Spatio-temporal modelling of the influence of climatic variables and seasonal variation on PM$_{10}$ in Malaysia using multivariate regression (MVR) and GIS

Abdulwaheed Tella$^a$, Abdul-Lateef Balogun$^a$ and Ibrahima Fayeb$^b$

$^a$Geospatial Analysis and Modelling (GAM) Research Laboratory, Department of Civil and Environmental Engineering, Universiti Teknologi PETRONAS (UTP), Seri Iskandar, Perak, Malaysia; $^b$Center for Intelligent Signal and Imaging Research & Fundamental and Applied Sciences Department, Universiti Teknologi PETRONAS, Seri Iskandar, Perak, Malaysia

**ABSTRACT**

In an era of rapidly changing climate, investigating the impacts of climate parameters on major air pollutants such as Particulate matter (PM$_{10}$) is imperative to mitigate its adverse effect. This study utilizes Geographic Information System (GIS), a multivariate regression model (MVR) and Pearson correlation analysis to examine the inter-relationship between PM$_{10}$ and major climate parameters such as temperature, wind speed, and humidity. Although the application of MVR for predicting PM$_{10}$ has been examined in previous studies, however, the spatial modelling and prediction of this air pollutant is limited. Accurate spatial assessment of pollutants' hazard susceptibility in relation to climate change can accelerate mitigation initiatives. Thus, to understand the behavior, seasonal pattern, and trend of PM$_{10}$ concentration which is vital for good air quality, GIS is essential for enhanced visualization and interpretation of the predicted occurrence of the pollutant. The acquired data were randomly divided into 80% and 20% for training and validation of the MVR model, respectively while GIS was used to model the spatial distribution of the predicted ambient PM$_{10}$ concentration, highlighting the hotspots of future PM$_{10}$ hazard. A positive correlation index was obtained between PM$_{10}$ with temperature and wind speed. However, humidity showed a negative correlation. The regression model showed high predictive performance of $R^2 = 0.298$, RMSE = 12.737, and MAE of 10.343, with the highest PM$_{10}$ concentration correlated with the warming event in the southwest monsoon. Temperature, wind speed, and humidity were identified as the most critical variables influencing PM$_{10}$ concentration in the study area, in descending order of importance. This study's outcome provides valuable spatio-temporal information on future climate change impact on PM$_{10}$ in the study area with the potential to support effective air quality management.

**ARTICLE HISTORY**

Received 16 October 2020
Accepted 14 January 2021

**KEYWORDS**

Climate; GIS; multivariate regression (MVR); pollution; PM$_{10}$; spatio-temporal
1. Introduction

Particulate matters (PM) are hazardous pollutants that have tremendous negative impacts on human health and environmental sustainability (Choubin et al. 2020). Exposure of people to atmospheric pollution impacts human health, causing heart diseases, respiratory diseases, and increased risk of stroke (World Health Organization 2016a), which has led to an estimated 3 million deaths annually (World Health Organization 2016b). Rapid urbanization characterized by population growth (United Nations 2015) is influencing the environment and affecting the air quality (Choubin et al. 2020; Liu et al. 2014). Recent projections indicate that the world’s population will exceed 9.5 billion by 2050 (United Nations 2015), and a high percentage of cities with over 100,000 citizens, majorly in the underdeveloped and developing countries, are unable to meet standard air quality guidelines (Osseiran and Chriscaden 2016).

Particulate matter with diameter 10 μm (PM10) are air pollutants that can easily penetrate the lung tract (de Rooij et al. 2017) and it is considered the most dangerous air pollutant in urbanized regions in the world (Yadav et al. 2014; Stoimenova et al. 2017; Ganguly et al. 2019; Cujia et al. 2019; Han et al. 2020). Exposure to PM10 for a short or prolonged period has severe impacts on global health, causing lung and heart diseases, and deaths if its concentration increases above its strata in the atmosphere (Czernecki et al. 2017; Feng et al. 2019; Yao et al. 2015; Zheng et al. 2015). For instance, a rise in the concentration of PM10 in Malaysia from 50 μg/m³ to 150 μg/m³ has been linked to an increase in respiratory diseases (12%), exacerbating asthmatic condition (19%) and rhinitis (26%) (Wong et al. 2017). This has been influenced by rapid urbanization leading to an increase in the population and industrialization (Qureshi et al. 2015), thereby depleting the air quality (Abdullah et al. 2017). PM10 is the dominant atmospheric pollutant in Peninsular Malaysia due to its higher level of concentration in the Air pollution Index (API) (Abdullah et al. 2017). In comparison to other air pollutants such as sulfur dioxide (SO2), ozone (O3), nitrogen (NOx), and carbon monoxide (CO), it has been mostly considered as the principal air pollutant for API calculation which indicates the Malaysian API level (Shaadan et al. 2015; Althuwaynee et al. 2020).

Despite the recent focus on PM2.5 (Wang and Wang 2019; Zaman et al. 2019; Li et al. 2019), further studies to deepen the understanding of the behavior, trend, seasonal patterns, and sources of PM10 remain crucial due to its adverse impacts on the respiratory and cardiovascular system (Al-Hemoud et al. 2017; Sun et al. 2018). Its potentiality in increasing the death rate for a 10 μg/m³ rise in concentration (Althuwaynee et al. 2020; de Rooij et al. 2017) is equally a source of concern.

1.1. Causes and enablers of air pollution (PM10)

PM10 occurs naturally or anthropogenically. Natural sources of PM10 include dust, sea salt, and carbon emitted from open burning (Czernecki et al. 2017). Emission from open burning in Malaysia and neighbouring countries, particularly Indonesia (Wong et al. 2017), is the main source of PM10 in the country (Khan et al. 2015; Latif et al. 2011). Human-induced sources of PM10 include vehicular emission, dust from soils, emission from industries, and power plants (Azarmi et al. 2016; Taheri
Shahraiyni and Sodoudi 2016). These anthropogenic sources have a considerable influence on the release of PM$_{10}$ to the atmosphere (Khan et al. 2015; Abdullah et al. 2017).

Climatic factors such as temperature, wind speed, and humidity have a significant impact on the formation, transportation, and deposition of atmospheric pollutants (Khan et al. 2015; Zhang and Ding 2017). Temperature and humidity affect the physical and chemical components of PM$_{10}$ (Khan et al. 2015). Also, the temperature has a reductive effect on particulate matters through the convectional method (Li et al. 2015), and it is an important variable for secondary PM$_{10}$ formation (Taheri Shahraiyni and Sodoudi 2016). Wind speed influences the transportation of ambient pollutants (Wang and Ogawa 2015; Ganguly et al. 2019) while relative humidity makes PM$_{10}$ heavier and also aids the dry deposition removal (Kleine Deters et al. 2017). Recent studies have highlighted a correlation between air pollution and seasonal variation since different pollutants exhibit diverse properties at different times of the year due to climatic variability (Wang and Ogawa 2015; Kleine Deters et al. 2017).

An accurate assessment of the correlation between climatic variables and air pollutants is crucial for projecting PM$_{10}$ concentration in cities (Althuwaynee et al. 2020). The knowledge of the nexus between air pollutants and climatic variables will ensure a proper understanding of how it influences air pollution. Thus, understanding this inter-relationship will assist in the mitigation, monitoring, and control of air pollution emission, transportation, and distribution. Also, identifying the climatic variables that exert the most significant impact on air pollutants and the period when the pollutants are most lethal will assist stakeholders to prioritize intervention plans in the right context. For example, understanding the climatic variables that exhibit high positive correlation and high predictive performance as a predictor will provide nuggets of information on the leading climatic factors causing air pollution.

However, it is noteworthy that the interdependence of air pollutants and climatic conditions vary across regions, which makes it difficult to generalize trends. For example, while the temperature is highly correlated with PM$_{10}$ in Seoul in Korea (Kim 2019), its correlation outcome in Auckland in New Zealand (Hernandez et al. 2017) is different. Similarly Giri et al. (2008) obtained a negative correlation between temperature and PM$_{10}$ in the Himalayan Kingdom of Nepal while (Sharma and Sharma 2016) got varied and positive correlation index in four different seasons in Northern India.

