

PAPER • OPEN ACCESS

Boosted Regression Tree (BRT) model for PM₁₀ concentrations prediction in Malaysia

To cite this article: R. Norazrin *et al* 2023 *IOP Conf. Ser.: Earth Environ. Sci.* **1135** 012041

View the [article online](#) for updates and enhancements.

You may also like

- [The broken ray transform: additional properties and new inversion formula](#)
Michael R Walker and Joseph A O'Sullivan
- [Moisture stress of a hydrological year on tree growth in the Tibetan Plateau and surroundings](#)
Keyan Fang, David Frank, Yan Zhao et al.
- [Bi-Ronchi-test proposal for wavefront sensing and aberration analysis based on the differential form of the Zernike polynomials](#)
J A Arriaga Hernández, B T Cuevas Otahola, A Jaramillo Núñez et al.



Free the Science Week 2023 April 2–9

Accelerating discovery through
open access!

 www.ecsdl.org [Discover more!](#)

The banner features a dark blue background with a futuristic, glowing blue interface. A hand is shown pointing at a central circular element that contains a white padlock icon, symbolizing open access. The text is in white and light blue, with the ECS logo and website URL in white.

Boosted Regression Tree (BRT) model for PM₁₀ concentrations prediction in Malaysia

Norazrin R.^{1,2,3}, Hazrul Abdul Hamid² and Ahmad Shukri Yahaya³

¹ Environmental Engineering, Faculty of Civil Engineering Technology, Universiti Malaysia Perlis, Kompleks Pusat Pengajian Jejawi 3, 02600 Arau, Perlis, Malaysia

²School of Distance Education, Universiti Sains Malaysia, 11800 Gelugor, Penang, Malaysia

³School of Civil Engineering, Engineering Campus, Universiti Sains Malaysia, 14300 Nibong Tebal, Penang, Malaysia

norazrin@unimap.edu.my

Abstract. Air pollution in urban areas is a highly complex problem, displaying strong seasonality and dependence on meteorological factors. Urban particulate matter with an aerodynamic diameter less than 10 μm (PM₁₀) has been identified as one of the most hazardous air pollutants to human health due to the fact that its size range overlaps with that of respirable particles. Models of air quality forecasting are used to provide forecasters with numerical guidance for issuing particulate matter concentration forecasts to human health exposure in a timely manner. The aim of the study was to propose a Boosted Regression Tree (BRT) model for predicting PM₁₀ concentrations in the short term. Multiple Linear Regression (MLR) and Boosted Regression Tree (BRT) models for short-term PM₁₀ predictions are provided, and performance indicators (IA, R², RMSE, MAE, and MAPE) are used to find the appropriate model. The Department of Environment Malaysia (DOE) provided seventeen years of daily average air quality monitoring data, including eight parameters (PM₁₀, wind speed, temperature, relative humidity, NO₂, SO₂, CO, and O₃) and five monitoring stations (Perai, Shah Alam, Nilai, Larkin, and Pasir Gudang). The BRT model gave good results for predicting the PM₁₀ concentrations for each station. The results indicated that for the Perai monitoring station (R² = 0.774), Shah Alam monitoring station (R² = 0.813), Nilai monitoring station (R² = 0.792), Larkin monitoring station (R² = 0.817) and the Pasir Gudang monitoring station (R² = 0.810). According to the findings, the BRT model should be employed in air pollution research, where it is projected to outperform other methodologies in terms of predictive performance. The findings would enhance the existing air pollution research and forecasting approaches.

1. Introduction

PM₁₀ was emitted in metropolitan areas by road automobiles, which may be split into two principal sources: emission of exhaust and non-exhaust pollution. The pollution caused by exhaust pollution consisted of particles of carbonaceous carbon, whereas the non-emission of exhaust was caused by the wear and tear of the tyres, clutch, and brakes, which produced particulate matter [1, 2]. The air quality, as assessed by the amounts of primary atmospheric contaminants has deteriorated considerably in recent years and due to Asia's rapid economic progress, is expected to worsen during the next 40 years. Air quality in the city is deteriorating because of the rise of industry and cities, according to the World



Health Organization (WHO). In metropolitan areas, the majority of air pollution is caused by emissions from motor vehicles. Road traffic contributes between 50 and 85 percent of particulate matter pollution in Metro Manila, Philippines, according to a survey. In the previous decade, the concentration of particles less than 10 μm in aerodynamic diameter (PM_{10}) in metropolitan areas dramatically surpassed the WHO's annual mean threshold of 20 $\mu\text{g}/\text{m}^3$ as a 24-hour mean [3].

