PAPER • **OPEN ACCESS**

The Analysis of Pollutant Concentrations Before and During Movement Control at Quarry Areas Using an Artificial Intelligent Technique

To cite this article: N Z Yahaya *et al* 2022 *IOP Conf. Ser.: Earth Environ. Sci.* **1013** 012011

The Analysis of Pollutant Concentrations Before and During Movement Control at Quarry Areas Using an Artificial Intelligent Technique

N Z Yahaya[1*], J Jalaludin[2], I A I Rais[3], U S Mohamad Hussin[4], J Yahaya[5], E S A Rahman[6]

[1]Faculty of Ocean Engineering Technology and Informatics, University Malaysia Terengganu, Kuala Terengganu, 21030, Malaysia
[2]Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Serdang, Selangor, Malaysia
[3]Pusat Penyelidikan Mineral Malaysia, Jabatan Mineral dan Geosains Malaysia, Ipoh, Perak, Malaysia
[4]Fakulti Pengurusan dan Teknologi Maklumat, Universiti Sultan Azlan Shah, Kuala Kangsar, Perak, Malaysia
[5]Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, Bangi, Selangor, Malaysia
[6]Department of Environment (Air Division), Ministry of Water and Environment, Putrajaya, Malaysia
*Corresponding author: nzaitun@umt.edu.my

Abstract. This paper investigated the influence and interactions of air pollution concentrations by using the stochastic boosted regression trees between variables for each station and the impact of the COVID-19 Movement Control Order at Ipoh City air quality station. The one-hour data were gathered from the Department of Environment from January until June 2019 and 2020. Two thousand two hundred thirty-one data of particles, gases (Nitrogen oxides, Sulphur Dioxide, Ozone, Carbon Monoxide) concentrations and meteorological data (wind speed, wind directions, temperature, and relative humidity) were captured. The BRT model development process with an algorithm using a comprehensive package, R Software and its packages to understand the variability and trends. It was found that the relationship between the number of samples and number of trees (*nt)* of 4372 for *oob* were found the best iterations obtained. The performance of the boosting model was assessed and found that the FAC2 was 0.91, the R^2 values were above 0.56 (R = 0.74), and the Index of Agreements (IOA) was 0.67, which fall ranges are within an acceptable for model performance. The Relative Variable Importance (RVI) that influenced $PM_{2.5}$ for non-MCO data was CO (18.9%), SO_2 (14.6 %), O_3 (12.9 %), and wd (10.66 %) while CO (22.6%), RH (13.4%), 14.7% and O_3 (12.1%) were RVI factors influenced to PM2.5 concentrations during MCO periods. Estimating the strength of interaction effects (SIE) between variables was 0.24 for CO-wind directions, followed by 0.19 for ozone-wind speeds and 0.15 for NO_2-CO. Results showed that the model developed was within the acceptable range and could be used to understand particles and identify important parameters that influence particle concentrations.

Keywords: Pollutants; Boosted Regression Trees; R-Packages; Relative Variable Importance, Strength of Interactions

1. Introduction

It was reported that a continuous occurrence of an unknown acute respiratory tract infection was first reported in Wuhan City, Hubei Province, China, on 12 December 2019, originating from the Hunan South China Seafood Market [1]. This coronavirus was initially named the 2019-novel coronavirus (2019-nCoV) by the World Health Organization (WHO). But then the WHO renamed this disease the coronavirus disease 2019 (COVID-19). Later after several efforts and studies taken by the WHO and Chinese scientists, this disease spread worldwide. On 11 March 2020, WHO announced the new coronavirus or COVID-19 illness as a "global pandemic" [2]. Concerning that, governments worldwide have taken several steps and actions to control the spreading of COVID-19. The measures include closing schools, industrial activities, restrictions to public transport, houses of worship closure and many other actions.

The COVID-19 pandemic is thought to originate from China by human contact with non-domestic animals. The origin of the COVID-19 virus is believed to have spread from the wild animal market in Wuhan, China [3, 4]. Infectious rates of diseases occur faster than in the past, and the rate of spread increases with increasing human contact. Between January 2020 and the middle of July 2020, COVID-19 infections were detected in more than 13 million worldwide, and 600 thousand people died. According to the World Health Organization [3], there have been 5,313,098 confirmed cases of COVID-19. In Malaysia, COVID-19 was first reported and imported from Wuhan, China, on 25 January 2020 and spread to eight positive patients within six days [5] since the first case. The situation became unfavourable on 11 March 2020 as the spreading of this virus had worsened across the country until then to control the spreading and mortality due to the virus, the Ministry of Health, Malaysia issued a Movement Control Order (MCO), which was then implemented across the country on 18th March 2020. MCO enforcement came under the Prevention and Control of Infectious Diseases Act 1988 and the Police Act 1967 and would help control the spread of the virus. Table 1 shows pollutant standards for the European Union and World Health Organization and Department of Environment (Air Division), Malaysia.