Considering the influence of climatic conditions on air pollution occurrence and management in other regions, it is crucial to accurately assess the correlation of climatic variables with PM$_{10}$ in Malaysia too, taking into consideration its peculiar seasonal variations and regular occurrence of wildfire and haze pollution.

1.2. Modelling air pollutants

Different methods exist for modelling air pollution. The deterministic method which uses air transport, release, and chemical modules (Zhou et al. 2019) is often utilized (Djalalova et al. 2015; Choubin et al. 2020). Statistical models such as linear regression have evolved to show higher predictive performance and accuracy than the deterministic method (Song et al. 2015). Regression has been used to accurately...
predict air pollution because of its simplicity in computation and implementation (Abdullah et al. 2017; Fong et al. 2018). The most widely adopted regression method is the multi-linear regression model which utilizes climatic factors as predictors for modelling air quality.

Previous studies have implemented statistical models for PM$_{10}$ concentration prediction. Nazif et al. (2019), Fong et al. (2018), and Abdullah et al. (2017) used multiple regression for forecasting air pollution in Malaysia. However, many of these studies do not model the spatial distribution of PM$_{10}$, which can aid the prompt detection of hazardous areas. Spatial modelling of pollutants is essential because it can help stakeholders and decision-makers to determine regions that have poor air quality.

Further, studies that consider seasonal variation in modelling the emission and distribution of PM$_{10}$ in Malaysia are scant. Investigating the emission trend of PM$_{10}$ in different months, years, and seasons will offer further insights on the impacts of climatic variables such as temperature, wind speed and humidity on the pollutant and enable evidence-based intervention. The main contributions of this paper are as follows:

i. The assessment of the impacts of climatic variables on PM$_{10}$ concentration, taking into consideration seasonal variation. This approach reveals the effect of climatic variables such as temperature, humidity, and wind speed on PM$_{10}$ variation in Malaysia.

ii. Development and validation of an algorithm to predict the concentration of PM$_{10}$. This will assess the potential of climatic variables to predict PM$_{10}$ concentration.

iii. Leverage advanced geospatial techniques to model the PM$_{10}$ hazard level. This ensures better visualization, analysis and interpretation of the model’s result.

2. Study area

Malaysia (Figure 1) is an Asian country with latitude and longitude 4.2105° N, and 101.9758°, respectively (Pour et al. 2020). The federal constitutional monarchy consists of 13 states and three federal territories. The country is separated by the South China Sea into two regions, namely peninsular and East Malaysia. The study area was once dominated by palm oil plantation. However, there is a rapid increase in population due to urban sprawl, especially in peninsular Malaysia (Andaya and Andaya 2016). Its population is over 30 million with a land area of 320,000 km$^2$ (123,553 sq mi) (Ab Rahman et al. 2013). It has an equatorial climate with a warm and humid climate throughout the year (Tang 2019). The temperature ranges from 25.5°C to 33°C and is moderated by the presence of the surrounding oceans while annual rainfall ranges between 200 cm and 400 cm.

Malaysia has two main seasons also referred to as ‘monsoons’. The northeast monsoon starts from November to March, while the southwest monsoon extends from May to September (Fong et al. 2018; Nazif et al. 2019). The two transitional periods are between April-May and October–November, called intermonsoon 1 and 2, respectively (Abdullah et al. 2017). According to Andaya and Andaya (2016), the monsoon winds are associated with rainfall variation; however, the wettest season corresponds to the northeast monsoon while the southwest monsoon is characterized...
by temperate weather. Changes in the monsoon pattern occur due to El Nino effects which are prevalent in the southwest monsoon (Ab Rahman et al. 2013). This causes a rise in temperature and a more extended period of warm weather (Andaya and Andaya 2016). According to Baker (2008), it cannot be fully established that there is a dry season in Malaysia because even in Southwest monsoon, there will be rainfall for at least 7 days which makes the two seasons either rainy or more rainy season.

3. Methodology

3.1. Data utilized

The data used in this study are the meteorological and air quality data obtained from the Department of Environment (DOE) Malaysia. Climatic variables include the hourly temperature (°C), humidity (%), wind speed (m/s), and the air quality data, which comprises the hourly concentration of PM_{10} (μg/m^3). These data, covering the period from 2012 to 2016, were acquired from five air monitoring stations (Figure 2). The stations are strategically mounted around industrialized and urbanized regions (Ahmat et al. 2015) across five states in Peninsular Malaysia. The Bukit Rambai and Nilai air quality stations are located in the industrial hub in Peninsular Malaysia; Cederawasih, Taman Inderawasih, Perai station in Pahang, the second most populated state; the Klang air quality station is situated in Klang valley, the trade centre, and industrialized region in the most populated state (Selangor) in Malaysia (Ahamad et al. 2014). The data was classified into months, years, and seasons for analysis. The seasons considered in this study are the southwest and northeast monsoon. The spatial distribution of these stations is displayed in Figure 2.
3.2. Data analysis

First, a thorough investigation was carried out on the PM$_{10}$ concentration to understand its distribution and behaviour in Malaysia. To achieve this, the PM$_{10}$ concentration is studied based on months (January–December), years (2012–2016), and seasons (Southwest and Northeast Monsoons). This ensures a better understanding of the temporal and seasonal behaviour of PM$_{10}$. Then, we observed and compared the relationship between the meteorological parameters and the atmospheric PM$_{10}$ concentration. This was done by studying the monthly, yearly, and seasonal distribution of the three climatic variables and PM$_{10}$. Pearson’s coefficient of correlation was used to investigate the inter-link between the climatic factors and PM$_{10}$ adopting the methodology of some related studies such as Rahman et al. (2019). Afterwards, the capability of multivariate regression to predict PM$_{10}$ using climatic variables as a predictor was investigated. The PM$_{10}$ modelling was done in python, while other analysis was carried out in Minitab software. Finally, the PM10 hazard susceptibility map of the study area was generated in the ArcGIS environment using the regression model. The map was generated using the coefficient prediction value of the model as earlier employed by Bozdağ et al. (2020) through the application of geostatistical analysis. Also, the spatial distribution of PM$_{10}$ concentration was generated based on the Air Pollution Index (API) format of the DOE Malaysia using spatial interpolation analysis in ArcGIS.
Spatial interpolation is a common geostatistical technique used to map air pollution (Jumaah et al. 2019). This approach involves using the point value of known regions to estimate unknown regions’ point value (Ma et al. 2019) within the spatial range of the known values. This method depends on the number and evenly distribution of point values, limiting its performance or spatial variability (Tian 2016; Kim et al. 2014). However, this approach is still widely utilized for spatial assessment and distribution of air pollutants (Bozdağ et al. 2020; Shukla et al. 2020; Ahmed et al. 2018; Ma et al. 2019). Among the different types of interpolation techniques, IDW is one of the most widely used. The IDW can be used for multivariate interpolation (Chen and Liu 2012). The performance of IDW has been proven in air pollution studies (Jumaah et al. 2019; Kumar et al. 2016). Thus, Inverse Distance Weighting (IDW) was employed in this study.

### 3.3. Multivariate regression (MVR)

Multivariate regression is used to determine the relationship between a dependent variable and multiple independent variables (Abdullah et al. 2016). The model accounts for the influence of the multiple control variables on a single dependent variable. It is also used to check the impacts of one or more independent variables on the dependent variable (Abba et al. 2017). An advantage of MVR is that the observable changes in the dependent variable can be explained by the explanatory variable and the contribution of each of these variables can be easily determined. The equation of multivariate regression is shown in Eq. (1)

\[
Y = b_0 + b_1X_1 + b_2X_2 + \ldots + b_nX_n + \varepsilon
\]  

(1)

where \(Y\) is the dependent variable, \(b_0\) is the constant y-intercept, \(b_1, b_2, \ldots, b_n\) are the regression slope coefficient for the independent variables, \(X\) is the independent variables and \(\varepsilon\) is the residuals.