Transportation growth and the use of fossil fuels for energy generation can benefit the economy. However, one of the primary challenges is controlling and monitoring emissions [3]. PM_{10} concentrations in Asian and Pacific regions continue to be the sources of the most air pollutants problems [4] categorized as the most damaging pollutant in Peninsular Malaysia and Southeast Asia [5, 6]. The high PM_{10} emissions were directly proportionate because to the expansion of industry and the quantity of vehicles on the road, as a result of which air pollution has increased [3]. It has been proven that high PM_{10} concentrations have negative effects on agriculture, environmental degradation, and biodiversity [7, 8, 9].

In 2016, Malaysia launched MyCREST (Malaysia Carbon Reduction and Environmental Sustainability Tool) to analyze carbon emissions and the long-term effects of the built environment. The interim goal will strengthen the concentration limit for air pollutants, and the standard will be fully implemented in 2020. One of the concerns is to ensure that Malaysia's air quality is good and that the new standards will reduce environmental pollution. There is growing concern that rapid industrial planning, anticipated economic expansion, and how many people, cars, and businesses there will expand as development proceeds, thereby posing environmental challenges and possibly degrading Malaysia's air quality [3, 10].

Predicting the future Malaysian PM_{10} concentrations requires predictive modelling [11]. Prediction models are an important instrument since they are designed to decrease the model's autocorrelation or inaccuracy. The statistical modelling has the potential for great precision in predicting PM_{10} concentrations [12]. To forecast the Air Quality Index, researchers have employed numerous models, some of which are described in the literature [13]. Multiple studies have utilised multiple linear regression (MLR) as a prediction technique due to its ease of implementation and calculation. [10, 14, 15, 16, 17, 18]. Linear models are simpler to understand than nonlinear models. Having a model that can be deciphered gives the user confidence [16]. In regression approaches, the statistical research included meteorological data and air pollution monitoring data as predictive tools. One of the modelling techniques employed is the use of a multiple linear regression (MLR) model to predict PM_{10} concentrations [10, 14, 15, 16, 19, 20].

The conventional methods based on statistics and probability are very complex and inefficient. Predictive models based on machine learning (ML) are more accurate and consistent [13]. Consequently, a short-term forecast must be established and utilized as a rapid warning system to inform the public of harmful air pollution events and to modify air pollution control strategies [21]. Therefore, it is essential to forecast the short-term PM_{10} concentrations level in Malaysia using statistical model forecasting. Yahaya et al. [22] examined the application of boosted regression trees (BRT) to predict daytime and nighttime ozone formation in Kemaman, Malaysia. According to the results, wind speed, wind direction, relative humidity, and temperature were the most influential parameters on ozone formation. The advantage of the BRT model is that it can use a bag fraction of the training dataset to improve prediction performance by randomly selecting parameters. The BRT models for predicting PM_{10} concentrations have a lack of literature. As an alternative to statistical model analysis, the BRT model was designed to forecast Malaysia's short-term PM_{10} concentration. The BRT method was used to construct a model for forecasting PM_{10} concentrations level in Malaysia employing a new methodology to air pollution, specifically for PM_{10} concentrations. The MLR model as a short-term forecasting tool is also included in the study for comparison with BRT models.