Table 1. Air Quality Standards for the European Union, World Health Organisation Standards and Department of Environment (Air Division), Malaysia

Pollutants/mean	*European Union Air Quality Standards Limit Value (µg/m^3)	**WHO Guidelines (µg/m^3)	***MAAQS (2018) -IT2 (µg/m^3)	US EPA 2003 (µg/m^3)
$PM_{2.5}$ (24-hr mean)	NA	15	50 µg/m^3	35 µg/m^3
$PM_{2.5}$(Annual)	25 µg/m^3	5 µg/m^3	25 µg/m^3	15.0 µg/m^3
PM_{10} (24-hr)	50 µg/m^3	45 µg/m^3	120 µg/m^3	150 µg/m^3
PM_{10}(Annual)	40 µg/m^3	15 µg/m^3	45 µg/m^3	NA
O_3 8-hr mean	120 µg/m^3	[a]100 µg/m^3	120 µg/m^3	0.070 ppm
O_3 1-hr mean	NA	[b]60 µg/m^3	200 µg/m^3	0.12 ppm
NO_2 (Hourly)	200 µg/m^3	25 µg/m^3	300 µg/m^3	100 ppb
NO_2 (Annual)	40 µg/m^3	10 µg/m^3	75 µg/m^3	53 ppb
SO_2 (Hourly)	350 µg/m^3	NA	300 µg/m^3	75 ppb
SO_2 (24-hour)	125 µg/m^3	40µg/m^3	90 µg/m^3	0.14 ppm
CO (1-hr mean)	NA	26 ppm	35 mg/m^3	35 ppm (40 mg/m^3)
CO (8-hr mean)	10 mg/m^3	9 ppm	10 mg/m^3	9 ppm (10 mg/m^3)

*Sources: EU Air Quality Directive (2008/50/EC) https://ec.europa.eu/environment/air/quality/standards.htm),
**WHO,2021, Air Quality Guidelines: Global update
[a] 100 µg/m3, 8-hour daily maximum - 99th percentile, (i.e. 3-4 exceedance days per year)
[b] 60 µg/m3 8-hour mean, peak season - Peak season is defined as an average, so daily maximum 8-hour mean O_3 concentration in the six consecutive months with the highest six-month running average O_3 concentration
*** Malaysia Air Quality Standards [6]

The Department of Environment (DOE) Malaysia reported that air quality levels were low and worse than before the MCO. According to the DOE, the primary pollutants such as Nitrogen dioxide (NO_2), Sulphur dioxide (SO_2), Carbon monoxide (CO), and particulate matter with 2.5-micrometre size at several locations in an urban environment such as Batu Muda, Cheras and Putrajaya were decreased dramatically. It is significantly shown that the reduction of NO_2 is the highest down to 62% lower than before MCO and followed by CO gases. Reducing NO_2 is the highest down to 62% lower than before MCO and followed by CO gases with a maximum of 22% reduction. The main cities in Malaysia also show the same trends whereby NO_2 reduction was highest recorded, followed by SO_2, CO, and particulate matter. It is good to see that air pollution has also decreased. The percentage of stations that recorded "good" air quality readings increased two-fold from 28% to 57% after MCO.

Abdullah *et al.* studied the air quality status during the 2020 Malaysia MCO due to the 2019 novel coronavirus (2019-nCoV) pandemic by comparing $PM_{2.5}$ levels [5]. The study revealed that the changes of fine particulate matter ($PM_{2.5}$) at 68 air quality monitoring stations were found that the $PM_{2.5}$ concentrations showed a high reduction of up to 58.4% during the MCO. Several red zone areas (N41 confirmed COVID-19 cases) had also reduced up to 28.3% in the $PM_{2.5}$ concentrations variation. A study by Latif *et al.* in Klang Valley air quality data in Malaysia indicated that there were significant reductions ($p<0.05$) of PM_{10}, $PM_{2.5}$, NO_2 and CO during the MCO compared with the same periods in 2019 and 2018 [7]. NO_2 recorded the highest percentage of reduction during the MCO with a percentage reduction between the drops of its precursors.