Although simple linear regression has been implemented in air pollution studies, it has a low predictive accuracy for particulate matters (Wang and Ogawa 2015). In contrast, multiple linear regression has a higher probability of getting a better model fit than simple linear regression due to the inclusion of more than one explanatory variable that determines the prediction of the dependent variable (Gupta 2019). Thus, the capability and strength of the explanatory variables in predicting the dependent variable using MVR will be determined. Moreover, with the application of MVR, the significance of the independent variables (e.g. climatic factors) in predicting the dependent variable (PM\(_{10}\)) will be determined. Understanding the multicollinearity between the independent variables is fundamental to determine the statistical significance of the independent variables (Allen 1997; Yoo et al. 2014; Daoud 2017) in predicting the dependent variable. Thus, pairwise correlation analysis was performed to determine the problem of multicollinearity in the independent variables. The following statistical indices (Table 1) were used to evaluate the model’s performance.
4. Results and discussion

The statistical description of the variables considered for this study is presented in Table 2. The total mean concentration of PM$_{10}$ recorded for 12 months is 55.77 µg/m$^3$. The total maximum mean concentration of PM$_{10}$ in 12 months (68.33 µg/m$^3$) exceeded the guidelines set by the DOE, Malaysia in controlling the PM$_{10}$ concentration, as shown in Table 3. The guidelines state that the annual concentration of PM$_{10}$ should not exceed 50 µg/m$^3$ (Department of Environment 2013). However, this is even higher than the limit set by the European Union (EU) (40 µg/m$^3$) (Khan et al. 2015) and the World Health Organization (20 µg/m$^3$) (World Health Organization 2006) to reduce poor air quality. Exceeding the guidelines may be influenced by the biomass and bush burning used in Indonesia for land cultivation. The smokes from this combustion are transported to Malaysia, a neighbouring country (Abdullah et al. 2020), under the influence of wind and the seasons (Kusumaningtyas and Aldrian 2016).

### 4.1. Variation in PM$_{10}$ concentration in Malaysia

The time series of the monthly mean concentration of PM$_{10}$ is displayed in Figure 3. The monthly mean concentration ranges from 44.08 µg/m$^3$ to 68.33 µg/m$^3$. Approximately 67% of the monthly concentration of PM$_{10}$ for five years is above the Malaysian Ambient Air Quality Guidelines (MAAQG). In terms of the monsoons, 20% of the high PM$_{10}$ concentration above 50 µg/m$^3$ occurred in the northeast monsoon while 40% of the high PM$_{10}$ concentration above 50 µg/m$^3$ occurred during the southwest monsoon.

Specifically, a trend of high concentration of PM$_{10}$ is observed in June, October, and March, respectively. The highest concentration of PM$_{10}$ occurs in June (68.33 µg/m$^3$), the second month that precedes the beginning of the Southwest monsoon while the lowest concentration of PM$_{10}$ falls in November (44.08 µg/m$^3$) which is the starting month for the Northwest monsoon. This finding aligns with the work of Hassan et al. (2020) whereby the highest PM$_{10}$ concentration started and was

| Indicators               | Description                                                                 | Equation                                                                 |
|-------------------------|-----------------------------------------------------------------------------|-------------------------------------------------------------------------|
| Root Mean squared Error (RMSE) | It shows the standard deviation of the variation in the predicted and true values. Thus, it indicates the potential of having mispredicted values. Lower values reflect better predictive performance. | RMSE = $\sqrt{\frac{1}{n} \sum_{i=1}^{n} (b_i - a_i)^2}$ |
| Coefficient of determinant (R$^2$) | It expresses the model’s fitness with the range of −1 to +1. A negative value indicates a low performance while positive values that extend towards +1 indicate a high predictive capability. | $R^2 = \frac{n(B\bar{a} - \bar{a}B)(\bar{b}\bar{a} - \bar{a}\bar{b})}{\left(\left[n\sum b^2 - (\sum b)^2\right]\left[n\sum a^2 - (\sum a)^2\right]\right)}$ |
| Mean Absolute Error (MAE) | MAE evaluate the predictive accuracy of the prediction by estimating the difference between the mean absolute value between the predicted and true value. | MAE = $\frac{1}{n} \sum_{i=1}^{n} |b_i - a_i|$ |
dominant in the southwest monsoon. Thus, high PM$_{10}$ concentration is correlated with the warming event in the monsoon. Also, the lowest concentration of PM$_{10}$ in the starting month of the northeast monsoon might be due to the abundant rainfall in this monsoon compared to the southwest monsoon. Rainfall enables the reduction of PM$_{10}$ in the atmosphere (Chu 2015) by wet deposition (Luan et al. 2019). Therefore, it is inferred that as temperature affects air quality, rainfall aids the removal of air pollutants in the atmosphere in the study area.

Table 2. Descriptive statistics of the dataset.

| Variables | Mean | Minimum | Maximum |
|-----------|------|---------|---------|
| PM$_{10}$ | 55.77 | 44.08 | 68.33 |
| SW | 60.05 | 44.19 | 72.01 |
| NE | 50.29 | 37.78 | 60.23 |
| Wind speed | | | |
| Months | 6.15 | 5.18 | 7.44 |
| SW | 5.97 | 5.08 | 6.72 |
| NE | 6.470 | 4.140 | 8.01 |
| Temperature | | | |
| Months | 29.25 | 27.61 | 30.65 |
| SW | 29.28 | 28.54 | 30.30 |
| NE | 29.04 | 28.38 | 30.33 |
| Humidity | | | |
| Months | 74.60 | 70.61 | 80.17 |
| SW | 72.65 | 69.85 | 76.33 |
| NE | 75.65 | 73.47 | 79.44 |

Note: SW: Southwest NE: Northeast.

Table 3. Correlation analysis of meteorological variables to PM$_{10}$ in air monitoring stations.

| Air quality stations | Wind Speed | Temperature | Humidity |
|----------------------|------------|-------------|----------|
| Bukit Rambai         | $-0.220$  | $0.083$     | $-0.060$ |
| Nilai                | $-0.560$  | $0.297$     | $-0.023$ |
| Jalan Tasek Ipoh     | $0.337$   | $0.600$     | $-0.632$ |
| Cederawasih, Taman Inderawasih, Perai | $0.661$ | $0.343$ | $-0.095$ |
| Kelang               | $0.497$   | $0.250$     | $-0.679$ |

Figure 3. Monthly time series analysis of PM$_{10}$. 
This has been earlier done by some studies (Özdemir and Taner 2014; Nazif et al. 2019) which used the boxplot to understand the concentration and trend of PM10 and to compare the PM10 concentration either in months or seasons.

The box plot, which is a graphical expression for devising comparative discussion and understanding the different variables (Abdullah et al. 2017) is shown in Figure 4. Noteworthy that a distribution is skewed when a tail is longer than the other. Thus, it can be inferred from Figure 4 that the concentration of PM10 is skewed to the left due to the tail pointing towards the negative direction. The negative skewness may indicate that there were more moderate to low concentration of ambient PM10 than the concentration represented by the positive skewness in Ng and Awang (2018). The box plot also depicts the effect of seasonal changes in the formation and accumulation of PM10 in the atmosphere. It shows the average monthly concentration of PM10 in different seasons. A higher average concentration of PM10 in the southwest monsoon (72.01 µg/m³) was recorded compared to the northeast monsoon (60.23 µg/m³). A higher concentration in the southwest monsoon that represents a drier season may be impacted by a rise in temperature in this monsoon (Kassomenos et al. 2014).

Basically, Malaysia is a tropical country with high temperatures (Abdullah et al. 2017). Anthropogenic activities such as combustion will undoubtedly affect the warming in the country especially during the dry season. According to Abdullah et al. (2020), the transboundary haze that affects the air quality in Malaysia does occur more often during the southwest monsoon with lesser rainfall. Temperature also caused a variation in the circulation of wind (a medium for transboundary pollution) in the atmosphere (Abdullah et al. 2017). Thus, the 9% higher concentration of PM10 in the warm southwest monsoon season than the wet northeast monsoon.

This finding is supported by some previous studies on the causes of variation in the concentration of air pollutants. Hassan et al. (2020) concluded that the high concentration of PM10 in Malaysia in the southwest monsoon is due to the warming effect in the dry season (southwest monsoon). Also, a high concentration of PM10

![Figure 4. Box plot of the mean monthly concentration of PM10 in the two monsoons.](image-url)
was observed in the warm season in the study of Kassomenos et al. (2014), which examined the variation in the atmospheric PM$_{2.5}$ and PM$_{10}$ in European cities. Nazif et al. (2019) also discovered that seasonal changes influence the PM$_{10}$ variation in the atmosphere with a higher concentration of PM$_{10}$ in the warm season. According to Wang and Ogawa (2015), warm-seasons impact the formation of atmospheric particles by enhancing the physical and chemical reactions of atmospheric precursors.

