Temporal Assessment on Variation of PM$_{10}$ Concentration in Kota Kinabalu using Principal Component Analysis and Fourier Analysis

MUHAMMAD IZZUDDIN RUMALING$^1$, FUEI PIEN CHEE$^{1,*}$, JEDOL DAYOU$^1$, JACKSON HIAN WUI CHANG$^{3,4}$, STEVEN SOON KAI KONG$^4$ and JUSTIN SENTIAN$^2$

$^1$Faculty of Science and Natural Resources (FSNR), Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia.

$^2$Climate Change Research Group (CCRG), FSNR, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia.

$^3$Preparatory Centre for Science and Technology, Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia.

$^4$Cloud and Aerosol Laboratory, Department of Atmospheric Science, National Central University, Taoyuan, Taiwan (ROC).

Abstract

Introduction: PM$_{10}$ (particulate matter with aerodynamic diameter below 10 microns) has always caught scientific attention due to its effect to human health. Predicting PM$_{10}$ concentration is essential for early preventive measures, especially for cities such as Kota Kinabalu. Temporal data clustering may enhance accuracy of prediction model by group data in time range. However, the necessity of temporal data clustering has yet to be studied in Kota Kinabalu.

Objective: This research is conducted to compare significance of meteorological and pollutant factors for PM$_{10}$ variation in clustered and unclustered data.

Methodology: This study is focused in Kota Kinabalu, Sabah. The data for meteorological factors (Ws, Wd, Hum, Temp) and pollutant factors (CO$_2$, NO$_2$, O$_3$, SO$_2$, PM$_{10}$) from 2003 to 2012 provided by Department of Environment are used for this research. Missing data are imputed using nearest neighbour method before it is clustered by monsoonal clustering. Unclustered and clustered datasets are analysed using principal

Article History

Received: 26 August 2019
Accepted: 9 October 2019

Keywords

Fourier Analysis; Particulate Matter; Principal Component Analysis; Temporal Clustering; 2-D Regression Analysis.

CONTACT
Fuei Pien Chee  fpchee06@ums.edu.my  Faculty of Science and Natural Resources (FSNR), Universiti Malaysia Sabah, Kota Kinabalu, Sabah, Malaysia.

© 2019 The Author(s). Published by Enviro Research Publishers.
This is an Open Access article licensed under a Creative Commons license: Attribution 4.0 International (CC-BY).
Doi: http://dx.doi.org/10.12944/CWE.14.3.08
component analysis (PCA) to check significance of factors contributing to PM$_{10}$ concentration.

**Findings:** PCA results show that temporal clustering does not have noticeable effect on the variation of PM$_{10}$ concentration. For all datasets, humidity and x-component wind speed have highest factor loading on PC$_1$ and PC$_2$ respectively. Further statistical analysis by 2-D regression shows that humidity ($\rho = -0.60 \pm 0.20$) and temperature ($\rho = 0.63 \pm 0.11$) have moderate to strong correlation towards PM$_{10}$ concentration. This may be due to high humidity level and strong negative correlation between temperature and humidity ($\rho = -0.91 \pm 0.03$). In contrast, both x- and y-component wind speed generally show weak correlation towards PM$_{10}$, with $\rho$ value of $0.09 \pm 0.14$ and $0.24 \pm 0.18$ respectively probably because of varying direction of particle dispersion. Fourier analysis further confirms this result by showing that human activity contributes major effect to variation of PM$_{10}$ concentration.

**Introduction**

Particulate matter is an ambient respiratory particle suspended in the air. It has been attracting scientific attention because of its effects on human health (Shahraiyni & Sodoudi, 2016). PM$_{10}$ (particulate matter with size less than 10 $\mu$m) is considered to be one of major air pollutants. One of major sources of PM$_{10}$ is from biomass burning, brought by haze. This has becoming a typical and reoccurring challenge in Southeast Asia since 1980s (Shaadan et al., 2015). Other than that, motor vehicle usage and industrial activities also contribute to emission of PM$_{10}$. PM$_{10}$ is more hazardous to human health compared to other pollutants such as carbon monoxide and ground-level ozone (Kim et al., 2015; Ny & Lee, 2010). One study suggested that PM$_{10}$ increases the risk of children (between 5 to 15 years old) with asthma to visit emergency department (Cadelis et al., 2014). Also, short to medium-term exposure to PM10 among infants leads to risk of respiratory syncytial virus (RSV) bronchiolitis (Carugno et al., 2018).