Various statistical tools have been used to analyse $PM_{2.5}$ data used to predict and forecast Malaysia and elsewhere. A study by Marsha and Larkin used a multiple linear regression scheme that relies on the previous day's $PM_{2.5}$ value and fire and smoke-related variables from satellite observations [8]. It was found that the statistical model obtained was able to explain 78% of the variance in daily ground-level $PM_{2.5}$ data during a smoke event in Western United States monitoring stations. A study by Ul-Saufie *et al.* (2013) reported that the regression method is one of the acceptable modelling techniques that has been used to predict PM_{10} concentration [9]. Compared to the traditional least-squares way, using this method, the influence of outliers can be minimised. The most recent studies have attempted to develop powerful computational intelligence models using machine learning algorithms (neural network) conducted by Abdullah *et al*. (2019) to use PM_{10} concentrations in Malaysia air quality data to predict the desired value [5].

Recent studies have demonstrated that a Boosted Regression Tree (BRT) model with stochasticity (SGB) can be used to obtain the best model that can deal with a high-level complexity and large datasets and yield substantial outcomes [10–13]. The boosting technique is applied to 'boost' the model accuracy of any given learning algorithm. Boosting is a general method that can be used, and it was first developed by Friedman in 2001 [14] and then added the stochastic element in the algorithm by Friedman in 2002 [15]. The difference in the stochastic BRT algorithm is that at each iteration, a subsample of the training data is randomly extracted from the complete training data set without being replaced. Furthermore, extensive work was performed by Carslaw and Taylor [10] and Yahaya [11] in the City of Leeds, UK. Their studies were among the first to have examined and used BRTs to analyse air pollution, traffic, and meteorological data intensively. Munir et al. (2021) compared different approaches for assessing the impact of COVID-19 lockdown on urban air quality in the City of Reading, United Kingdom [16]. In their study, they compared different approaches, namely: (1) Sequential approach – comparing pre-lockdown and lockdown periods 2020; (2) Parallel approach – comparing 2019 and 2020 for the equivalent time of the lockdown period; and (3) Machine learning modelling approach as a method to analyse COVID-19 impact to the air quality.

The BRT stochastic procedure [15] as applied to the gradient boosting algorithm is described as follows: Let $\{y_i, x_i\}_i^N$ of known (y,x) values be the entire training data sample and $\{\pi(i)\}_i^N$ be a random permutation of the integers {1,.......N}. Then a random subsample of size $\tilde{N}<N$ is given by $\{y_{\pi(i)}, x_{\pi(i)}\}_i^{\tilde{N}}$. The SGB algorithm is then:

1. $\hat{f}_0(x) = arg\ arg\ min_{\gamma} \sum_{1=1}^{N} \quad \psi(y_i, \gamma)$ (1)
2. $Form = 1\ toMdo$:
3. $\{\pi(i)\}_1^N$ = rand_perm $\{i\}_1^N$ (2)
4. $\tilde{y}_{im}$= - $[\frac{\partial\varphi(y_i,F(x_i))}{\partial F(x_i)}]_{F(x)=F_{m-1}(x)}$, I = 1, $\tilde{N}$ (3)
5. $\{R_{lm}\}_1^L)$ = L – terminal node tree ($\{\tilde{y}_{\pi(i)m}, x_{\pi(i)m}\}_1^{\tilde{N}}$ (4)
6. $\gamma_{lm} = arg\ min_{\gamma} \sum_{x_{\pi(i)\in R_{lm}}} \quad \psi\left(y_{\pi(i)}, F_{m-1}\left(x_{\pi(i)}\right) + \gamma\right)$ (5)
7. F_m (x) = F_{m-1}(x) + $\upsilon.\gamma_{lm}$ 1(x∈ R_{lm}) (6)
8. $endFor$
9. End Algorithm

According to Friedman (2002), the smaller the fraction $f = \tilde{N}/N$, the more the random samples used in successive iterations will differ [15]. Making the value of f smaller reduces the amount of data available to train the base learner at each iteration. This condition causes the variance associated with the individual base learner estimates to increase. Among the outputs of the BRT is the ability to examine RVI and estimate the SIE obtainable between variables. It is expected that the outcomes from this analysis are; First, the RVI in percentage (%) that will rank which variables most influence the concentrations of pollutants at different categories. This information is vital for the manager and planner to understand which pollutants need attention, especially during strategy planning, in mitigating the pollutant levels taking into place; Second, to estimate the SIE on a scale of 0 – 1, which will explore the interaction of variables.