Figure 5 shows the trend of the PM$_{10}$ concentration with variations in the annual PM$_{10}$ level from 2012 to 2016. The trend of the annual concentration of PM$_{10}$ shows that they exceed the recommended Malaysian Air Quality Guidelines (RMG) of 50 $\mu$g/m$^3$. The highest concentration values are recorded in 2013 (56.81 $\mu$g/m$^3$) and 2015 (63.73 $\mu$g/m$^3$). The relatively higher concentration in 2013 and 2015 is likely due to the haze pollution in Malaysia, which exacerbates poor air quality. PM$_{10}$ has been linked to haze episodes, a significant challenge in Southeast Asia since the 1980s, due to the combustion of biomass (Shaadan et al. 2015). The transboundary pollution from Indonesia wildfire has been the main causative factor of haze pollution in Malaysia (Wen et al. 2016). The smokes generated from the wildfire are dispersed and transported through wind which pollutes the atmosphere across borders. Studies have shown that wind speed influences the transportation of ambient atmospheric pollutants to other regions (Zhang et al. 2014). The haze event in 2015 caused a decline in the air quality in Malaysia Peninsular Sulong et al. (2017). Also, according to a study by Hassan et al. (2020), severe El Niño in dry seasons affects the weather, especially in Malaysia peninsular, which in turn influences the PM$_{10}$ variation. This study's outcome also corroborates the earlier research by How and Ling (2016) in finding the impact of particulate matters on API during haze and non-haze periods. Their study identified particulate matter as the major contributor to air quality variation in Malaysia, highlighting a higher concentration of PM$_{10}$ during the haze period.

Figure 5. Annual time-series analysis of PM$_{10}$.
in comparison to the non-haze period. The lower concentration in some years (Figure 5) is likely to have been influenced by the El Niño effect coupled with the transboundary and regional emission sources.

Thus, PM$_{10}$ concentration tends to be more dominated in the southwest monsoon with high warming in western Malaysia. In the northeast monsoon, there exists a lower concentration of PM$_{10}$ due to abundant rainfall. Climatic variables, such as temperature and wind, influence the concentration and variation of atmospheric PM$_{10}$, in conjunction with the El Niño effect.

### 4.2. Inter-relationship between climatic variables and PM$_{10}$

It is noteworthy that analyzing the association between climatic variables and PM$_{10}$ is intricately complex and sometimes not linearly related (Czernecki et al. 2017). Nevertheless, a Pearson correlation analysis was carried to examine the linear association between the climatic variables and the PM$_{10}$ concentration.

Most of the fixed monitoring stations are located around residential areas, which according to Rani et al. (2018) contribute to high accumulation of atmospheric PM$_{10}$. Thus, according to the authors, this location aids the high Air Pollution Index (API) value due to the atmospheric distribution of air pollutants. This is due to meteorological conditions that determine the emission and distribution strength, spatial location, and the seasons of this emission (Folberth et al. 2015). Pearson’s correlation was used to ascertain the degree of linear correlation between the climatic variables and PM$_{10}$. The correlation analysis was performed based on the geographic location of the stations and seasons. The correlation index of the climatic variables (humidity, temperature, and wind speed) with PM$_{10}$ is shown in Tables 3 and 4.

We observed that temperature showed a positive correlation with PM$_{10}$ concentration in all the five stations, which agrees with some studies in some other countries. For instance, Lee and Kim (2018) obtained a positive correlation between temperature and PM$_{10}$ in Seoul, Korea while analyzing the effect of climatic variables on PM$_{10}$ variation. Similarly, Dotse et al. (2016) obtained a positive correlation with PM$_{10}$ while investigating the impacts of haze on PM$_{10}$ concentration in Brunei Darussalam. The outcome supports the notion that temperature influences the PM$_{10}$ formation and its variability in the atmosphere. This is due to the impacts of high temperature inducing combustion activities and evaporation from the earth’s surface in Malaysia. The correlation result explains the reason for the urinary correlation between temperature and PM$_{10}$ in this study.

Analysis of the seasons reveals that temperature exhibits a higher correlation index with PM$_{10}$ during the wet northeast monsoon than the warmer southwest monsoon. Studies such as Huang et al. (2016) and Hernandez et al. (2017) had earlier concluded that the concentration of PM$_{10}$ declines in extremely hot weather conditions.

### Table 4. Correlation analysis of meteorological variables to PM$_{10}$ on a seasonal basis.

| Meteorological parameters | Northeast monsoon | Southwest monsoon |
|---------------------------|-------------------|-------------------|
| Temperature               | 0.368             | 0.134             |
| Humidity                  | -0.530            | 0.079             |
| Wind speed                | 0.532             | 0.257             |

A. TELLA ET AL.
However, humidity tends to show a negative correlation with PM$_{10}$ concentration in all five air monitoring stations (Table 3). This is logical because high humidity is associated with high precipitation, which decreases PM10 concentration in the atmosphere because high humidity is associated with high rainfall (Ng and Awang 2018; Gvozdić et al. 2011).

This outcome aligns with the study of Wang and Ogawa (2015) in Nagasaki Japan to study the effect of meteorological factors on particulate matter. A negative correlation is also gotten in the study of Afzali et al. (2014) in Pasir Gudang, Johor. The authors used monitoring stations to determine the effects of meteorological variables on PM$_{10}$. Comparatively, a negative correlation was observed in the southwest monsoon while a positive but very low correlation index exists in the northeast monsoon. This result indicates that the relationship between PM$_{10}$ and humidity is not linear since an increase in humidity reduces the concentration of PM$_{10}$. Zhang et al. (2015) reported a similar relationship between humidity and PM$_{10}$ in three metropolitan cities in China. Two of these cities (Shanghai and Guangzhou) exhibit a negative correlation between humidity and PM$_{10}$.

The wind speed shows a positive to negative correlation with PM$_{10}$ concentration in Bukit Rambai and Nilai. Also, a positive correlation is observed in the monsoons, with the highest correlation in the northeast monsoon. Generally, high wind speed shows a positive correlation with PM$_{10}$ in the study area. High wind speed can increase the velocity of transportation of air pollutants to other regions as opposed to low wind speed. An increase in wind speed accelerates air pollutant dispersion in the atmosphere (Afzali et al. 2014). This implies that low wind speed influences the accumulation of PM$_{10}$ in the atmosphere. The negative correlation of wind speed and PM10 in some of the air quality stations may be due to the large ventilation of the air masses in these stations and the deracination of the air pollutants over a remote area. As shown in the study of Cichowicz et al. (2020), wind speed exhibited an inverse relationship with PM$_{10}$ in some air monitoring locations in Poland. That is, higher wind speed corresponds to a lower PM10 concentration and vice-versa. Also, the location of the stations is another influencing factor considering the conurbation and emission from vehicle fleet. This aligns with the outcome of earlier studies such as Czernecki et al. (2017) study of the impacts of atmospheric conditions of PM10 in Poland.

Thus, the result of the correlation analysis at the five monitoring stations indicates that high wind speed and temperature are major contributors to the dispersion of atmospheric pollutants in Malaysia, confirming the findings of Folberth et al. (2015) and Rani et al. (2018). Also, both temperature and wind exhibit a positive correlation with PM$_{10}$ in the northeast monsoon and southwest monsoon.

4.3. Prediction of PM$_{10}$ concentration using MLR

Due to the numerous assumptions about the regression model, which can lead to bias or unacceptable estimation of dependent variables, multicollinearity among the independent variables, was investigated. The multicollinearity explains the degree of correlation between the independent variables. A high correlation between the
independent variables pose a problem, while a moderate correlation can be solved (Daoud 2017). Table 5 shows the relationship between the independent variables.

According to Table 5, all the three variables are uncorrelated except the relationship between the wind speed and temperature. This aligns with a study in Malaysia also by Zakaria et al. (2017), with correlation index of 0.633 between the wind speed and temperature. We performed the regression model, then exclude one of these variables to observe the changes in the model output. This is done in order to resolve the multicollinearity, because the omission of a variable that is correlated to the other in the model can be used to solve the problem of multicollinearity (Daoud 2017). Therefore, the regression model was performed out for all the variables before the omission.