Located in Malaysia, Kota Kinabalu experiences monsoon season twice a year: northeast monsoon (NEM) and southwest monsoon (SWM). NEM originates from China and northern Pacific, while SWM originates from deserts in Australia. NEM occurs between November to March and is characterized by cool temperature (Ho et al., 2013). Meanwhile, SWM occurs between May to September as climate becomes warmer. According to Teong et al., (2017), SWM recorded greater amount of rainfall compared to NEM. Outside monsoon season is known as inter-monsoon, happening on April and October (Naing et al., 2011). Inter-monsoon during April is characterized by hot and dry climate, while inter-monsoon in October is cooler and has higher rainfall (Teong et al., 2017).

Short-term prediction of PM$_{10}$ concentration is essential as it enables community to reduce health risks by early preventive measures. Meanwhile, long-term forecasting model is crucial for urban planning, transportation networks, industrial and residential area management in such a way that minimizes health risks to public (Shahraiyni & Sodoudi, 2016). This is important for developing city such as Kota Kinabalu. Prediction model may become more accurate when seasonal factor such as monsoon is considered. This is because climate may affect PM10 concentration differently, depending on season. However, it has not been studied yet whether monsoonal clustering may show difference in PM10 variation with respect to meteorological and pollutant factors. Hence, this study aims to find out whether monsoonal clustering for Kota Kinabalu, Sabah shows difference with unclustered data.

**Data and Methods**

**Study Area and Data**

This research focuses on Kota Kinabalu (5.98°N, 116.07°E, altitude: 13 m), capital city of Sabah, Malaysia. As shown in Figure 1, Kota Kinabalu...
is located at west coast of Sabah. The climate in Kota Kinabalu is characterized by hot and humid climate, influenced by circulation of monsoons (Ho et al., 2013). Kota Kinabalu experiences uniformly high and invariant temperature and high humidity. Heavy rainfall recorded usually occurs in the afternoon (Djamila et al., 2011). In terms of economic activity, Kota Kinabalu is the gateway to the islands of Borneo. Consequently, activities such as trade, industry and tourism are concentrated in Kota Kinabalu (Noor et al., 2014).

![Fig.1: Location of Kota Kinabalu](image)

Air quality monitoring station, located in SMK Putatan, Kota Kinabalu is operated by Alam Sekitar Sdn. Bhd. (ASMA), a private company working under Department of Environment Malaysia (DOE). PM\textsubscript{10} concentration in μg m\textsuperscript{-3} is measured in this station using tapered element oscillating microbalance (TOEM), with temporal resolution of 1 h. Along with other meteorological data Ws (wind speed), Wd (wind direction), Hum (relative humidity), Temp (ambient temperature) and air pollutant concentration data for CO (carbon monoxide), NO\textsubscript{2} (nitrogen dioxide), O\textsubscript{3} (ozone), SO\textsubscript{2} (sulfur dioxide), PM\textsubscript{10} concentration data is made available by Air Quality Division under DOE. For this research, data from 2003 to 2012 is studied. Statistics of PM\textsubscript{10} concentration data collected in CA0030 from 2003 to 2012 is summarized in Table 1.

**Table 1: Summary of statistics of PM\textsubscript{10} concentration data from 2003 to 2012**

<table>
<thead>
<tr>
<th>Statistics</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>35.73</td>
</tr>
<tr>
<td>Standard Deviation</td>
<td>19.16</td>
</tr>
<tr>
<td>Minimum</td>
<td>5.00</td>
</tr>
<tr>
<td>Maximum</td>
<td>495.00</td>
</tr>
<tr>
<td>Median</td>
<td>32.00</td>
</tr>
<tr>
<td>1st quartile</td>
<td>23.00</td>
</tr>
<tr>
<td>3rd quartile</td>
<td>44.00</td>
</tr>
<tr>
<td>Coefficient of Variation (%)</td>
<td>53.63</td>
</tr>
</tbody>
</table>
As Wd is an angular quantity, it is not usually used for principal component analysis or regression analysis (Gvozdic et al., 2011). Furthermore, higher magnitude of Wd does not reflect higher intensity of that property, which may present inaccuracy in both principal component analysis and regression analysis. Thus, Ws and Wd are transformed into east-west component (Wx) and north-south component (Wy) of wind speed using equations (1) and (2). Positive values indicate wind blowing eastward and northward respectively (Gvozdic et al., 2011).