2. Methods

Pakar Scieno TW Sdn Bhd recorded the data of the major air pollutants such as PM_{10}, $PM_{2.5}$, SO_2, NO_2, CO, O_3 on an hourly basis at the Continuous Air Quality Monitoring Station (CAQMS) by the orders of the Malaysian DOE. The hourly dataset of air pollution from the 1 March 2019 (before MCO) to – 1 July 2020 (during MCO) was obtained from DOE for this study. The data was gathered from the DOE from March until June in 2019 and the same period in 2020 to determine the variability of particulate matter ($PM_{2.5}$ and PM_{10}) and gases such as Sulphur dioxide (SO_2), Nitrogen dioxide (NO_2) and other gases. The data were compared between the selected areas based on COVID-19 most affected areas along with meteorological variables analysis. From the zoning map, the sampling point falls into Institution and Community Facilities. Two DOE monitoring stations were selected which are SMK Jalan Tasek (4° 37' 50.0448", 101° 7' 4.224") and SK Jalan Pegoh (4° 33' 16.1058", 101° 4' 54.192"). The stations are marked as blue in Figure 1.

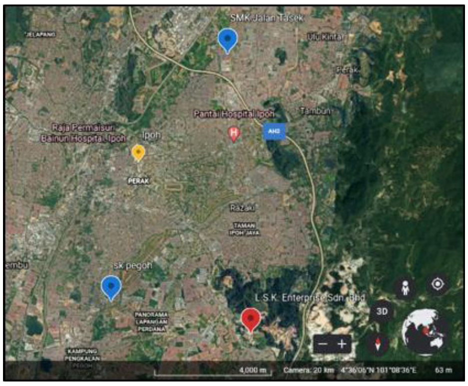

Figure 1. Location of CAQM station at Pegoh and Tasek station in City of Ipoh, Perak, Malaysia (www.google.com/maps/)

3. Results and Discussion

The current famous software used for air quality data analysis is open-source R-package. It was developed by the Development Core Group R (2008). It was used to run and analyse the data obtained from DOE using an extended R Software version 4.1.0 [17] and Openair packages [18, 19]. The list of variables involved in the algorithm setting is shown in Table 2. R is free software, robust and interpreted language with powerful interactive analytical features, making it ideal for quickly developing statistical and data analysis applications [19]. The generalised boosting machine (gbm) package in R was utilised to study the data and produce the BRT algorithms. It was found that the relationship between the number of samples and number of trees (*nt)* of 4372 for *oob* were found the best iterations obtained. The performance of the boosting model was assessed and found that the FAC2 was 0.91, the R^2 values were above 0.56 (R = 0.74), and the Index of Agreements (IOA) was 0.67, which fall ranges are within an acceptable for model performance.

Table 2. Variables names and descriptions used to model boosting algorithm setting

Response Name		Variable names and descriptions	Units
1.	$PM_{2.5}$	Particulate Matter	(µg/m^3)
Explanatory Variables			
1.	O_3	Ozone	(ppm)
2.	NO	Nitric Oxide	(ppm)
3.	NO_2	Nitrogen Dioxide	(ppm)
4.	NO_x	Nitrogen Oxides	(ppm)
5.	SO_2	Sulphur Dioxide	(ppm)
6.	CO	Carbon Monoxide	(ppm)
7.	temp	Ambient Temperature	(^{o}C)
8.	humid	Relative Humidity	Percentage (%)
9.	ws	Wind Direction and Wind Speed at 10m height	(km/hr)
10.	wd	Wind Direction	(^{o})

The objective here was to find the best combination of parameters (*lr, tc*, and *nt*) that minimised the predictive error (minimal error for predictions to independent samples). The stochastic gradient machine technique was also applied in the BRT model; this technique introduces randomness into the boosted model, typically increases the accuracy and speed, and reduces overfitting [15]. At this stage, the sequential model-fitting process builds on fitted trees. A series of BRT analyses using the gradient boosted model (*GBM* in different settings was performed to examine the *GBM* before determining *tc*, *nt*, and *lr*. The GBM uses three methods to estimate the optimal number of iterations after the GBM model was fitted to identify the optimal number of trees. An independent test set (*test*) method, out-of-bag estimation (*OOB*), and v-fold cross-validation optimisation (CV). Similar to Friedman's MART software [20, 21], it determines the optimal number of iterations by using the independent test set method involving the use of a single holdout base dataset. BRT models with different *numbers of trees* (from 1000 to 10000) were simulated for datasets with *nt* = 10000, *lr* = 0.001, independent depth = 5, and CV fold = 10. A similar shrinkage value (*lr*) of 0.001 was suggested by Ridgeway [20, 21] and used by Carslaw and Taylor [10] and Yahaya [11]for analysing an air pollution dataset.

