Universiti Malaysia Terengganu Journal of Undergraduate Research *eISSN: 2637-1138*
Volume 2 Number 3, July 2020: 57-66

SEASONAL VARIATION OF PARTICLE NUMBER COUNT (PNC) AT THE COASTAL ENVIRONMENT IN PENINSULAR MALAYSIA USING BOOSTED REGRESSION TREES (BRTS)

SOBIRATUL NADIA ABDULLAH, NOOR ZAITUN YAHAYA*, AND WAN RAFIZAH WAN WAN ABDULLAH

Faculty of Ocean Engineering Technology and Informatics, Universiti Malaysia Terengganu, Terengganu, Kuala Nerus, Terengganu, Malaysia.

**Corresponding author: nzaitun@umt.edu.my*

Abstract: The concentrations of airborne particulate matter (PM) is often measured as a mass concentration. However, the other way to express particulate matter is by using the Particle Number Count ([PNC]) concentrations. This study aims to analyse the seasonal variation of airborne particulate matter in terms of [PNC] by using R packages and the Boosted Regression Trees (BRTs) technique. The study was conducted at IOES, Universiti of Malaya in Bachok, Kelantan. The monitoring was important to understand the variability of seasonal effects due to different seasons. In this work, only the datasets for three seasons (Inter Monsoon, North East Monsoon and South-West Monsoon) were analysed involving 25,958 data. The air quality monitoring equipment involved was the particle counter Environment Dust Monitor *GRIMM Model 180* and a weather station for recording the meteorological parameters. The data analysis was completed by using R software and its package for evaluating seasonal variability and providing the statistical analysis. The relationship between variables was studied by using the Boosted Regression Tree (BRT) technique. The interaction between independent variables towards the [PNC] in different seasons was discussed. The best setting result of BRT model evaluation R^2 is 0.22 (North-East Monsoon), 0.87 (Intern monsoon 1), and 0.59 for South West Monsoon which indicated that the model developed is acceptable except for NEM and intern monsoon seasons. Temperature (57 %) and wind direction (67%) were found to be the highest factor influenced by the formation of [PNC] concentrations in this area. Finally, good results indicated that BRT technique is an acceptable way to analysed air pollution data.

Keywords: Particle number count, concentration, seasonal analysis, boosted regression tree

Introduction

Air pollution can be defined as the modification of the natural characteristics of the indoor or outdoor atmosphere due to the contamination by physical, chemical or biological agents. Several air pollutants including the particulate matter (PM), carbon monoxide (CO), nitrogen dioxide (NO_2), sulfur dioxide (SO_2) and ozone (O_3) are considered as the major threats to public health. These pollutants are consistently associated with the global increase in cases of fatal respiratory diseases. Prolonged exposure to the fine particle fractions present in ambient air poses risks to human health even at a low concentration. Hence, monitoring of source apportionment, concentrations and chemical composition of particles concerning their size is essential in minimizing the effects.

The main aim of this paper is to present the analyses of concentrations particle number count [PNC] on the east coast of Malaysia and to understand factors that influenced the concentrations of [PNC]. The location for east coast environment in Malaysia is located at the Institute of Ocean and Earth Science at Bachok, Kelantan, Malaysia which facing the South China Sea has been chosen. This location is selected because there is a facility such as instruments and machines needed to detect [PNC].

Particle Matter (PM) is a combination of solid and liquid particles in varying number, physical and chemical properties, and origin (Pope and Dockery, 2009). Airborne PM which includes a wide range of particle size (coarse, fine particles and ultrafine) is usually expressed in terms of its aerodynamics diameter and it is often categorized as PM_{10} and $PM_{2.5}$ and PM_1. Otherwise, the presence of PM can also be described according to [PNC] which represents the number of particles in the atmosphere per cubic centimetre (#/cm³). The particle can also be explained or described according to Number Count (Number concentration, [PNC]) to quantify the characteristics of particles in the atmosphere expressed in particles per cubic centimetre (#/cm³). Condensation nucleus counter counts the total number of particles and defined size range from 3nm – 10 µm.