Multiple linear regression (MLR) was used to predict the concentration of PM$_{10}$. The predictors considered for this model are the climatic variables which include temperature, humidity, and wind speed. The dataset was divided into training and testing. 80% of the data was used to train the data, while the remaining 20% was used to test and validate the prediction in python software.

The statistical indices used are $R^2$, MAE, and RMSE. The generated MVR equation for the three predictors is represented in Eq. (2).

\[
\text{PM10} = -251.20 + 1.31 \text{Wind Speed} + 4.03 \text{Temperature} - 0.89 \text{Humidity} \quad (2)
\]

Figure 6 shows the residual plots of the PM$_{10}$ regression model prediction. The acceptability and reliability of a linear regression model are usually validated by residual plots (Law and Jackson 2017). Residual value is the difference between the observed and the predicted value. The residuals (error) are essential for determining the performance of a statistical model (Abdullah et al. 2017). The adequacy of a linear regression model is a factor of homoscedasticity (Law and Jackson 2017). It explains if the residuals are the same and evenly distributed. The normal probability plot of the residuals showed a normally distributed variance (e.g. it shows how well the data is distributed). Normal distributed data is obtained when the data are close to the straight line (Adio et al. 2019).

In this study, the plot of the normal probability is linear (Figure 6), thus, we postulate that the error terms are normally distributed. The versus fits show the fitness of the model, which is used to test the goodness of fit (Bowen 2018). In order to show the validity of the assumption of this model, a scatterplot of the residuals is plotted against the fitted value. The distance from 0 explains how good, or bad the prediction is, that is if the regression model is under-predicted, overpredicted, or correctly predicted.

Our model showed a symmetrically distributed prediction because the clustering tends towards the centre of the plot, which is zero (0). From this, we establish that

**Table 5. Correlation of independent variables.**

| Variables | Temperature | Humidity | Wind speed |
|-----------|-------------|----------|------------|
| Temperature | 1           |          |            |
| Humidity   | -0.92       | 1        |            |
| Wind speed | 0.56        | -0.71    | 1          |
our regression model is acceptable and appropriate to an extent for the variables used. This is because the points are randomly distributed around the horizontal axis.

Also, the regression model is at 95% confidence interval with a p-value of 0.000 and standard error (S) of 12.8.

The statistical models used to rank the performance of the prediction of PM$_{10}$ concentration is shown in Table 6. Generally, the MVR’s predictive accuracy is acceptable with very low RMSE and a positive and high coefficient of determinants. The coefficient of determinants revealed that approximately 30% of the variations in PM$_{10}$ is influenced by the climatic variables. A low RMSE and MAE values are indicative of the model’s acceptable predictive performance. Thus, this model has a RMSE and MAE of 12.737 and 10.343, respectively. Some previous studies have also established the accuracy of MVR for air pollution prediction. The study of Karatzas et al. (2018) to predict urban air quality in Poland recorded a high predictive performance for MVR, outperforming the Artificial Neural Networks (ANN) results. The RMSE of the model ranges from 6.690–22.391, while the model has MAE of 6.380 to 15.050. Afzali et al. (2014) also obtained an acceptable predictive performance ($R^2 = 0.18$) in Johor using a multiple regression model. Thus, MVR has proven to have better performance for the prediction of PM$_{10}$ concentration in the study area.

A histogram representing the actual and the predicted concentration of PM$_{10}$ is presented in Figure 7. Analysis of Figure 7 shows the difference between the observed and predicted, with little margin.

Temperature and wind are critical factors that showed a strong influence on the prediction. This is because the regression model chooses the coefficients of these two variables, as shown in equation 2. The temperature has the most considerable absolute value, followed by wind speed. To further confirm the effects of the individual variables on the linear regression model, the Pareto chart was used. According to Harvey and Sotardi (2018), the Pareto chart is used to determine the relative importance of parameters to the model result. It consists of variables with their association against the standardized effects (Adio et al. 2019). The horizontal axis represents the t-test at 95% confidence level. The vertical axis consists of the predictors used for the regression model. At a 95% confidence level, the value of the t-test is 1.963 with a vertical dash line representing the significance level. Figure 8 showed that all the predictors are significant. In order of significance, temperature, wind speed, and humidity showed a higher value, respectively.
To further investigate the significance of the variables in the output model, an exclusion of a variable was carried out. Table 7 shows the regression output of the variable omission. The exclusion of one of the variables affects the model outcome. The exclusion of temperature reduces the model performance with a $R^2$ of 0.237 compared to the overall model performance of 0.298. The MAE and RMSE of 10.618 and 13.280, respectively, also exceed the model performance of the three variables, suggesting that temperature is an essential variable in the model. The omission of humidity also reduces the performance of the model to a $R^2$ of 0.238. Also, the MAE of 10.782 and RMSE of 13.272 exceed the model outcome of the three parameters, meaning that the performance of the model reduces due to the exclusion of humidity. Also, the regression model performance reduces with the exclusion of wind speed given output with $R^2$ of 0.240, MAE of 10.678 and RMSE of 13.255.

**4.4. Spatial prediction of PM$_{10}$**

After the regression model has been tested, the coefficients of the independent variables were used to produce the PM$_{10}$ hazard susceptibility map. The hazard susceptibility map was further classified into five classes, including the very high, high, medium, low, and very low. This was generated in the ArcGIS interface using a natural break (Jenks) classification. The map of the independent variables is shown in Figure 9.

![Figure 7](image.png)

Figure 7. Histogram showing the relationship between the actual and predicted values of PM$_{10}$.

To further investigate the significance of the variables in the output model, an exclusion of a variable was carried out. Table 7 shows the regression output of the variable omission.

| Model | $R^2$ | RMSE  | MAE  |
|-------|-------|-------|------|
| MVR   | 0.298 | 12.737| 10.343|

Table 6. The model performance of the MVR.
Figure 10 shows the PM10 hazard map. The PM10 hazard map is classified into five classes ranging from very low, low, moderate, high, and very high classes. The regions with a high concentration of PM10 are more susceptible to greater health impacts than others with low concentration. For instance, in Malaysia’s API table (APIMS 2020), areas with high air pollution are susceptible to dangerous health conditions such as lung and respiratory complications. Similarly, very high PM10 hazard level is risky for public health, with the possibility of causing deaths if not well managed (Manisalidis et al. 2020). Conditions in areas with a moderate concentration of PM10 may deteriorate to an unhealthy level while low and very low hazard classifications pose no significant threats to the community’s health.

The very high hazard classes covered 8153.82 km² (20.57%), high covered 3263.01 km² (8.23%), moderate covered 10,897.76 km² (27.49%), low covered 16,005.31 km² (40.38%), and very low PM10 hazard classes covered 1318.69 km² (3.33%). The higher percentage of the PM10 hazard in the study area is majorly between moderate (27.49%) and low (40.38%) PM10 hazard level (Table 8). In this study, the climatic variables have shown the effectiveness of predicting PM10. However, further work is required to include some other spatial data such as road networks.

4.5. Spatial distribution of \( PM_{10} \)

An Air Pollution Index (API) susceptibility map for the study area was produced using data acquired from 2012 to 2016. To get the API, the sub-index value of PM10...
was calculated using the formula in Table 9 (Rani et al. 2018). Most often, PM$_{10}$ is the dominant pollutant in the API value (Department of Environment 2019). For instance, PM$_{10}$ was discovered to be the most contributing factor to the haze event in 1997. Thus, the API values were based on the PM$_{10}$ sub-index (Department of Environment Malaysia 1997). However, not quite long, PM$_{2.5}$ concentration became the highest among other pollutants and is now used to determine the API value. But due to data accessibility constraints and considering the sustained interest in the impacts of PM10, this current study utilized PM10 to determine the API value.

Figure 10 represents the generated API risk map of the area of study. The API values range from 41 to 58 denoting a good to moderate air quality. According to the Department of Environment (2019), this does not pose any threat to human health. This generated API map (Figure 11) has a significant correlation with the PM$_{10}$ hazard level above (Figure 10). Also, according to the Air Pollution Index of Malaysia (APIMS), the API level ranges from moderate to good (http://apims.doe.gov.my/public_v2/api_table.html). Thus, plans and regulations should be utilized to maintain good air quality for the sustainability of human and the environment.

