Austria;

2008.

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

Schapire RE. The strength of weak learnability. Machine Learning 1990; 5(2): 197–227

Sharifah MSA. The Coastal Zone in Malaysia. Processes, Issues and Management Plan. Background Paper, Malaysian National Conversation Strategy. Economic Planning Unit, Kuala Lumpur. 1992.

Shridhar V, Khillare PS, Agarwal T, Ray S. Metallic species in ambient particulate matter at rural and urban location of Delhi. Journal of Hazardous Material 2010; 175: 600-607.

Wilks DS. Statistical Methods in the Atmospheric Sciences. 2nd Ed. United State: Elsevier; 2006.

Vakeva M, Hameri K, Puhakka T, Nilsson ED, Hohti H, Makela JM. Effects of meteorological processes on aerosol particle size distribution in an urban background area. Journal of Geophysical Research 2000; 105: 9807-9821.

World Health Organization (WHO). WHO Air Quality Guidelines for Particulate Matter, Ozone, Nitrogen Dioxide and Sulphur Dioxide. World Health Organization. 2006.

Yahaya NZ, Tate JE, Tight MR. Studying particle number noncentrations (PNC) in an urban street canyon: Using boosted regression trees, BRT. Proc. 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). Proc. Presented The European Aerosol Conference 2011, Manchester University. 2011b.

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

Yahaya NZ, Ghazali NA, Ahmad S., Asri MAM, Ramli NA. Analysis of Daytime and Nighttime Ground Level Ozone Concentrations Using Boosted Regression Tree Technique. Journal of Environmental Asia 2017; 10(1): 118-129.