An aerosol can be physically described by mass, surface area or number concentrations as a function of particle size. Airborne PM includes a wide range of particle size (coarse, fine particles and ultrafine) is usually expressed in terms of its aerodynamics, as the diameter of a unit density sphere that has the same settling velocity as the particles. PM is most often categorized as PM_{10} and $PM_{2.5}$ and PM_1 (AQEG, 2005). Large particles are sized from 2.5µm u to 30µm aerodynamic size diameter. Airborne particles are considered to have three size models, (Harrison, 2000) divided the airborne particles into three distinct models, which are divided according to size, namely, transient mode of nucleation mode, accumulation mode and coarse particle mode. Particle size 0.1 µm in aerodynamic diameter are called the nucleation mode mainly formed from diesel vehicles, metal smelting operations, condensation of hot vapour from combustion sources and from chemical conversion of gases to particles in the atmosphere. Nucleation mode particles are emitted to the atmosphere as primary particles from combustion sources, both stationary and mobile.

The weather in Malaysia can be divided into two monsoon regimes. The Southwest Monsoon (SWM) starts in late May to September and the weather relatively drier throughout the season.

The following monsoon regime, the Northeast Monsoon (NEM) begins in November until March every year. The NEM brings heavy rainfalls, particularly to the East Coast of Peninsular Malaysia and Western Sarawak. The transition period in between the monsoon is signified as the Inter-monsoon (I1) period. The East Coast is sunny and dry for much of the year. The breeze from the South China Sea regulates the humidity. However, as the NEM strikes between November and February, it brings heavy rainfalls and surges in wave-activity on this side of the peninsular. In this weather, boat crossings and tourism activity to the islands are limited until March.

In addition, airborne particles suspended in the air contain solid and liquid of varying size and chemical compositions. The presence of this pollutant in the atmosphere gives rise to many serious respiratory problems. $PM_{2.5}$ and PM_1 which can easily penetrate the airways and lungs worsen the heart and lung diseases. Elderly and young children are the most at-risk group. The increase in the number of premature death is also linked to exposure to a high concentration of PM (Linda, 2009). In general, East Coast in Malaysia normally receives rainfall during Northeast Monsoon. However, during the Southwest Monsoon and monsoon transition, the state still receives rainfall rates are relatively high. By obtaining the raw data from the Institute of Ocean & Earth Science (IEOS), Bachok Kelantan located at the East Coast Malaysia is doing a comparison of Particulate Matter (PM) concentration, meteorological aspects are the main purpose to be considered in this study. Different seasonal trends may develop a different graph of Boosted Regression Trees algorithm.

Recent research trends demonstrated that a Boosted Regression Trees (BRT) model with stochastic approach can be used to obtain the best model from a high level of complexity and large datasets (Carslaw & Taylor, 2009; Yahaya *et al*., 2011a; Yahaya *et al*., 2011b; Yahaya, 2013; Noor Zaitun *et al.,* 2018; N.Yahaya *et al.,* 2019). Friedman (2001) developed the gradient boosting algorithm, which called for all of the training data observations to be included in the function estimation process at each iteration.

According to Friedman (2001), no matter how dimensionally large the predictor variable space is, or how many variables are used for the prediction, the model subcomponents can be represented by a two-dimensional graphical representation which can be easily plotted and interpreted. BRTs can be an effective tool to understand the fundamental behaviour of [PNC] through the derivation of a [PNC] boosting algorithm for solving the quantitative regression problems. Then, in 2002, Friedman added a stochastic element to the initial boosting algorithm that involved taking a random sample of observations for each iteration of *t.* Thus the performance of the initial technique was improved by adding an element of randomness to the algorithm and creating the stochastic gradient boosting machine (GBM) or stochastic BRT (Friedman, 2002).