2. Methodology

The study focuses solely on a single pollutant, defined as particles less than 10 μm in aerodynamic diameter (PM_{10}). A total of five air monitoring stations were obtained from the Malaysian Department of Environment in the cities of Perai, Shah Alam, Nilai, Larkin, and Pasir Gudang with actual daily average monitoring data for ground-level air quality from 1999 to 2015. The study's dependent parameter is particulates with an aerodynamic diameter less than 10 microns (PM_{10} ; g/m^3), while the independent parameters are sulphur dioxide (SO_2 ; ppm), nitrogen dioxide (NO_2 ; ppm), ground-level ozone (O_3 ; ppm), carbon monoxide (CO ; ppm), relative humidity (RH; %), temperature (T; $^\circ\text{C}$), and wind speed (WS; km/hr). Eighty percent the data were employed for training, while twenty percent were used for validation. Boosted Regression Trees (BRT) and Multiple Linear Regression (MLR) were utilized to predict the PM_{10} concentrations level based on the daily average PM_{10} concentrations.

2.1. Boosted Regression Trees (BRT)

The BRT was selected because the BRT model may be employed without making any assumptions about the atmospheric processes involved [23]. To construct air quality models, it is necessary to assemble a number of simpler models capable of simulating complicated interactions between parameters and able to increase the accuracy of forecast by using randomly selected parameters [23, 24].

2.1.1. BRT Model. Each node of a decision tree involves a different option an attribute (feature), each rule is represented by a branch, while each leaf represents a potential result. BRT are one of a number of techniques designed to enhance the efficiency of a single model by using several models together. A complete version of the BRT model is simple trees are fitted to variables in a forward and sequential way in an additive regression model [25]. BRT utilised two methods: i. Trees of Regression (models in which recursive binary division is used to establish relationships between dependent and independent factors) ii. Improving by boosting (used a large number of simple models in combination to boost prediction ability). A boosting ensemble approach which a group of predictors is generated consecutively rather than individually. Ensemble is the combination of numerous models as the basis model to achieve a better outcome with the way of enhancing. The goal is to define and reduce a loss function. The formula for the mean square error is:

$$\text{Mean Square Error} = \frac{1}{n} \sum_{i=1}^n (PM_{10,D1} - \hat{PM}_{10,D1})^2 \quad (1)$$

where, $PM_{10,D1}$ = Observation data, $\hat{PM}_{10,D1}$ = Prediction data

Improving by boosting approach leverages the reasoning that future forecasters learn from the errors of their predecessors. The predictors can be selected from several models, such as decision trees and regression coefficient. The forecasters are gaining knowledge from the errors committed by their predecessors; as a result, it takes fewer time and iterations to arrive at accurate forecasts [26]. There are three major steps involved in obtaining the BRT model and analyses. Data collection and preparation is the initial step. The second step is the investigation and creation of the dataset's boosting algorithm. The BRT output, interpretations, and algorithm settings are determined in the third stage. BRT makes use of aspect of chance to increase performance of using prediction randomly picked bag portion of the training dataset to fit each successive tree. The algorithm settings for the primary model input parameters are chosen in order to create the BRT model (learning rate, trees number and interaction depth) [26, 27]. Table 1 provides a description of the terminology utilised in BRT analysis.

Table 1: Brief summary of the Boosted Regression Tree (BRT).

Notation	Function
Shrinkage	Small values would enhance prediction performance by minimizing the influence of additional fitted trees. (Ranges from 0 – 1)
n.trees	Error could be reduced by increasing the number of trees. (Range 100 – 5500)
Interaction depth	The number of splits that began with a single node. (Between 1 and 5)
Bag fractionation	Standard expansion of the following tree is 0.5. fifty percent of the training data is chosen at random.
n.minobsinnode	Minimum number of samples in terminal nodes of a tree (10)

2.2. Multiple Linear Regression (MLR)

Linear models are simpler than nonlinear models and straightforward to interpret, hence the MLR model was utilized [16] and deliver accurate short-term and long-term PM₁₀ concentration predictions [10, 14, 15, 19, 20, 28]. The next day's PM₁₀ concentration level (PM_{10,D1}) is predicted using MLR. PM₁₀ concentration level on the same day, temperature, relative humidity, wind speed, nitrogen dioxide, sulphur dioxide, carbon monoxide, and ground-level ozone are the model's parameters. The stepwise procedure was utilized, which is a variant of the forward selection in which all variables in the model are examined to determine if their significance has been decreased to less than the tolerance level and non-significant variables have been deleted. In stepwise regression, there are two degrees of significance: one is used to insert new variables while the other eliminates existing ones.