\[ W_x = W_s \sin W_d \]  
\[ W_y = W_s \cos W_d \]

Nearest Neighbour Method
Some data from monitoring stations are missing either due to power outage, calibration, or relocation of monitoring station. Due to missingness of data, it must be imputed before it is used for this research. Data imputation is crucial for further analysis that requires complete dataset such as principal component analysis and Fourier analysis. One of the most common applicable method known as nearest neighbour method (Dominick et al., 2012; Li & Liu, 2014) is used in this paper. Given starting point of a gap \((x_1, y_1)\) and end point of the gap \((x_2, y_2)\), the interpolant \(y\) at corresponding time \(x\) can be calculated using equation (3) (Zakaria & Noor, 2018).

\[ y = \begin{cases} 
  y_1, & \text{when } x < \frac{x_1 + x_2}{2} \\
  y_2, & \text{when } x \geq \frac{x_1 + x_2}{2}
\end{cases} \]

Clustering and Unclustered Data
Data is clustered when data is grouped into a set with same characteristics. In contrast, data is unclustered when they are ungrouped. For this research, two sets of temporal clusters are tested and compared with unclustered data. Due to geographic location of Kota Kinabalu, this city experiences monsoonal season. Thus, monsoonal clustering is studied. In this case, data is divided into four groups: NEM (November to March), IM\(^4\) (April), SWM (May to September), and IM\(^10\) (October). Data is grouped based on the month the data is observed. The number of samples for all clusters are summarized in Table 2.

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Number of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unclustered data (total)</td>
<td>87672</td>
</tr>
<tr>
<td>Monsoonal cluster:</td>
<td></td>
</tr>
<tr>
<td>NEM</td>
<td>36312</td>
</tr>
<tr>
<td>IM(^4)</td>
<td>7200</td>
</tr>
<tr>
<td>SWM</td>
<td>36720</td>
</tr>
<tr>
<td>IM(^10)</td>
<td>7440</td>
</tr>
</tbody>
</table>

Principal Component Analysis (PCA)
PCA is a method used to identify the most significant variable that contributes to variation (Dominick et al., 2012). This method is usually used to transform a set of variables to \(n\) new variables, known as principal component (PC). PCs account for more significant variation compared to observed variable \(X\). Given that \(l\) is loading factor, principal component of \(i\)th component is given by equation (4) (Gvozdic et al., 2011).

\[ PC_i = l_1 X_1 + l_2 X_2 + \ldots + l_n X_n \]

Explain variation is the total variation of data included in principal component. Usually, explained variation is highest in \(PC_1\), followed by \(PC_2\),
and so on. Gvozdic et al., (2011) and Dominick et al., (2012) had demonstrated that the first two principal components $PC_1$ and $PC_2$ are used to create factor loading plot. This is because both PCs usually account at least 90% of variation of data, as evaluated in cumulative explained variation.

**Fourier Analysis**

Fourier analysis involves transforming data in time domain into frequency domain using fast Fourier transform (FFT) (Kim & Son, 2011). The results obtained from Fourier analysis gives important insights on time series data (Gvozdic et al., 2011). Output from frequency domain displays cyclical behaviour of time series data as seasonal or diurnal data. Due to the nature of meteorological and air pollutant data as time series data, Fourier analysis is a suitable method in analysing the data (Gvozdic et al., 2011). A function of time series data $y(t)$ is expressed in the form of Fourier series in equation (5), where $\bar{y}$ is arithmetic mean of data $y(t)$, $T$ is measurement period, $A_k$ and $B_k$ are Fourier coefficients and $k$ is a harmonic number.

$$y(t) = \bar{y} + \sum_{k=1}^{\infty} \left[ A_k \cos \left( \frac{2\pi k t}{T} \right) + B_k \sin \left( \frac{2\pi k t}{T} \right) \right]$$  \hspace{1cm} ...(5)

Data transformed into frequency domain is plotted into Fourier spectrum, containing amplitudes $C_k$ (as shown in equation (6)) as a function of frequency or period (Gvozdic et al., 2011).

$$C_k = \sqrt{A_k^2 + B_k^2}$$  \hspace{1cm} ...(6)

Due to ability of Fourier analysis to show cyclic behaviour of time series data, this is used to verify whether variation of time series data is mainly natural or human-induced (Choi et al., 2008). It is assumed that higher amplitudes in seasonal and yearly cycle corresponds to the major effect of natural cycle while higher amplitudes in daily and weekly cycle corresponds to the major effect of anthropogenic activities (Choi et al., 2008).