The used of a particle boosting model as a statistical tool for predicting [PNC] formation is a new approach to the analysis of particles data. Compared to other techniques, the BRT models have better predictive advantage owing to their ability to select significant variables, fit precise functions, recognise and model interactions. Besides, the prediction techniques used to determine the relative importance of the different variables in influencing the particles level could indicate the variables that contribute the most to particles pollution. The developed model is expected to be useful for decision-makers especially to marine and coastal people because it can enable them to consider which variables play an important role in high concentrations of pollutants in this area. This can be used to mitigate the air pollution issues such as haze episode. Compared with other traditional techniques, this method helps deal with large or big data and has an advantage furthermore it has an enhancement from using a traditional method such as the least-squares technique or multiple regressions to the newest technique with the IR 4.0 element.

Interpretation of the complex boosting model can be performed with the right tools. Friedman (2001) indicates that interpretation via visualisation is useful for graphical renderings of the value of the derived approximation ((x)). Considering the predictors, x, the variables that are used in BRT model fitting are represented as 'x' while the response variables are represented as 'y'. The interpretation of BRT output can be graphically presented in three different ways, namely, relative influence variables, partial dependence plots between variables and two-way variables interactions.

In the fitting process, identifying the correlation between the response and the predictors is crucial. According to Friedman (2001), boosting applies the concept of the 'variable influence' to replace the regression coefficient. The focus is to identify specific input variables that have the most significant influence on the variation of response. To the extent that (x) qualitatively reflects the nature of the target function $F(x_i)$, such tools can determine the correlation between the variable y and the variable x. Later, Friedman developed an extension of a variable's 'relative influence' for boosted estimation by using tree-based methods (Friedman, 2002). This method estimates the approximation of the relative influence of a variable x_j, which is applied to Equation 1 as

$$\hat{J}_j^2 = \Sigma_{splits\ on\ x_i} I_t^2, \tag{1}$$

where I_t^2 is the incremental improvement by splitting x_j at that point. The extension averages the relative influence of variable x_j across all the trees generated by the boosting algorithm. The measure of variable influence can be determined depending on how many times a variable is selected for splitting and weighted by the improvement in the overall model. Later, Friedman and Popescu (2005) emphasized that the relevance of the respective input variables (x_1, x_2, x_3, ...x_n) to the predictive model is the main interest in a descriptive statistic of predictive learning. The measurement of importance, Jl (x), of input variables, xl, at each prediction point, x, can be represented in Equation 2.

$$J_l = I_l(x) + \Sigma_{xl \in rk} I_k(x)/m_k \tag{2}$$

where $I_l(x)$ is the importance of the linear predictor involving x_l, and the second term sums the importance of the relative variables that contain $x_l(x_l \in r_k)$ where each is divided by the total number of input variables m_k. The relative influence of each variable obtained from BRT analysis is scaled such that the total is expressed as a percentage. The interaction effect is another important feature in BRT. It explains the interactions between variables. To determine the variable interactions, the number of splits (size) of the individual trees is varied and the degree to which predictors interact in determining the response can be identified. If there is no interaction effect between variables, the value will be zero. A larger value corresponds to a stronger interaction effect between the variables. Similarly, if the quantity of H differs from zero to the extent, it indicates that interacts with either one or more other variables. Two-way interactions reflect the interaction between two variables compared with the purely additive combination. For instance, two-way interaction of two variables, predictions were made across regular interval combinations while keeping all other variables at their mean level. In this case, the interaction between variables by graphically can be examined using *Akima* Package.

Materials and Methods

Data Collection Station

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 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. Bachok is categorised as rural where agriculture is the main activity of the residents and where fishing activities take place along the coastal area

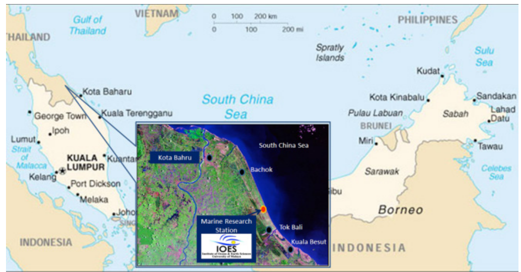

Figure 1: IOES monitoring station at Bachok, Kelantan, Malaysia

Data Collection and Instrumentations.