2.2.1. MLR Model. Multiple regression models are related to linear regression ones, except that they include additional factors and can be used to propose more complex linkages than a straight line [29]. The random parameters PM₁₀ concentration level on the same day, temperature, relative humidity, wind speed, nitrogen dioxide, sulphur dioxide, carbon monoxide, and ground-level ozone are independent, while the next day PM₁₀ is dependent. Then, the model of multiple linear regression is expressed by

$$PM_{10,D1} = \beta_0 + \beta_1 PM_{10,D0} + \beta_2 Temp_i + \beta_3 RH_i + \beta_4 WS_i + \beta_5 NO_{2,i} + \beta_6 SO_{2,i} + \beta_7 CO_i + \beta_8 O_{3,i} + \varepsilon_i \quad i = 1, 2, \dots, n \quad (2)$$

Where $\beta_0, \beta_1, \dots, \beta_8$ are the regression model coefficients, ε_i is the term for random error. The random error term's assumptions are the average of ε is 0, $E(\varepsilon) = 0$. For all independent parameter configurations (PM_{10,D0}, T, RH, WS, NO₂, SO₂, CO and O₃) the variance of ε is constant, $V(\varepsilon) = \sigma_c^2$. The random errors have a normal distribution and are uncorrelated.

2.3. Performance Indicators (PI) The performance of the model is determined by computing the indicators of performance. The PI used were root mean square error (RMSE), mean absolute percentage error (MAPE), coefficient of determination (R²), index of agreement (IA), and normalized absolute error (NAE). A model for the prediction of PM₁₀ concentrations in Malaysia was chosen based on the evaluation of the obtained result.

3. Results and Discussions

Predicting of air pollution is a valuable approach for urban air quality management. Utilizing simulation models is prevalent, resulting in a range of ways. Predictions of PM₁₀ concentrations can be made using a variety of methods.

3.1. Descriptive analysis

The daily average of PM₁₀ concentrations data was subjected to descriptive statistics, and the results are helpful in characterising PM₁₀ concentrations at each monitoring site between 1999 and 2015. PM₁₀ levels in Perai, Shah Alam, Nilai, Larkin, and Pasir Gudang are listed in Table 2 along with their descriptive statistics.

Table 2: The 1999–2015 descriptive analysis of PM₁₀ concentrations

Air monitoring station	Perai station	Shah Alam station	Nilai station	Larkin station	Pasir Gudang station
N	6209	6209	6178	6209	6179
Minimum, $\mu\text{g}/\text{m}^3$	13	14	15	13	15
Maximum, $\mu\text{g}/\text{m}^3$	317	587	327	349	461
Mean, $\mu\text{g}/\text{m}^3$	53.77	53.61	59.19	43.17	51.21
Std. Deviation, $\mu\text{g}/\text{m}^3$	24.80	26.89	23.39	18.54	19.54
Coefficient of variation	0.46	0.50	0.40	0.43	0.38
Skewness	1.67	4.64	3.15	4.05	4.65
Kurtosis	5.55	53.38	20.82	37.01	57.04