Figure 9. Independent variables (a) wind speed, (b) temperature, (c) humidity.
5. Conclusion

This study investigated the impacts of three climatic factors on PM$_{10}$ concentration in Malaysia, considering seasonal variation. The choice of PM$_{10}$ is due to being the dominant pollutant in deriving the API value adopted in Malaysia. Considering the recent availability and potential accessibility to PM2.5 data, it is recommended that future studies focus on PM2.5. Correlation analysis and a multivariate linear
A regression model was used to determine the relationship between the climatic variables (such as temperature, wind speed, and humidity), and PM\textsubscript{10}. A relatively high and positive correlation index was gotten for temperature and wind speed; however, humidity negatively correlates with PM\textsubscript{10}. The regression analysis also showed a high predictive performance, which suggests an inter-link between the climatic variables and PM\textsubscript{10}. The results of the MVR model showed, a positive correlation and high accuracy with statistical indices of R\textsuperscript{2} (0.30), RMSE (12.74), and MAE (10.34) was obtained. The critical factors were identified, and the Pareto chart was used to determine the model’s performance vis-a-vis the significance of each criterion. Although all the variables exhibit a significant influence on PM\textsubscript{10}, however, the chart indicates that temperature and wind speed are the most influential factors impacting PM\textsubscript{10} in the study area. This corroborates the outcome of the correlation analysis. Given the nexus between the climatic variable and PM\textsubscript{10} established in this study, the projected variation in the mass concentration of PM\textsubscript{10} due to climate change in the future is feasible. According to the spatial modelling result of PM\textsubscript{10} hazard level, a large percentage of the study area has very low (3.33%), low (40.38%), moderate (27.49%), high (8.23%) and very high (20.57%) of PM\textsubscript{10} hazard level in the study area. The API map also showed a good and moderate air quality, which is not dangerous to human health. However, the increasing trend of climate change and urbanization in the World should be considered for a possible increase in PM\textsubscript{10} concentration also in the future.

Some other findings of this study are: the annual mean concentration of PM\textsubscript{10} exceeds the MAAQG guidelines. Also, the season plays a vital role in influencing the accumulation and concentration of PM\textsubscript{10}. According to this study, the PM\textsubscript{10} concentration in the southwest monsoon is 9% greater than that of the northeast monsoon. Given the haphazard behaviour of the climatic variables with respect to the seasons

![Figure 11. API map of the study area.](image-url)
and climate change effects, a need to formulate and re-formulate policies/guidelines to monitor the PM$_{10}$ concentration with respect to these effects is necessary. Considering the effects of climate change on warming one of the causative factors of bushfire in Malaysia, awareness about the multiple impacts of this on the atmosphere and health should be made to citizens. This will ensure a fast means to curb bad air quality from this source. Finally, carbon emission should be regulated in urbanized and industrialized areas, while green infrastructure should be encouraged and maximize for solving urban and climate impacts in this context.

**Acknowledgements**

The authors gratefully acknowledge the Department of Environment (DoE), Malaysia, for providing the air quality data used in this study and the Federal Department of Town and Country Planning (PLANMalaysia) for providing spatial and attribute data of the study area.

**Disclosure statement**

No potential conflict of interest was reported by the authors.

**ORCID**

Abdulwaheed Tella [http://orcid.org/0000-0002-4380-3343](http://orcid.org/0000-0002-4380-3343)
Abdul-Lateef Balogun [http://orcid.org/0000-0002-0418-3487](http://orcid.org/0000-0002-0418-3487)

**Data availability statement**

The data that support the findings of this study are available from Department of Environment (DoE), Malaysia. Restrictions apply to the availability of these data, which were used under license for this study. Data are available from the first and/or corresponding author with the permission of the Department of Environment (DoE), Malaysia.

**References**

Ab Rahman AK, Abdullah R, Balu N, Shariff FM. 2013. The impact of La Niña and El Niño events on crude palm oil prices: an econometric analysis. Oil Palm Ind. Econ J. 13:38–51.
Abba SI, Hadi SJ, Abdullahi J. 2017. River water modelling prediction using multi-linear regression, artificial neural network, and adaptive neuro-fuzzy inference system techniques. Procedia Comput Sci. 120:75–82.
Abdullah S, Ismail M, Fong SY. 2017. Multiple linear regression (MLR) models for long term PM$_{10}$ concentration forecasting during different monsoon seasons. J Sustain Sci Manag. 12: 60–69.
Abdullah S, Ismail M, Fong SY, Najah A-M. 2016. Evaluation for long term PM 10 concentration forecasting using multi linear regression (MLR) and principal component regression (PCR) models. EnvironmentAsia.
Abdullah S, Napi NNLM, Ahmed AN, Mansor WNW, Mansor AA, Ismail M, Abdullah AM, Ramly ZTA. Ramly. 2020. Development of multiple linear regression for particulate matter (PM10) forecasting during episodic transboundary haze event in Malaysia. Atmosphere 11(3):289. DOI:10.3390/atmos11030289
Adio SO, Asif M, Mohammed A-RI, Baig N, Al-Arfaj AA, Saleh TA. 2019. Poly (amidoxime) modified magnetic activated carbon for chromium and thallium adsorption: statistical analysis and regeneration. Process Saf Environ Prot. 121:254–262.

Afzali A, Rashid M, Sabariah B, Ramli M. 2014. PM10 pollution: its prediction and meteorological influence in Pasir Gudang, Johor. IOP conference series: earth and environmental science. IOP Publishing. p. 012100.

Ahmad F, Latif MT, Tang R, Juneng L, Dominick D, Juahir H. 2014. Variation of surface ozone exceedance around Klang Valley, Malaysia. Atmos Res. 139:116–127. doi: 10.1016/j.atmosres.2014.01.003

Ahmat H, Yahaya AS, Ramli NA. 2015. PM10 analysis for three industrialized areas using extreme value. JSM. 44(2):175–185.

Ahmed SO, Mazloum R, Abou-Ali H. 2018. Spatiotemporal interpolation of air pollutants in the Greater Cairo and the Delta, Egypt. Environ Res. 160:27–34.

Al-Hemoud A, Al-Dousari A, Al-Shatti A, Alkhayat A, Behbehani W, Malek M. 2017. Health impact assessment associated with exposure to PM10 and dust storms in Kuwait. Atmosphere. 9.

Allen MP. 1997. The problem of multicollinearity. In: Understanding regression analysis. p. 176–180.

Althuwaynee OF, Balogun A-L, Al Madhoun W. 2020. Air pollution hazard assessment using decision tree algorithms and bivariate probability cluster polar function: evaluating inter-correlation clusters of PM10 and other air pollutants. GISci Remote Sens. 57(2):207–226.

Andaya BW, Andaya LY. 2016. A history of Malaysia. Macmillan International Higher Education.

APIMS. 2020. API table [hourly]. Department of Environment (DOE), Malaysia [accessed 2020 October 2]. http://apims.doe.gov.my/public_v2/api_table.html.

Azarmi F, Kumar P, Marsh D, Fuller G. 2016. Assessment of the long-term impacts of PM 10 and PM 2.5 particles from construction works on surrounding areas. Environ Sci Processes Impacts. 18:208–221.

Baker J. 2008. Crossroads: a popular history of Malaysia and Singapore. 2nd ed. Marshall Cavendish International Asia Pte Ltd.

Bowen B. 2018. Goodness of fit via residual plots in item response theory. University of South Carolina, Columbia

Bozdağ A, Dokuz Y, Gökçek ÖB. 2020. Spatial prediction of PM10 concentration using machine learning algorithms in Ankara, Turkey. Environ Pollut. 263:114635.

Chen F-W, Liu C-W. 2012. Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan. Paddy Water Environ. 10(3):209–222.

Choubin B, Abdolshahnejad M, Moradi E, Querol X, Mosavi A, Shamshirband S, Ghamisi P. 2020. Spatial hazard assessment of the PM10 using machine learning models in Barcelona, Spain. Sci Total Environ. 701:134474.

Chu J. 2015. Can rain clean the atmosphere? Massachusetts Institute of Technology (MIT). http://news.mit.edu/2015/rain-drops-attract-aerosols-clean-air-0828.