A one-minute data datasets of the [PNC] and selected meteorological factors (temperature, pressure, humidity, wind speed and wind direction) compiled for a one-year (1st January to 31st December 2016) were obtained from the Institute of Ocean and Earth Sciences (IOES) Station, Bachok, Malaysia. [PNC] were monitored using a particle counter (Model EDM180, GRIMM, Germany) of 31 channels of the [PNC] (Dp = 0.265–34 µm), in units of counts/litre. The EDM180 uses a patented laser with a 90° scattering angle. 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. Data sets were then averaged to 10 minutes average that is used throughout the study to synchronized with meteorological data which saved in 10-minutes average.

The front panel of Environmental Dust Monitor GRIMM180 device has been utilized to detect [PNC] level in the air. The sample air was directed into the measuring cell through the aerosol inlet. Inside the measuring cell, detection of the particles in the sample air was done through light scattering technique. Classification of particle size was determined based on the counted scattering light pulse of every single particle and the scattering light signal intensity. Figure 2 shows the monitoring device used.

Figure 2: Environmental Dust Monitor GRIMM180 source by NAQC-EPA Monitoring Conference Denver, Colorado (2012)

All GRIMM180 laser aerosol spectrometers and dust monitors are important instruments for detecting the particle in the range of 0.25μm – 30 μm. It uses a laser diode (the wavelength in the visible range at 655nm) as a light source. The laser diode can be operated and its intensity is modulated in Multiplex Mode. An illumination optic focuses the laser beam to a flat elliptical strip. The sample air is focused aerodynamically and then led as particle flow through the inner area of the measuring volume. During measurements, the particle concentration of the sample air is normally so low. Thus, only one particle is statically seen in the measuring volume. The SO_2 and NO_X 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 temperature, pressure, humidity, wind speed and wind direction.

Data (n = 25, 958) on the [PNC], SO_2, NO_X and selected meteorological conditions were recorded in a Microsoft Excel spreadsheet and analysed statistically. Two types of [PNC], namely, fine particle number count (F[PNC]; Dp = 0.265–0.900 μm) and coarse particle number count (C[PNC]; Dp = 2.75–9.25 μm) were considered in the analyses. The statistical analysis of air pollution data was performed by using the R programming language developed by the Development Core Group R (2008). The R provides a range of statistical and graphical techniques such as statistical tests, linear and nonlinear modelling, time-series analysis, classification, and others. The first section will present the open-air package developed by Carslaw (2012), which is available at http://www.open-air-project.org/. In the second part, the stochastic boosted regression trees (BRTs) approach and its application in the air pollution fields are reviewed. In this study, the open-air packages (Carslaw and Ropkins, 2013) and the generalised boosting machine (gbm) package (Ridgeway, 2010) were used. The ability of R to explore comprehensive data is proven (Carslaw & Taylor, 2009). The analysis of [PNC] and other variables are presented in the following sections:

i. The seasonal trends of [PNC] concentration in the coastal environment.
ii. The Analysis of [PNC] using Boosted Regression Trees (BRTs)
iii. Factors that influence the seasonal of PM and gaseous

Variables can be classified into a dependent or independent variables. The dependent variables are the variable that depends on the outcome of the independent variables (Yahaya, 2013). In this study, the [PNC] is considered as the dependent variable and the meteorological factors are considered as the independent variables.

Results and Discussion

Table 1 shows a summary of the descriptive statistics of [PNC] in IOES Bachok, Kelantan. [PNC] data has been separated into three seasons which are the Southwest Monsoon (SWM) that occurred during May until August, Northeast Monsoon (NEM) occurred in November until February and Inter-monsoon 1 (I1) occurred in March until April 2016. However, this research had the limit data available therefore only three-season available in this study. The minimum and maximum concentrations for [PNC] were 1658 count/litre and 6,439,013 counts/litre respectively. Besides that, the 1st quartile and

3rd quartile were 69,436 count/L and 290,355 L/minute respectively while the median value of [PNC] was 141,179 L/minute and the mean of [PNC] is 361084 count/L.