The region surrounding the Perai station is classed as an industrialized zone. According to Table 2, the average PM₁₀ concentration for the Perai is 53.77 $\mu\text{g}/\text{m}^3$. At the Perai monitoring station, the PM₁₀ distribution is right skewed with skewness of 1.67 and kurtosis of 5.55, indicating a low risk of extreme events. The PM₁₀ concentrations at the Perai monitoring site exhibit only slight scatter around the mean, with a standard deviation of 24.80 $\mu\text{g}/\text{m}^3$ and a coefficient of variation of 0.46. Sekolah Kebangsaan Taman Tun Dr. Ismail Jaya is site location for Shah Alam's urban station. The average PM₁₀ level in Shah Alam is 53.61 $\mu\text{g}/\text{m}^3$. As shown by the skewness value of 4.64, the PM₁₀ distribution is right-skewed. Shah Alam has a standard deviation of 26.89 $\mu\text{g}/\text{m}^3$ and a coefficient of variation of 0.50. The Shah Alam monitoring station recorded a range of PM₁₀ concentrations, from 14 to 587 $\mu\text{g}/\text{m}^3$. The Nilai monitoring station is located in an industrial zone with an average PM₁₀ concentration of 59.19 $\mu\text{g}/\text{m}^3$. The PM₁₀ distribution is more likely to have concentrations of PM₁₀ at the Nilai monitoring station, as measured by its right skewness (3.15), and its kurtosis (20.82). The average PM₁₀ concentration at the Larkin monitoring station is 43.17 $\mu\text{g}/\text{m}^3$, making it an industrial zone. The PM₁₀ distribution at the Larkin monitoring site is right-skewed (4.05), with a kurtosis of 37.01. Pasir Gudang has an average PM₁₀ concentration of 51.21 $\mu\text{g}/\text{m}^3$ and is classed as an industrial region. The PM₁₀ distribution is right skewed with a skewness of 4.65 and a kurtosis of 57.04, indicating that extreme events are likely to occur at the Pasir Gudang monitoring station.

3.2. Boosted regression tree model of PM_{10}

BRT models are presented to forecast the next day PM_{10} concentrations level at five air monitoring stations in Malaysia. The collected findings are reported in Table 3.

Table 3: Boosted regression tree model of PM_{10}

BRT Model	R^2
Perai station Trees (500), interaction.depth (5), shrinkage and n.minobsinnode (0.01, 10)	0.774
Shah Alam station Trees (500), interaction.depth (4), shrinkage and n.minobsinnode (0.01, 10)	0.813
Nilai station Trees (900), interaction.depth (5), shrinkage and n.minobsinnode (0.01, 10)	0.792
Larkin station Trees (1500), interaction.depth (2), shrinkage and n.minobsinnode (0.01, 10)	0.817
Pasir Gudang station Trees (500), interaction.depth (5), shrinkage and n.minobsinnode (0.01, 10)	0.810

As shown in Table 3, the BRT model findings imply that Larkin monitoring station has the most optimal trees with boosting (1500 trees). 500 trees are indicated for the Perai, Shah Alam, and Pasir Gudang monitoring stations, followed by 900 trees for the Nilai monitoring station. To increase the prediction performance, the model is configured with shrinkage values of 0.001, 0.01, and 0.1. The shrinkage represents each tree's and stump's contribution to the final model. The smaller shrinkage values between zero and one will necessitate more iterations and result in improved performance [27]. The shrinkage obtained for BRT models is 0.01 at all monitoring stations (Perai, Shah Alam, Nilai, Larkin, and Pasir Gudang). Larkin monitoring station and Shah Alam monitoring station have good BRT models with R^2 values of 0.817% and 0.813%, respectively. Suleiman et al. [23] findings at London Marylebone Road and London Bloomsbury were consistent with those of his earlier research. The values of the BRT model for estimating particulate matter fall between 0.87 and 0.96. Several air quality studies use the BRT model to estimate particulate matter [23], ozone concentrations [22, 27], and nitrogen dioxide concentrations [24]. All studies recommend the BRT model as a good predictive model for a variety of fields.

3.3. Multiple linear regression model of PM_{10}

Using the stepwise Ordinary Least Square (OLS) method, the MLR model was generated daily average PM_{10} values for five monitoring locations from January 1999 to December 2015. The projected PM_{10} concentrations for the following day and a summary of the model are shown in Table 4 accordingly.