**Results**

**Unclustered Data**

Unclustered data is analysed by inputting all data into PCA. All PCs are tabulated in Table 3, along with explained variation (EV) and cumulative explained variation (CEV). The variables with highest loading factor contribute the most in variation of corresponding principal components, and their values are highlighted in Table 3.

<table>
<thead>
<tr>
<th>PC</th>
<th>PC$_1$</th>
<th>PC$_2$</th>
<th>PC$_3$</th>
<th>PC$_4$</th>
<th>PC$_5$</th>
<th>PC$_6$</th>
<th>PC$_7$</th>
<th>PC$_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wx</td>
<td>-0.1115</td>
<td>0.9503</td>
<td>-0.2540</td>
<td>-0.1411</td>
<td>0.0031</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Wy</td>
<td>0.1668</td>
<td>0.2782</td>
<td>0.9450</td>
<td>0.0404</td>
<td>-0.0122</td>
<td>0.0005</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Hum</td>
<td>0.9433</td>
<td>0.0924</td>
<td>-0.2042</td>
<td>0.2449</td>
<td>0.0010</td>
<td>0.0002</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Temp</td>
<td>-0.2644</td>
<td>0.1045</td>
<td>-0.0249</td>
<td>0.9583</td>
<td>0.0132</td>
<td>-0.0012</td>
<td>-0.0003</td>
<td>-</td>
</tr>
<tr>
<td>CO</td>
<td>0.0050</td>
<td>-0.0010</td>
<td>0.0129</td>
<td>-0.0120</td>
<td>0.9998</td>
<td>0.0061</td>
<td>-0.0077</td>
<td>-0.0006</td>
</tr>
<tr>
<td>NO$_2$</td>
<td>-</td>
<td>-</td>
<td>0.0002</td>
<td>-</td>
<td>0.0083</td>
<td>-0.1009</td>
<td>0.9939</td>
<td>-0.0437</td>
</tr>
<tr>
<td>O$_3$</td>
<td>-0.0006</td>
<td>-</td>
<td>-</td>
<td>0.0011</td>
<td>-0.0053</td>
<td>0.9948</td>
<td>0.1013</td>
<td>0.0073</td>
</tr>
<tr>
<td>SO$_2$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0010</td>
<td>-0.0116</td>
<td>0.0427</td>
<td>0.9990</td>
</tr>
</tbody>
</table>

| EV (%) | 83.12 | 9.15 | 6.27 | 1.42 | 0.04 | - | - | - |
| CEV (%)| 83.12 | 92.27| 98.54| 99.96| 100.00| 100.00| 100.00| 100.00|

Note: Values smaller than 0.0001 are not shown

The explained variation is largest in first principal component and gets smaller in successive principal components. The first two principal components are considered as they take account into more than 90% of total variation of data. $PC_1$ and $PC_2$ are plotted in factor loading as shown in Figure 2. The most
significant variable that contributes to variation in PC₁ is humidity, while x-component wind speed is the most significant in PC₂. This shows that 92.27% of data variation is due to both humidity and wind speed. While other variables are significant in other principal components, they are not as significant because these principal components consider only less than 10% of variation.

PC₁ accounts for 83.12% of variation of dataset. At PC₁, humidity has the highest loading factor with the value of 0.9433, followed by temperature (-0.2644) and y-component wind speed (0.1668). As for PC₂ (EV = 9.15%), x-component wind speed has the highest loading factor with the value of 0.9503. Because PC₁ has significantly higher EV compared to PC₂, humidity shows to have the highest significance in variation. Meanwhile, pollutant factors (CO, NO₂, O₃, SO₂) are very close to origin of factor loading plot, which reflects insignificance in variation.

Clustered Data
All monsoonal clusters (NEM, IM₄, SWM, IM₁₀) are analysed using PCA, and the results are tabulated in Table 4. As cumulative explained variable for first two principal components are 92.19 ± 0.09 %, only PC₁ and PC₂ are shown in Table 4. Factor loading for all clusters are plotted as shown in Figure 3.