The other important variables are meteorological parameters. The minimum and maximum value of the humidity were 50.86% and 89.99%, respectively. The minimum and maximum value of temperature were 18.19 ºC and about 49.99 ºC, respectively. The minimum and maximum value of pressure occur for the three-season were 924.5 Pa and 1020.7 Pa, respectively. Their wind direction rotates around the angle (from 0^o to 360^o) which recorded a minimum value for both wind speed and wind direction, 14.69 (m/s) and 360 (^{o}).

The [PNC] mean concentration was reported lower with 221,267 particles/L and the maximum concentration which was higher with 6,439,013 particles/litre compared to the value reported in Yahaya et al., (2018) in 2015 obtained from the same IOES station. This observation shows that the [PNC] data at the coastal environment shows similar pattern comparing two calendar year.

Table 1: Summary of Descriptive of [PNC] in IOES Bachok, Kelantan

ITEMs	HUMIDITY (%)	TEMP (^{o}C)	PRESSURE (hPa)	WS (km/hr)	WD (^{o})	[PNC] (#/L)
Minimum	50.86	18.19	924.5	0	0	1658
1st Quant.	72.69	23.87	1004.8	1.28	56.28	69436
Median	76.89	25.81	1006.3	1.98	169.54	141179
Mean	77.44	25.81	1006.3	2.782	147.27	361084
3rd. Quant.	82.63	27.9	1007.7	3.59	222.8	290355
Max	89.99	49.99	1020.7	14.69	360.00	6,439,013

Figure 2 shows further details of the [PNC] temporal trends of the [PNC] concentrations through-out the sampling period

Further analysis was conducted by plotting the call time variation plots and the plots are shown in Figure 3. The time variation function in R is a useful function to demonstrate the temporal variation of variables, averaged either by the time of day or day of the week. The advantage of time variation plot is it has a 95% confidence interval in the mean. Normalized variable data is a result of each data has been divide by its means.

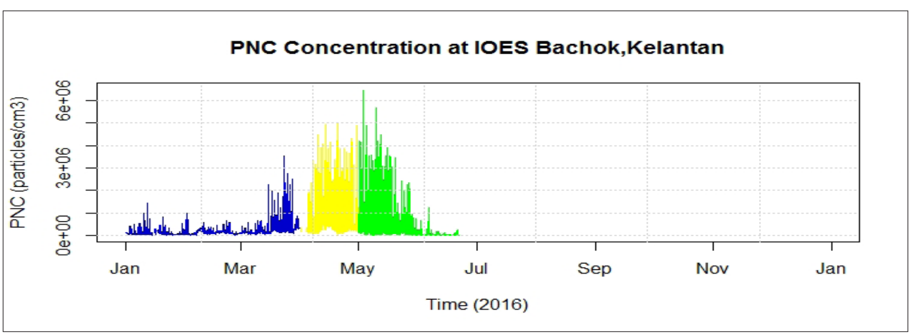

Figure 2: [PNC] concentration at IOES Bachok, Kelantan from 01 January 2016 to 31 December 2016

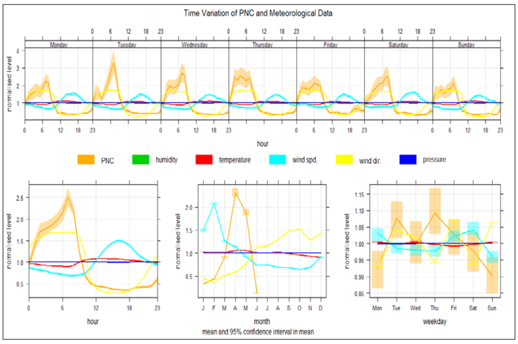

Figure 3: The time variation plot of [PNC] in an hour, month and weekday in IOES Bachok, Kelantan