Table 4: Multiple linear regression model of PM₁₀

MLR Model	R ²
Perai station The next day PM ₁₀ = 14.099 – 195.940NO ₂ - 9.198CO + 0.900PM _{10, D0}	0.638
Shah Alam station The next day PM ₁₀ = 90.251 + 1.040WS- 0.933Temp - 0.696 RH- 473.212SO ₂ + 111.774NO ₂ + 0.754PM _{10, D0}	0.678
Nilai station The next day PM ₁₀ = 42.422 - 0.668WS + 0.559Temp - 0.468RH + 0.704PM _{10, D0}	0.662
Larkin station The next day PM ₁₀ = 22.390 – 0.516WS – 0.102RH – 4.156CO + 0.773PM _{10, D0}	0.750
Pasir Gudang station The next day PM ₁₀ = 13.951 + 0.403Temp - 0.112RH + 282.644NO ₂ - 98.418O ₃ + 0.638PM _{10, D0}	0.753

As indicated in Table 4, the MLR models indicate that the PM₁₀ concentrations on the same day (PM_{10, D0}) most to the development of the forecasting models at the majority of stations. In agreement with Dominick et al. [19], the results indicate that PM₁₀ are the most influential factor in the MLR model prediction. At the Perai monitoring station, the PM₁₀ concentrations the following day are mostly influenced by the PM₁₀ concentrations from the previous day, in addition to nitrogen dioxide and carbon monoxide. The neighbourhood around the Perai observation post consists of both industrial and residential zones. It is possible that burning activities and industrial combustion contributed to the high levels of NO₂ in this area. The MLR model revealed the lowest R² value (R²=0.638) for the Perai monitoring station when compared to the other stations. To predict the PM₁₀ concentrations in Perai, Ul-Saufie et al. [30] acquire the greatest R² value of 0.942%. The MLR model relies heavily on the day's weather, as well as the concentrations of sulphur dioxide, nitrogen dioxide, and particulate matter PM₁₀ at the Shah Alam monitoring station. According to the MLR model, the current day's PM₁₀, temperature, wind speed, and relative humidity are the most important factors in predicting PM₁₀ concentrations at the Nilai monitoring station. Similar results were obtained by Ul-Saufie et al. [20] indicates that wind speed, temperature, and relative humidity all have significant positive and negative loading on the MLR model for the Nilai monitoring station. PM₁₀, temperature, nitrogen dioxide, and ground level ozone concentrations from the previous day are the key factors in predicting PM₁₀ levels for the following day at the Pasir Gudang monitoring station. The combustion of motor vehicles is the leading contributor to PM₁₀, nitrogen dioxide and carbon monoxide [20].

3.4. Performance Indicator

Evaluation of the performance of prediction models with observed datasets requires validation of statistical models. Five performance indicators (PI) were used to assess which of the MLR and BRT models provided the most optimal prediction. Based on the observed results, the best model is the BRT model, which accurately predicts PM₁₀ for all monitoring sites. According to the findings, the BRT model for Perai is relatively stable (R² = 0.774), Shah Alam (R² = 0.813), Nilai (R² = 0.792), Larkin (R² = 0.817) and Pasir Gudang (R² = 0.810). Results are in line with previous studies by Yahaya et al. [22], Suleiman et al. [23], Carslaw and Taylor, [24], and Asri et al. [27], as the BRT model is beneficial and advantageous when dealing with complicated information and selecting the appropriate parameters for the prediction model. The research also reveals the possibility of employing the BRT model as a

statistical model to predict PM₁₀ concentrations, which might be useful in the field of air quality research.

Table 5: Performance Indicator for MLR and BRT models

PI	Perai		Shah Alam		Nilai	
	BRT	MLR	BRT	MLR	BRT	MLR
MAE	6.966	8.642	10.362	12.944	11.055	13.393
RMSE	11.497	14.525	16.413	21.464	15.975	20.421
MAPE	15.907	19.593	18.251	21.442	16.959	20.042
IA	0.928	0.883	0.942	0.890	0.935	0.877
R^2	0.774	0.638	0.813	0.678	0.792	0.662
PI	Larkin		Pasir Gudang			
	BRT	MLR	BRT	MLR		
MAE	7.625	8.282	8.000	9.446		
RMSE	12.207	14.800	14.014	17.217		
MAPE	14.665	16.334	13.849	15.882		
IA	0.945	0.906	0.941	0.888		
R^2	0.817	0.750	0.810	0.753		