Cichowicz R, Wielgosiński G, Fetter W. 2020. Effect of wind speed on the level of particulate matter PM10 concentration in atmospheric air during winter season in vicinity of large combustion plant. J Atmos Chem. 77(1-2):35–48.

Cujia A, Agudelo-Castañeda D, Pacheco-Bustos C, Teixeira EC. 2019. Forecast of PM10 time-series data: a study case in Caribbean cities. Atmos Pollut Res. 10(6):2053–2062.

Czernicki B, Półrolniczak M, Kolendowicz L, Marosz M, Kendzierski S, Pilguj N. 2017. Influence of the atmospheric conditions on PM10 concentrations in Poznań, Poland. J Atmos Chem. 74(1):115–139.

Daoud JI. 2017. Multicollinearity and regression analysis. J Phys Conf Ser. 949:012009.

de Rooij MMT, Heederik DJJ, Borlée F, Hoek G, Wouters IM. 2017. Spatial and temporal variation in endotoxin and PM10 concentrations in ambient air in a livestock dense area. Environ Res. 153:161–170.
Department of Environment Malaysia, DOE. 1997. A guide to air pollutant in Malaysia (API). Kuala Lumpur, Malaysia: Ministry of Science, Technology and Environment.

Department of Environment. 2013. General information of air pollutant index. http://www.doe.gov.my/portalv1/wp-content/uploads/2013/06/General-Information-of-Air-Pollutant-Index.pdf.

Department of Environment, DOE. 2019. Information about API. Air Pollutant Index of Malaysia (APIMS). http://apims.doe.gov.my/public_v2/aboutapi.html.

Djalalova I, Delle Monache LD, Wilczak J. 2015. PM2.5 analog forecast and Kalman filter post-processing for the Community Multiscale Air Quality (CMAQ) model. Atmos Environ. 108:76–87.

Dotse S-Q, Dagar L, Petra MI, De Silva LC. 2016. Influence of Southeast Asian Haze episodes on high PM10 concentrations across Brunei Darussalam. Environ Pollut. 219:337–352.

Feng W, Li H, Wang S, Halm-Lutterodt NV, An J, Liu Y, Liu M, Wang X, Guo X. 2019. Short-term PM10 and emergency department admissions for selective cardiovascular and respiratory diseases in Beijing, China. Sci Total Environ. 657:213–221.

Folberth GA, Butler TM, Collins WJ, Rumbold ST. 2015. Megacities and climate change—A brief overview. Environ Pollut. 203:235–242.

Fong SY, Abdullah S, Ismail M. 2018. Forecasting of Particulate Matter (PM10) concentration based on gaseous pollutants and meteorological factors for different monsoons of urban coastal area in Terengganu. J Sustainability Sci Manage. 13:3–18.

Ganguly R, Sharma D, Kumar P. 2019. Trend analysis of observational PM10 concentrations in Shimla city, India. Sustainable Cities Soc. 51:101719.

Giri D, Adhikary PR, Murthy VK. 2008. The influence of meteorological conditions on PM10 concentrations in Kathmandu Valley. Int J Environ Res. 2:49–60.

Gupta R. 2019. Multiple linear regression using Python [accessed 2020 May 16]. https://www.c-sharpcorner.com/article/multiple-linear-regression/.

Gvozdić V, Kovač-Andrić E, Brana J. 2011. Influence of meteorological factors NO2, SO2, CO and PM10 on the concentration of O3 in the urban atmosphere of Eastern Croatia. Environ Model Assess. 16:491–501.

Han L, Zhao J, Gao Y, Gu Z, Xin K, Zhang J. 2020. Spatial distribution characteristics of PM2.5 and PM10 in Xi’an City predicted by land use regression models. Sustainable Cities Soc. 61:102329.

Harvey HB, Sotardi ST. 2018. The pareto principle. J Am Coll Radiol. 15(6):931.

Hassan H, Latif MT, Juneng L, Amil N, Khan MF, Yik DJ, Abdullah NA. 2020. Interaction of PM10 concentrations with local and synoptic meteorological conditions at different temporal scales. Atmos Res. 241:104975.

Hernandez G, Berry T-A, Wallis S, Poyner D. 2017. Temperature and humidity effects on particulate matter concentrations in a sub-tropical climate during winter.

How CY, Ling YE. 2016. The influence of PM2.5 and PM10 on Air Pollution Index (API). Environmental Engineering, Hydraulics and Hydrology: Proceeding of Civil Engineering, Universiti Teknologi Malaysia, Johor, Malaysia. 3:132.

Huang L, Chen M, Hu J. 2016. Twelve-year trends of PM10 and visibility in the Hefei metropolitan area of China. Adv Meteorol. 2016: 1121–1129.

Jumaah HJ, Ameen MH, Kalantar B, Rizeei HM, Jumaah SJ. 2019. Air quality index prediction using IDW geostatistical technique and OLS-based GIS technique in Kuala Lumpur, Malaysia. Geomatics Nat Hazards Risk. 10(1):2185–2199.

Karatzas K, Katsifarakis N, Orlowski C, Sarzyński A. 2018. Revisiting urban air quality forecasting: a regression approach. Vietnam J Comput Sci. 5(2):177–184.

Kassomenos PA, Vardoulakis S, Chaloulakou A, Paschalidou AK, Grivas G, Borge R, Lumbreras J. 2014. Study of PM10 and PM2.5 levels in three European cities: analysis of intra and inter urban variations. Atmos Environ. 87:153–163.

Khan MF, Latif MT, Juneng L, Amil N, Nadzir MSM, Hoque HMS. 2015. Physicochemical factors and sources of particulate matter at residential urban environment in Kuala Lumpur. J Air Waste Manag Assoc. 65:958–969.
Kim MJ. 2019. Changes in the relationship between particulate matter and surface temperature in Seoul from 2002–2017. Atmosphere. 10(5):238.

Kim S-Y, Yi S-J, Eum YS, Choi H-J, Shin H, Ryou HG, Kim H. 2014. Ordinary kriging approach to predicting long-term particulate matter concentrations in seven major Korean cities. Environ Anal Health Toxicol. 29:e2014012.

Kleine Deters J, Zalakeviciute R, Gonzalez M, Rybarczyk Y. 2017. Modeling PM2.5 urban pollution using machine learning and selected meteorological parameters. J Electr Comput Eng. 2017: 1–14. DOI: 10.1155/2017/5106045

Kumar A, Gupta I, Brandt J, Kumar R, Kumar Dikshit A, Patil RS. 2016. Air quality mapping using GIS and economic evaluation of health impact for Mumbai City, India. J Air Waste Manag Assoc. 66(5):470–481.

Kusumaningtyas SD, Aldrian E. 2016. Impact of the June 2013 Riau province Sumatera smoke haze event on regional air pollution. Environ Res Lett. 11(7):075007.

Latif MT, Azmi SZ, Noor ADM, Ismail AS, Johny Z, Idrus S, Mohamed AF, Mokhtar MB. 2011. The impact of urban growth on regional air quality surrounding the Langat River Basin, Malaysia. Environmental. 31(3):315–324.

Law M, Jackson D. 2017. Residual plots for linear regression models with censored outcome data: a refined method for visualizing residual uncertainty. Commun Stat - Simul. Comput. 46(4):3159–3171.

Lee J, Kim K-Y. 2018. Analysis of source regions and meteorological factors for the variability of spring PM10 concentrations in Seoul, Korea. Atmos Environ. 175:199–209.

Li Y, Chen Q, Zhao H, Wang L, Tao R. 2015. Variations in PM10, PM2.5 and PM1.0 in an urban area of the Sichuan Basin and their relation to meteorological factors. Atmosphere. 6(1):150–163.

Li J, Li X, Wang K. 2019. Atmospheric PM2.5 concentration prediction based on time series and interactive multiple model approach. Adv Meteorol. 2019: 1–11. doi: 10.1155/2019/1279565.

Liu X, Ma L, Li X, Ai B, Li S, He Z. 2014. Simulating urban growth by integrating landscape expansion index (LEI) and cellular automata. Int J Geog Inf Sci. 28(1):148–163.

Luan T, Guo X, Zhang T, Guo L. 2019. Below-cloud aerosol scavenging by different-intensity rains in Beijing city. J Meteorol Res. 33(1):126–137.