A BRT algorithm was fitted to the hourly [PNC] concentrations for each dataset to investigate how coastal [PNC] concentrations are influenced by ambient particulate matter and meteorological conditions (wind speed, wind direction, temperature, and humidity). In the first stage of the BRT analysis, the model setting parameters such as shrinkage and learning rate, interaction depth in the BRTs algorithm was tested and determined by testing 25 combinations of a different number of trees, shrinkage and tree complexity and model performance checking were performed by comparing between observed and modelled and also using an error bias method. It was found that the best iterations for the shrinkage was at 0.001 and the best setting for tree complexity was at 5. The modelled [PNC] data has been compared with the observed data in the error bias method and to select and prove the best algorithm setting can model the [PNC] data with minimum error. Besides, the number of trees for each season is 10000.

It was found that the Fraction of Prediction of 2 (FAC2) for each season was about 0.98 (SWM), 0.77 (NEM) and 0.86 (I1) which fall within an acceptable value (range of 0.5 to 2.0). The Pearson correlation (R) from the best model were 0.75 for I1, 0.58 for SWM marked acceptable correlation values while 0.21 value for NEM which were in the lower range.

The relative importance of the predictor variables was determined using formulae developed by Friedman (2001) which was implemented in the *gbm* in *R package*. In *gbm*, the measure is based on the number of times a variable is selected for splitting and is weighted by the improvement in the overall model. The relative importance analysis focuses on the most important variables obtained from the *gbm* iteration, which omitted the time system and in-street airflow (airflow velocity and direction). These variables (prevailing wind speed and directions, traffic flow and speeds, and meteorological variables) have also been commonly used by other air pollution studies and are known to influence the [PNC] and air pollution data; however, no studies have reported which variables have the most influence on the particles. Table 2 shows the relative influence of independent variable extracted from the BRT output.

Table 2: Relative importance of the influence variables at the IOES, Bachok with a different season in percentage (%)

Parameter	South West Monsoon (%)	Northeast Monsoon (%)	Inter Monsoon 1 (%)
Pressure	17	9	8
Temperature	38	52	18
Wind speed	5	9	7
Wind direction	40	30	67

In general, the wind direction factor was found to be the highest factor with 67% that influenced to the concentration of [PNC] during I1 season while temperature factor was the highest percentage that influenced during dry NEM with 52%. Generally, during this season, the East Coast environment experienced heavy rainfall. This may be due to the weather during the NEM which is wet and bring heavy rainfall. Therefore, the temperature is likely the highest factor influencing the [PNC] compared to other seasons.

Figure 4 shows that the wind speed is the highest relative influence which is 40.98% during SWM. The SWM is the drier season in the East Coast of Malaysia as most areas receive minimum monthly rainfall and the wind direction became the highest influence may be due to the weather. Although, the study area is influenced by the wind circulation system of the seasonal monsoon where the southwesterly wind is dominated the dry seasons and it can be characterized by comparatively stable atmospheric conditions in the equatorial region. Figure 4(a-c) show the relative influences of the independent variables on the [PNC] level during three different seasons.