Table 4: Principal Component Analysis (PCA) for 4 clusters divided based on monsoonal season

<table>
<thead>
<tr>
<th>Cluster</th>
<th>NEM</th>
<th>IM₄</th>
<th>SWM</th>
<th>IM₁₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>PC</td>
<td>PC₁</td>
<td>PC₂</td>
<td>PC₁</td>
<td>PC₂</td>
</tr>
<tr>
<td>Wx</td>
<td>-0.1687</td>
<td>0.9124</td>
<td>-0.1475</td>
<td>0.9678</td>
</tr>
<tr>
<td>Wy</td>
<td>0.1499</td>
<td>0.3918</td>
<td>0.2012</td>
<td>0.2109</td>
</tr>
<tr>
<td>Hum</td>
<td>0.9386</td>
<td>0.1120</td>
<td>0.9303</td>
<td>0.1247</td>
</tr>
<tr>
<td>Temp</td>
<td>-0.2611</td>
<td>0.0383</td>
<td>-0.2688</td>
<td>0.0585</td>
</tr>
<tr>
<td>CO</td>
<td>0.0047</td>
<td>-0.0004</td>
<td>0.0058</td>
<td>0.0013</td>
</tr>
<tr>
<td>NO₂</td>
<td>-0.0006</td>
<td>-0.0007</td>
<td>-0.0007</td>
<td>-0.0006</td>
</tr>
<tr>
<td>O₃</td>
<td>-0.0007</td>
<td>-0.0007</td>
<td>-0.0007</td>
<td>-0.0006</td>
</tr>
<tr>
<td>SO₂</td>
<td>-0.0007</td>
<td>-0.0007</td>
<td>-0.0007</td>
<td>-0.0006</td>
</tr>
<tr>
<td>EV (%)</td>
<td>83.96</td>
<td>8.35</td>
<td>85.36</td>
<td>6.84</td>
</tr>
<tr>
<td>CEV (%)</td>
<td>83.96</td>
<td>92.31</td>
<td>85.36</td>
<td>92.20</td>
</tr>
</tbody>
</table>

Note: Values smaller than 0.0001 are not shown
Based on Table 3 and 4, all PC<sub>1</sub> and PC<sub>2</sub> have similar most significant variables. The values are consistent, with small discrepancies. According to EV, PC<sub>1</sub> accounts for 83.35 ± 1.63% of total variation of data, with humidity having highest factor loading (0.9389 ± 0.0064), followed by ambient temperature (-0.2705 ± 0.0120) and y-component wind speed (0.1705 ± 0.0242). As for PC<sub>2</sub> with EV of 8.84 ± 1.58%, x-component wind speed has shown to have highest factor loading (0.9479 ± 0.0248). Similar to unclustered data, other variables are not as significant because PC<sub>3</sub> and above consider less than 10% of data variation for all clusters.

**Regression Analysis**

2-D regression is plotted in Figure 4 to further analyse the variation of each parameter in response to PM<sub>10</sub> concentration for every cluster. Pollutant factors (CO, NO<sub>2</sub>, O<sub>3</sub>, SO<sub>2</sub>) are not included because their variation is not significant as reflected by factor loading plot in Figure 2 and 3. According to 2-D regression in Figure 4, humidity and temperature generally show moderate to strong correlation with PM<sub>10</sub> with Pearson coefficient of correlation, ρ of -0.60 ± 0.20 and 0.63 ± 0.11 respectively. However, this is not the case for humidity in IM4 cluster with the value of ρ = -0.37. Furthermore, humidity and temperature show strong negative correlation with each other (ρ = -0.91 ± 0.03) for all clusters.

As for wind speed, both x- and y-components generally show weak to almost no correlation towards PM10 with values of ρ = 0.09 ± 0.14 and ρ = 0.24 ± 0.18 respectively. In addition,
both components also exhibit weak correlation towards each other (except during SWM), humidity (except during NEM and IM$_{10}$) and ambient temperature.

Fourier Analysis
Fourier spectrum in Figure 5 shows three major peaks corresponding to three major cycles that affect variation of PM$_{10}$ concentration. Seasonal cycle is the strongest for 6-month cycle. This is attributed to monsoonal cycle. However, seasonal cycle is not as strong compared to daily and half-daily cycle. Other cycles such as weekly and monthly cycle do not show strong effect on variation of PM$_{10}$, as reflected by Fourier spectrum.