3.1. BRT Outputs

3.1.1. Relative Variable Importance (RVI)

The Relative Variable Importance (RVI) of the predictor variables was determined using the formulae developed by [14] and implemented in the *gbm* library to estimate the relative importance of predictor variables. At each split in each tree, *gbm* computes the improvement in the split-criterion (MSE for regression). In these methods, the measure is calculated based on the number of times a variable is selected. The effort is used for splitting and is weighted by the improvement in the overall model. The variables with the most significant average decrease in Mean Square Error (MSE) are considered most important. The relative importance analysis focuses on the essential variables obtained from the *gbm* iteration, particles, gases and meteorological variables to the formation of $PM_{2.5}$ in this study. Table 3 shows a summary of relative variable importance (RVI) comparing non-MCO (2019) and MCO situations (2020) data. It indicates that CO (18.9 %) gases are the most dominant gases followed by SO_2 (14.6 %) of RVI values before during non-MCO period influenced $PM_{2.5}$ concentrations followed by meteorological parameters (relative humidity and wind directions). Similar during MCO, CO remained the highest variables to affect $PM_{2.5}$ with 22.6 % RVI, followed by O_3 with 12.1 % and meteorological parameters (temperature and relative humidity). These variables have also been commonly used by other air pollution studies and are known to influence air pollution data, especially during MCO and Non-MCO conditions. The highest differences of RVI are shown by SO_2 levels, which reduced from 14.6% during non-MCO to only 5.7 % during MCO periods. The RVI for O_3 and NO are not so different to give an impact during the pandemic.

Table 3. Relative Variable Importance (RVI) in percentage (%) comparison between 2019 and 2020 (MCO) data for Pegoh station in Ipoh, Perak City

Pollutant	CO	SO_2	NO_x	NO	NO_2	O_3	Temp	RH	WD	WS
2019	18.9	14.6	6.61	6.00	5.11	12.9	9.75	10.66	10.54	4.82
2020	22.6	5.7	3.4	6.6	8.2	12.1	14.7	13.4	7.74	5.6
Differ	+3.7	-8.9	-3.21	+0.6	-3.09	-0.8	+4.95	+2.74	-2.8	+0.78

3.1.2. Strength of Interaction Effects (SIE) and Interaction Indexes

The Strength of Interactions Effects (SIE) between parameters is another vital output in BRT that can explain the level of interactions between variables. De'ath (2007) described how two-way interactions could be used to determine the relative importance of considering the interaction between two variables compared with the purely additive combination [22]. For instance, the two-way interaction of two variables, x_1 and x_2, made predictions across regular interval combinations while keeping all other variables at their mean level. The variable interactions can be determined in BRT by varying the number of splits (size) of the individual trees, determining the degree to which predictors interact in determining the response [11]. The variable interactions can be determined in BRT by varying the number of splits (size) of the individual trees, determining the degree to which predictors interact in determining the response. These functions implement bivariate interpolation onto a grid for irregularly spaced input data. Bilinear or bicubic spline interpolation is applied using different versions of algorithms from *Akima* R-package [23, 24].

Table 4 shows the strength of the interaction index between variables to the level of $PM_{2.5}$ at the station. It can be seen that carbon monoxide gas is the most dominant pollutants and contributes to the highest Index with 0.23 (SO_2 – CO) followed by 0.18 (for CO–WS) compared to lower interactions were obtained during MCO (2020 data). Higher indexes may also be a good indication of the effects of motor vehicle emissions induced to the atmosphere before MCO. In contrast, SO_2 gases, which usually are emitted from industrial activities or diesel engine operation, is found the topmost dominant influencing the index levels during MCO with 0.17 (SO_2 – O_3), 0.16 (NO-SO_2), 0.15 (NO_x – SO_2) and 0.14 (SO_2 – O_3). The primary *sources* of SO_2 *emissions* are fossil fuel combustion at power plants, refineries, and other industrial facilities (cement, quarrying activities or other industries).

Table 4. Summary of The Strength of Interactions Effects (SIE) between parameters in this study for Pegoh station

	NO - NOx	NOx- SO_2	NO- SO_2	SO_2 -O_3	NO_2 -CO	SO_2 -CO	NO_2 -O_3	SO_2 - Temp	CO- WS
2019	0.10	0.04	0.07	0.05	0.14	0.23	0.08	0.13	0.18
2020	0.01	0.15	0.16	0.17	0.06	0.07	0.13	0.06	0.07
Differ	0.09	-0.11	-0.09	-0.12	0.08	0.16	-0.05	0.07	0.11

Further analyses were done to explore the SIE in a graphical form. The level of concentrations is plotted, and the lined contour displays the same concentration effects within the same concentrations. Figure 2 a-d and 3a-d illustrate the SIE and Index of Interactions between parameters for Pegoh station in 2020. The SIE and Index of Interactions between parameters for Pegoh station in 2019. As described, the highest interactions between CO and SO_2 gases with 0.23 out of 1 indicated that these pollutants usually are emitted from motor vehicles emissions. These results also strongly suggest the potential impact on people's health due to the high level of interactions, especially when finding a strategy to reduce air pollution in an urban environment. It was also indicated that lower interaction had been shown between gases during MCO conditions, which shows there are fewer interactions of gases during MCO time due to lower levels of pollutants.