Ma J, Ding Y, Cheng JCP, Jiang F, Wan Z. 2019. A temporal-spatial interpolation and extrapolation method based on geographic Long Short-Term Memory neural network for PM2.5. J Cleaner Prod. 237:117729.

Manisalidis I, Stavropoulos E, Stavropoulos A and Bezirtzoglou E. 2020. Environmental and health impacts of air pollution: A review. Frontiers in public health. 8.

Najid SK, Zaharim A, Razali AM, Zainol MS, Ibrahim K, and Sopian K. 2009. Analyzing the east coast Malaysia wind speed data. Inter J of Ener and Envir. 3: 53–60.

Nasif A, Mohammed NI, Malakahmad A, Abualqumboz MS. 2019. Multivariate analysis of monsoon seasonal variation and prediction of particulate matter episode using regression and hybrid models. International Journal of Environmental Science and Technology. 16: 2587–600.

Ng K, Awang N. 2018. Multiple linear regression and regression with time series error models in forecasting PM10 concentrations in Peninsular Malaysia. Environ Monit Assess. 190(2):63.

Osseiran N, Chriscarden K. 2016. Air pollution levels rising in many of the world’s poorest cities. Geneva: WHO.

Özdemir U, Taner S. 2014. Impacts of meteorological factors on PM10: Artificial neural networks (ANN) and multiple linear regression (MLR) approaches. Envirowen 15: 329–36.

Pour SH, Wahab AKA, Shahid S. 2020. Physical-empirical models for prediction of seasonal rainfall extremes of Peninsular Malaysia. Atmos Res. 233:104720.

Qureshi MI, Rasli AM, Awan U, Ma J, Ali G, Faridullah, Alam A, Sajjad F, Zaman K. 2015. Environment and air pollution: health services bequeath to grotesque menace. Environ Sci Pollut Res. 22:3467–3476.
Rahman A, Luo C, Khan MHR, Ke J, Thilakanayaka V, Kumar S. 2019. Influence of atmospheric PM2.5, PM10, O3, CO, NO2, SO2, and meteorological factors on the concentration of airborne pollen in Guangzhou, China. Atmos Environ. 212:290–304.

Rani NLA, Azid A, Khalit SI, Juahir H, Samsudin MS. 2018. Air pollution index trend analysis in Malaysia, 2010–15. Pol J Environ Stud. 27: 801–807.

Shaadan N, Jemain AA, Latif MT, Deni SM. 2015. Anomaly detection and assessment of PM10 functional data at several locations in the Klang Valley, Malaysia. Atmos Pollut Res. 6(2):365–375.

Sharma RC, Sharma N. 2016. Influence of some meteorological variables on PM10 and NOx in Gurgaon, Northern India. Am J Environ Prot. 4:1–6.

Shukla K, Kumar P, Mann GS, Khare M. 2020. Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi. Sustainable Cities Soc. 54:101997.

Song Y, Qin S, Qu J, Liu F. 2015. The forecasting research of early warning systems for atmospheric pollutants: a case in Yangtze River Delta region. Atmos Environ. 118:58–69.

Stoimenova M, Voynikova D, Ivanov A, Gocheva-Ilieva S, Iliev I. 2017. Regression trees modeling and forecasting of PM10 air pollution in urban areas. AIP Conference Proceedings, p. 030005. AIP Publishing LLC.

Sulong NA, Latif MT, Khan MF, Amil N, Ashfold MJ, Wahab MIA, Chan KM, Sahani M. 2017. Source apportionment and health risk assessment among specific age groups during haze and non-haze episodes in Kuala Lumpur, Malaysia. Sci Total Environ. 601–602:556–570.

Sun XW, Chen PL, Ren L, Lin YN, Zhou JP, Ni L, Li QY. 2018. The cumulative effect of air pollutants on the acute exacerbation of COPD in Shanghai, China. Sci Total Environ. 622–623:875–881.

Taheri Shahraiyi H, Sodoudi S. 2016. Statistical modeling approaches for PM10 prediction in urban areas; A review of 21st-century studies. Atmosphere. 7:15.

Tang KHD. 2019. Climate change in Malaysia: trends, contributors, impacts, mitigation and adaptations. Sci Total Environ. 650:1858–1871.

Tian B. 2016. GIS Technology Applications In: Environmental and Earth Sciences (1st ed.). CRC Press. doi: 10.1201/9781315366975. Edition: 1st Edition ; First Published: 2016 eBook Published: 29 September 2016 Pub. Location: Boca Raton Imprint: CRC Press DOI: 10.1201/9781315366975 Pages: 278

United Nations. 2015. World population prospects: the 2015 revision. United Nations Econ Soc Aff. 33:1–66.

Wang J, Ogawa S. 2015. Effects of meteorological conditions on PM2.5 concentrations in Nagasaki, Japan. Int J Environ Res Public Health. 12(8):9089–9101.

Wang X, Wang B. 2019. Research on prediction of environmental aerosol and PM2.5 based on artificial neural network. Neural Comput Appl. 31(12):8217–8227.

Wen Y, Nor AFbM, Sulaiman Zb. 2016. Transboundary air pollution in Malaysia: impact and perspective on haze. Nova J Eng Appl Sci. 5: 1–11, ISSN: 2292-7921, doi: 10.20286/nova-jes-050103.

Wong LP, Alias H, Aghamohammadi N, Ghadimi A, Sulaiman NMM. 2017. Control measures and health effects of air pollution: a survey among public transportation commuters in Malaysia. Sustainability. 9(9):1616.

World Health Organization. 2006. Health risks of particulate matter from long-range transboundary air pollution. Copenhagen: WHO Regional Office for Europe.

World Health Organization. 2016a. Air pollution levels rising in many of the world’s poorest cities. World Health Organization [accessed 2020 April 14]. https://www.who.int/en/newsroom/detail/12-05-2016-air-pollution-levels-rising-in-many-of-the-world-s-poorest-cities.

World Health Organization. 2016b. Ambient air pollution: a global assessment of exposure and burden of disease.

Yadav R, Sahu LK, Jaffrey SNA, Beig G. 2014. Temporal variation of Particulate Matter (PM) and potential sources at an urban site of Udaipur in Western India. Aerosol Air Qual Res. 14(6):1613–1629.
Yao L, Lu N, Yue X, Du J, Yang C. 2015. Comparison of hourly PM2.5 observations between urban and suburban areas in Beijing, China. IJERPH. 12(10):12264–12276.

Yoo W, Mayberry R, Bae S, Singh K, He QP, Lillard JW, Jr. 2014. A study of effects of multicollinearity in the multivariable analysis. Int J Appl Sci Technol. 4(5):9–19.

Zakaria UA, Saudi ASM, Abu IF, Azid A, Balakrishnan A, Amin NA, Rizman ZI. 2017. The assessment of ambient air pollution pattern in Shah Alam, Selangor, Malaysia. J Fundam Appl Sci. 9:772–788.

Zaman K, Nurul KZ, Kanniah DK, Mohamed N, Pahrol MA. 2019. Assessment of population exposure to PM2.5 and long-term mortality in Malaysia.

Zhang J, Ding W. 2017. Prediction of air pollutants concentration based on an extreme learning machine: the case of Hong Kong. Int J Environ Res Public Health. 14(114): 1–19.

Zhang F, Cheng H-r, Wang Z-w, Lv X-p, Zhu Z-m, Zhang G, Wang X-m. 2014. Fine particles (PM2.5) at a CAWNET background site in Central China: chemical compositions, seasonal variations and regional pollution events. Atmos Environ. 86:193–202.

Zhang H, Wang Y, Hu J, Ying Q, Hu X-M. 2015. Relationships between meteorological parameters and criteria air pollutants in three megacities in China. Environ Res. 140: 242–254.

Zheng S, Pozzer A, Cao CX, Lelieveld J. 2015. Long-term (2001–2012) concentrations of fine particulate matter (PM2.5) and the impact on human health in Beijing, China. Atmos Chem Phys. 15: 5715–25.

Zhou Y, Chang F-J, Chang L-C, Kao IF, Wang Y-S, Kang C-C. 2019. Multi-output support vector machine for regional multi-step-ahead PM2.5 forecasting. Sci Total Environ. 651: 230–240.