Discussion
Based on PCA and factor loading, significant variables contributing to PM$_{10}$ variation for monsoonal cluster does not have much difference when comparing with unclustered dataset. This is reflected by small discrepancies between factor loadings in every cluster as well as unclustered data. As clustered data shows almost similar PCA result with unclustered data, monsoonal clustering is not necessary when it comes to prediction of PM$_{10}$ concentration in Kota Kinabalu. This is further supported by Fourier spectrum as shown in Figure 5. Based on assumption by Choi et al. (2008), PM$_{10}$ concentration cycle is mainly affected by human activities. Daily activities such as traffic congestion, industries and night markets emit PM$_{10}$ particles (Chang et al., 2018), consequently contributing
to temporal variation of PM$_{10}$ concentration. High amplitude of daily and half-daily cycle implies that the effect of human-induced activities is dominant over seasonal phenomenon (Choi et al., 2008).

Statistical analysis shows that PM$_{10}$ concentration in Kota Kinabalu decreases as humidity increases. In addition, PM$_{10}$ concentration shows weak negative correlation with humidity as Pearson coefficient of correlation is -0.60 ± 0.20. According to raw data, humidity in Kota Kinabalu almost always exceeds 40%. Also, monthly average humidity in Kota Kinabalu is always above 77% based on Figure 4. At high humidity level, water in air almost reaches saturation. As a result, water becomes more easily condensed and rainfall becomes easier to occur (Lou et al., 2017). This causes washout effect and wet deposition by rain on PM$_{10}$. At high humidity level, particulate matter absorbs water molecule and becomes heavy enough to be deposited (Munir et al., 2017). At the beginning of rainfall, small raindrop amount and dust emission from the raindrops may raise PM$_{10}$ concentration until certain extent (Lou et al., 2017). This may result in weak correlation between PM$_{10}$ concentration and humidity.

Moderate to strong correlation is also exhibited between PM$_{10}$ concentration and ambient temperature. This is reflected by Pearson coefficient of correlation of 0.63 ± 0.11. This may be due to atmospheric instability and enhanced kinetic motion of particles (Munir et al., 2017). High ambient temperature allows resuspension of particles into air. Furthermore, ambient temperature is strongly and negatively correlated towards humidity (ρ = -0.91 ± 0.03). The decrease in humidity inhibits deposition of particles, thus further increases PM$_{10}$ concentration in the air.

Both x- and y-components of wind speed are weakly correlated towards PM$_{10}$ concentration. This is observed by values of ρ = 0.09 ± 0.14 for eastern component and ρ = 0.24 ± 0.18 for northern component. It is possible that wind speed facilitates the dispersion of the particles (Munir et al., 2017). However, the dispersion is not fixed and can be both towards and away from Kota Kinabalu, which contributes to weak correlation between wind speed and PM$_{10}$ concentration.

**Conclusion**

In Kota Kinabalu, clustering method is not necessary when prediction of PM$_{10}$ concentration is made. This is because monsoonal clustering shows small difference with unclustered data. Fourier analysis further confirms this result by showing that seasonal cycle of PM$_{10}$ concentration is not as strong as daily and half-daily cycle. According to regression analysis, humidity and temperature both affect PM$_{10}$ concentration based on respective values of ρ = -0.60 ± 0.20 and ρ = 0.63 ± 0.11. This may
be due to high humidity level and strong negative correlation towards each other. On the other hand, components of wind speed show weak correlation towards PM$_{10}$, probably due to random dispersion of particles towards and away from Kota Kinabalu.

Results from this research suggests that prediction model for PM$_{10}$ concentration in Kota Kinabalu does not require temporal clustering. However, future research requires study to investigate whether this applies for other parts of Sabah. Also, other types of clustering such as humidity level have yet to be studied up until now.

Acknowledgement
The authors would like to thank Universiti Malaysia Sabah for supporting this research by providing grant (SBK0324-2018, SGI0054-2018 and GUG378-2019) and Department of Environment Malaysia for providing meteorological and pollutant data for research purpose.

Funding
The authors received financial support from Universiti Malaysia Sabah via research grants (SBK0324-2018, SGI0054-2018, GUG0378-2019) for research, authorship, and publication of this article.

Conflict of Interest
The authors do not have any conflict of interest.

References