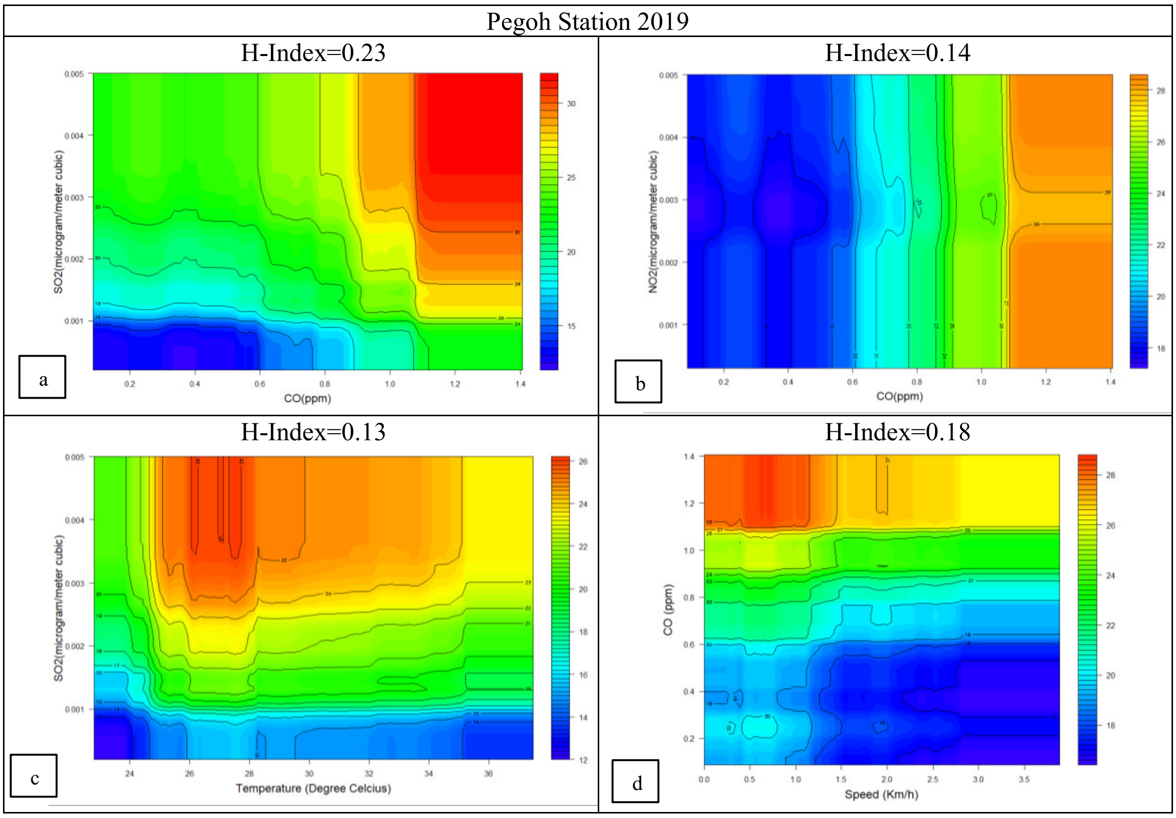

Figure 2a-d. The Strength of Interactions Effects (SIE) and Index of Interactions between parameters for Pegoh station in the year 2019