4. Conclusion

Malaysian PM₁₀ concentrations are predicted using the BRT and MLR Model. At the majority of monitoring stations, the MLR models suggest that the PM₁₀ concentrations on that same day (PM_{10,D0}) were the most influential factor in the MLR prediction models. $R^2 = 0.753$ indicates the best MLR models for the Pasir Gudang monitoring station. Comparing the BRT and MLR prediction models, BRT is the more appropriate model for short-term PM₁₀ concentration prediction than MLR. The BRT model provides more accurate PM₁₀ concentration predictions for each monitoring station, with R^2 values more than 0.774% and IA values larger than 0.928. Thus, the BRT model is a new approach of prediction employed in air quality investigations, particularly for PM₁₀ concentrations in Malaysia.

Acknowledgement

The Malaysian Ministry of Higher Education (MOHE) and the Universiti Malaysia Perlis (UniMAP) funded this research. We appreciate the Department of Environment Malaysia for providing the air quality data used in this study.

References

- [1] Lawrence S *et al* 2016 *Environmental Pollution* **210** 419-428
- [2] Samiksha S *et al* 2017 *Environmental Pollution* **222** 477-485
- [3] Jamalani M A *et al* 2018 *Journal of Fundamental and Applied Sciences* **10(1S)** 313-324
- [4] Zhou M *et al* 2014 *Environmental Pollution* **186** 1–6
- [5] Mohamed Noor N *et al* 2015 *Revista de Chimie-Bucharest* **66(9)** 1443-1447
- [6] Latif M *et al* 2014 *Science of the Total Environment* 482–483(2) 336–348
- [7] Sulong N A *et al* 2017 *Science of the Total Environment* (601–601) 556–570
- [8] Fotourehchi Z 2016 *Atmospheric Pollution Research* **7** 201-206
- [9] Hassan N A *et al* 2015 *Asia-Pacific Journal of Public Health* 1–11
- [10] Abdullah S *et al* 2019 *Atmosphere* **10(11)** 667
- [11] Ul-Saufie A Z *et al* 2015 *Journal of Environmental Science and Technology* **8(2)** 59-73
- [12] Shahraiyni T H and Sodoudi S 2016 *Atmosphere* **7(2)** 1-24
- [13] Kumar K and Pande B P 2022 *International Journal of Environmental Science and Technology* 1-16
- [14] Fong, S. Y *et al* 2018 *Journal of Sustainability Science and Management* **5** 3–17

- [15] Abdullah S *et al* 2017 *Journal of Sustainability Science and Management* **12(1)** 60–69
- [16] Pires J C M *et al* 2008 *American Journal of Environmental Sciences* **4(5)** 445–453
- [17] Perez P 2012 *Atmospheric Environment* **60** 271–276
- [18] Ul-Saufie A Z *et al* 2012 *Journal of Applied Sciences* **12(14)** 1488–1494
- [19] Dominick D *et al* 2012 *Atmospheric Environment* **60** 172–181
- [20] Ul-Saufie A Z *et al* 2013 *Atmospheric Environment* **77** 621–630
- [21] Brunelli U *et al* 2007 *Atmospheric Environment* **41** 2967–2995
- [22] Yahaya N Z *et al* 2017 *Environment Asia* **10(1)** 118–129.
- [23] Suleiman A *et al* 2016 *Environmental Modeling and Assessment* 731–750
- [24] Carslaw D C and Taylor P J 2009 *Atmospheric Environment* **43(22–23)** 3563–3570
- [25] Elith J *et al* 2008 *The Journal of Animal Ecology* **77(4)** 802–13
- [26] Elith J and Leathwick J 2013 Boosted Regression Trees for Ecological Modeling 1-22
- [27] Asri M A M *et al.* 2015 *Statistics* **82** 32419–32424.
- [28] Abdullah S *et al* 2016 *Environment Asia* **9(2)** 101–110
- [29] Mendenhall W and Sincich T 1995 *Statistics for Engineering and The Sciences (Fourth edition)*. Prentice Hall Inc.
- [30] Ul-Saufie A Z *et al* 2011 *International Journal of Applied Science and Technology* **1(4)** 42-49