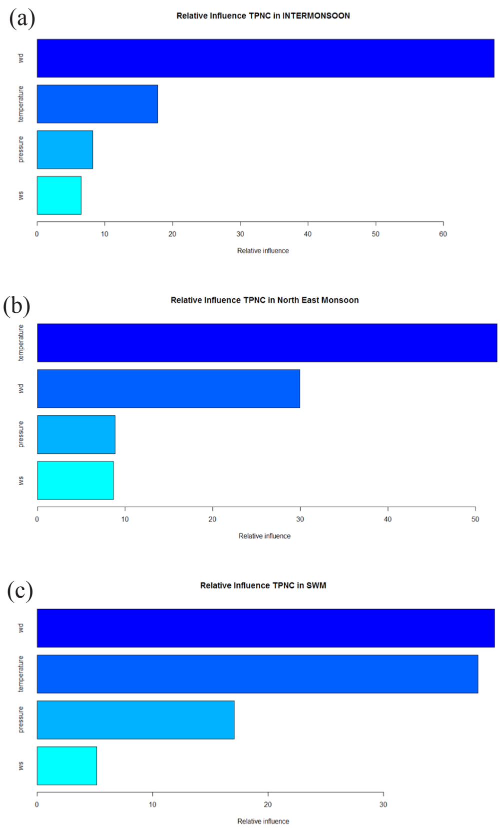

Figure 4: Relative importance of the independent variable at IOES Bachok, Kelantan

The visualization of two – ways variable is plotted by using AKIMA package to examine the interaction between variables. A 2D variable interaction of temperature and relative humidity was plotted to examine the interaction index between two variables and demonstrated how these two variables affected [PNC]. In addition, 2D variable interaction of wind speed and wind direction was also plotted to predict how these two variables affected [PNC] at IOES, Bachok, Kelantan.

Figure 5 shows the two ways interaction between wind direction and wind speed for the [PNC] at IOES, Bachok in during (a) I1, (b) NEM and (c) SEM. There are two areas with the highest [PNC] for NEM for instance. The level of [PNC] was produced at 230,000 counts/minute with the wind direction starting at 200° to 260 ° with the wind speed of 2 m/s. The SWM season recorded the maximum level of [PNC] produced at 1,200,000 count/litre with the wind direction starting at 150 °C to 250 °C and when the wind speed was between 1.5 – 3.0 m/s. Similarly, higher [PNC] concentrations were observed during NEM when the wind blew from the land which can be resulted from land activity such as agriculture activity contributed to a high level of particles in this areas. Similar results were also reported by Yahaya et al., (2018).

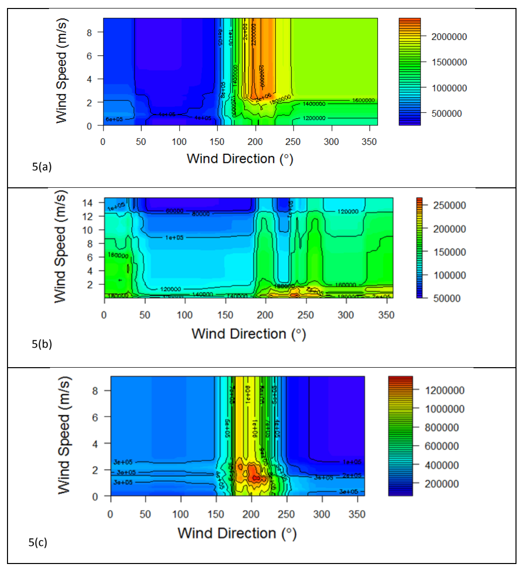

Figure 5: The two ways interaction between wind direction and wind speed for the [PNC] at IOES, Bachok in during (a) I1, (b) NEM and (c) SEM

Conclusion

In conclusion, the findings show that the [PNC] has significantly increased during the South West Monsoon and Inter Monsoon 1. The increased in [PNC] may due to the open burning of the recent period and human or agricultural activities around the area. The highest relative influence at Bachok for North East Monsoon was the humidity, which was 52%. For the Inter Monsoon 1 season, the wind direction was found to be the highest influence factor to [PNC] which was 67% and during the South West Monsoon season, the wind direction was the highest contributor with more than 40% influenced the concentrations of [PNC]. The interaction between predictor variables to the [PNC] also has been analyzed in the partial dependence plot and two ways predictors plot to able explain the interaction between the influenced variables to [PNC] in Bachok, Kelantan. The findings show that the R-squared (R^2) value for BRT model was 0.21 for North East Monsoon (NEM), 0.75 for Intern Monsoon 1 (I1), and 0.58 for South West Monsoon (SWM), respectively. The model developed is considered acceptable for Intern Monsoon 1. It was determined that the Temperature (57 %) and wind direction (67%) were the most significant factor influencing to the value of [PNC] concentrations in the studied area.