Pegoh Station 2020

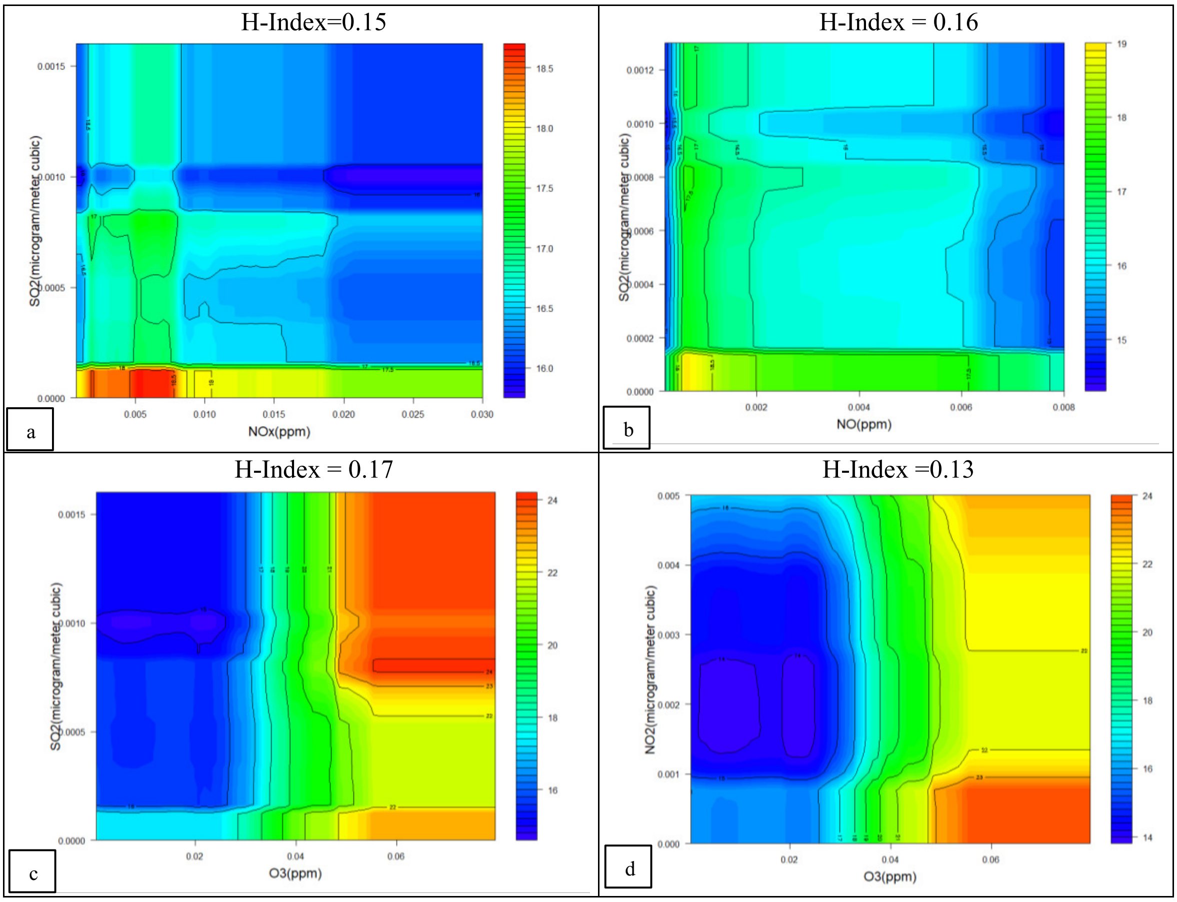

Figure 3 a-d. The Strength of Interactions Effects (SIE) and Index of Interactions between parameters for Pegoh station in the year 2020

4. Conclusions

During the COVID-19 pandemic quarantine, the restricted movement of motor vehicles, transportations, and closure of industries greatly influenced the concentration of air pollutants. Some measures taken during the COVID-19 pandemic period may be turned into permanent habits over time. Comparing the 2019 and 2020 MCO data, it is clear that CO (18.9 %) gas is the most dominant gas followed by SO_2 (14.6 %) of RVI values before during non-MCO period influenced $PM_{2.5}$ concentrations followed by meteorological parameters (relative humidity and wind directions). Similar during MCO, CO remained the highest variable to control $PM_{2.5}$ with 22.6 % RVI, followed by O_3 with 12.1 % and meteorological parameters (temperature and relative humidity). These variables have also been commonly used in other air pollution studies and influence air pollution data, especially during MCO and non-MCO conditions. The highest difference of RVi is shown by SO_2 levels, which reduced from 14.6% during non-MCO to only 5.7 % during MCO periods.

Studying the characteristics of PM_{10}, CO, and O_3 is vital for improving the quality of life, physical health, and inhabitants' living environment. However, the factors that influence these phenomena are still in study in most cities and expanded. The COVID-19 outbreak may be a turning point in combating global climate change as the behaviours of travelling and people's movements are restricted. The improvement in air quality can be considered as the most crucial gain after the pandemic. Air pollution and climate change have long term effects through complex interactions in the atmosphere and ground-level surface of pollutants, especially in urban environments.

It is valuable to understand the relationship between parameters of the RVI, SIE, and IOI to the concentrations of particulate matter. Especially when observing the pattern of before and after the MCO was implemented in certain states or countries. The findings add our knowledge of the mechanism of air pollution and provide a scientific reference for implementing targeted control measures. This study provides an insight into how the gaseous interaction in the atmosphere happens, which might be harmful to human health and the environment in general. For sustainable environmental and air quality, gains from measures should be continued.