The used of particles boosting model as a statistical tool for predicting [PNC] formation is a new approach to the analysis of particles data. The BRT model has many advantages compared with other techniques, such as its capability to select relevant variables, fit accurate functions, and automatically identify. Besides, the prediction techniques used to determine the relative importance of the different variables in influencing the particles level could indicate the variables contributing the most to particles pollution. The developed model is expected to be useful for decision-makers especially to marine and coastal people because it can enable them to consider which variables play an important role in high concentrations of pollutants in this area. This can be used to mitigate the air pollution issues such as haze episode. Compared with other traditional techniques, this method helps deal with large or big data and has advantages furthermore it has an enhancement from using a traditional method such as the least-squares technique or multiple regressions to the newest technique with the IR 4.0 element.

Acknowledgements

The authors would like to thank Universiti Malaysia Terengganu (UMT) for the opportunity in supporting the implementation of this study. Not to forget The Institute for Ocean and Environment Sciences (IOES), Universiti of Malaya for instrumentations and data collections.

References

AQEG 2005. (2005). Particulate Matter in the United Kingdom. *Department for Environment, Food and Rural Affairs, UK.*

Carslaw, D. C., & K. Ropkins, (2012). Openair – an R package for air quality data analysis. Environmental Modelling & Software.

Carslaw, D.C. & Taylor, P.J. (2009). Analysis of air pollution data at mixed source location using boosted regression trees. *Atmospheric Environment, 43*, 7053-7063.

Chen R. J., Chen B. H., & Kan H. D. (2010). A health-based economic assessment of particulate air pollution in 113 Chinese cities. *China Environmental Science*, *30*(3), 410–5.

Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics, 29*, 1189-1232.

Friedman, J. H. (2002). Stochastic gradient boosting. *Computational Statistics & Data Analysis, 38*, 367-378.

Harisson, R. M. (2000). Studies of the source apportionment of airborne particulate matter in the United Kingdom. *Journal of Aerosol Science, 31*, 106-107.

Linda Smith. (2009). Ambient Air Quality Standards (AAQS) for Particulate Matter. California Environmental Protection Agency. *Air Resources Board*.

Noor Zaitun Yahaya, Siew Moi Phang, Azizan Abu Samah, Intan Nabila Azman & Zul Fadhli Ibrahim. (2018). Analysis of Fine and Course Particle Number Count Concentrations Using Boosted Regression Tree Technique in Coastal Environment. *Journal of Environment Asia, 11*(3), 221-234.

Yahaya, N., & Zulfadhli, I. (2019). The used of the Boosted Regression Tree Optimization Technique to analyse an Air Pollution data. *Journal of Recent Technology and Engineering (IJRTE), 8*(4), 1565-1575.

Pope III, C. A., Ezzati, M., & Dockery, D. W. (2009). Fine-particulate air pollution and life expectancy in the United States. *New England Journal of Medicine*, *360,* 376–86.

R Development Group. (2008). R: A Language and environment for Statistical Computing. *In:* Computing. R. F. F. S. (ed.). Vienna, Austria.

Ridgeway, G. (2010). GBM: Generalized Boosted Regression Models. . *R Package Version 1.6-3.1.*

Yahaya, N. Z., Tate, J. E., & Tight, M. R. (2011a). Studying Particle Number Concentrations (PNC) in an Urban Street Canyon: Using Boosted Regression Trees BRT. Proc. The International Conference on Humanities, Social Sciences and Science Technology 2011, Manchester University UK.

Yahaya, N. Z., Tate, J. E., & Tight, M. R. (2011b). Analyzing Roadside Particle Number Concentrations using Boosted Regression Trees (BRT). Proc. Presented The European Aerosol Conference 2011, Manchester University.

Yahaya, N. Z. (2013). Spatial and Temporal variations of ultrafine particles in urban Environment. Phd Thesis Institute for Transport Studies, University of Leeds, United Kingdom.