Acknowledgments
The authors wish to express their gratitude towards the Department of Environment (Air Division), Ministry of Water and Environment, Malaysia for producing the data, Atmospheric, Air Quality and Climate Change Research Group, University Malaysia Terengganu, University Putra Malaysia, University Kebangsaan Malaysia, Mineral Research Centre, Minerals and Geoscience Department Malaysia, Universiti Sultan Azlan Shah, Malaysia and The Clean Air Forum Society of Malaysia for their support.

References
[1] Guo YR, Cao QD, Hong ZS, et al. The origin, transmission and clinical therapies on coronavirus disease 2019 (COVID-19) outbreak- An update on the status. *Mil Med Res* 2020; 7: 1–10.
[2] WHO. *Coronavirus disease 2019 (COVID-19) Situation Report-94*. 2020.
[3] WHO. Ambient (Outdoor) Air Quality and Health. *World Health Organization*, https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health (2018).
[4] Shah AUM, Safri SNA, Thevadas R, et al. COVID-19 outbreak in Malaysia: Actions taken by the Malaysian government. *Int J Infect Dis* 2020; 97: 108–116.
[5] Abdullah S, Mansor AA, Napi NNLM, et al. Air Quality Status during 2020 Malaysia Movement Control Order (MCO) Due To 2019 Novel Coronavirus (2019-nCoV) Pandemic. *Sci Total Environ* 2020; 729: 139022.
[6] DOE Malaysia. New Malaysia Ambient Air Quality Standard, https://www.doe.gov.my/portalv1/en/category/info-umum/info-kualiti-udara (2020).
[7] Latif MT, Dominick D, Hawari NSSL, et al. The concentration of major air pollutants during the movement control order due to the COVID-19 pandemic in the Klang Valley, Malaysia. *Sustain Cities Soc* 2021; 66: 102660.
[8] Marsha A, Larkin NK. A statistical model for predicting PM2.5 for the western United States. *J Air Waste Manag Assoc* 2019; 69: 1215–1229.
[9] Ul-Saufie AZ, Yahaya AS, Ramli NA, et al. Future daily PM10 concentrations prediction by combining regression models and feedforward backpropagation models with principle component analysis (PCA). *Atmos Environ* 2013; 77: 621–630.
[10] Carslaw DC, Taylor PJ. Analysis of air pollution data at a mixed source location using boosted regression trees. *Atmos Environ* 2009; 43: 3563–3570.
[11] Yahaya NZ. *Temporal and Spatial Variations of Ultra-fine Particles in the Urban Environment*. University of Leeds, 2013.
[12] Yahaya NZ, Ghazali NA, Ahmad S, et al. Analysis of daytime and nighttime ground level ozone concentrations using boosted regression tree technique. *Environ Asia* 2017; 10: 118–129.
[13] Yahaya NZ, Ibrahim ZF, Yahaya J. The used of the Boosted Regression Tree Optimization Technique to Analyse an Air Pollution data. *Int J Recent Technol Eng* 2019; 8: 1565–1575.
[14] Friedman JH. Greedy Function Approximation: A Gradient Boosting Machine. *Ann Stat* 2001; 29: 1189–1232.
[15] Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal* 2002; 38: 367–378.
[16] Munir S, Luo Z, Dixon T. Comparing different approaches for assessing the impact of COVID-19 lockdown on urban air quality in Reading, UK. *Atmos Res*; 261. Epub ahead of print 15 October

2021. DOI: 10.1016/J.ATMOSRES.2021.105730.
[17] R Core Team. R: a language and environment for statistical computing. *R Foundation for Statistical Computing*, https://www.gbif.org/tool/81287/r-a-language-and-environment-for-statistical-computing (2018).
[18] David Carslaw. *The openair manual open-source tools for analysing air pollution data*, https://github.com/davidcarslaw/openair. (2019).
[19] Carslaw DC, Ropkins K. openair — An R package for air quality data analysis. *Environ Model Softw* 2012; 27–28: 52–61.
[20] Ridgeway G. Generalized Boosted Models: A Guide to The GBM Package, https://rdrr.io/cran/gbm/man/gbm.html (2007).
[21] Ridgeway G. Generalized Boosted Models: A Guide to the GBM Package, https://www.researchgate.net/publication/215990688_Generalized_Boosted_Models_A_Guide_to_the_GBM_Package (2005).
[22] De'ath G. Boosted Trees for Ecological Modeling and Prediction. *Ecology* 2007; 88: 243–251.
[23] Akima H. Algorithm 761: Scattered-data surface fitting that has the accuracy of a cubic polynomial. *ACM Trans Math Softw* 1996; 22: 362–371.
[24] Akima H. A Method of Bivariate Interpolation and Smooth Surface Fitting for Irregularly Distributed Data Points. *ACM Trans Math Softw* 1978; 4: 148–159.