Assessment of 24-hour moving average PM2.5 concentrations in Bangkok, Thailand against WHO guidelines

Sirapong Sooktawee
: Department of Environmental Quality Promotion
Suwimon Kanchanasuta (suwimon.kan@mahidol.ac.th)
Mahidol University Faculty of Public Health https://orcid.org/0000-0002-6824-6133
Natthaya Bunplod
Mahidol University Faculty of Public Health

Research Article

Keywords: Bangkok, PM2.5, Euclidean distance, Cross-correlation, Similarity

Posted Date: September 30th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-2063119/v1

Version of Record: A version of this preprint was published at Sustainable Environment Research on January 24th, 2023. See the published version at https://doi.org/10.1186/s42834-023-00165-y.

Abstract

Currently, the particulate matter with diameter less than 2.5 micron ($PM_{2.5}$) pollutant has gained more concerned as can be seen from the WHO revised the air quality guideline value. The 24-hour average concentration has been strengthened from 25 μg m^{-3} to 15 μg m^{-3}. However, the continuous $PM_{2.5}$ monitoring system provides data on an hourly basis, which can be averaged at a 24-hour value compare with the WHO air quality guidelines. The value given by the moving average technique can be stored at the leftmost, center or rightmost hour. Three moving average $PM_{2.5}$ time series would differ from the hourly observed $PM_{2.5}$ data. Similarity testing by cross-correlation and Euclidean distance was performed to present a suitable 24-hour moving average time series for hourly data. The 24-hour moving average time series recorded at center is more suitable than the leftmost and rightmost 24-hour moving average time series in terms of shape and distance. It has less time lag and distance to the hourly $PM_{2.5}$ time series. Comparing the 24-hour moving average time series to the WHO interim targets and the guideline value reveals $PM_{2.5}$ concentration level lower than the guideline value (15 μg m^{-3}) about 40% during the nighttime, whereas the proportion during daytime is around 28%. Also, the NAAQS of Thailand for 24-hour $PM_{2.5}$ was changed from 50 μg m^{-3} to 37.5 μg m^{-3} corresponding to the interim targets 3 and 2, respectively. From this study, concentrations higher than the NAAQs level will increase from 10 to about 22%. The increase in the number of exceedances based on the same data means the state of air quality is similar. Therefore, residents may misunderstand and know the air quality becomes more severe. The government should spend more effort to reduce emissions and ambient air concentrations than earlier endeavors.

1 Introduction

Air pollution problems, especially particulate matter less than 2.5 micron ($PM_{2.5}$) issues, have been experienced and remain a concerned in many countries such as Australia, Brazil, China, India, Iran, Japan, Portugal, Thailand, and the United States [1–10] The $PM_{2.5}$ situation is important for sustainable development. Its annual mean is considered an indicator for target 11.6, reducing the adverse per capita environmental impact of cities by 2030, which is a part of the United Nations Sustainable Development Goals (SDGs) [11]. The $PM_{2.5}$ problem arising in many areas has inevitably affected the health of humanity. The World Health Organization (WHO) strengthened air quality guidelines in 2021 to challenge the global community in enhancing air quality and reducing health burden. The guideline values suggested long term and short term exposure in 2006 were revised in 2021. The $PM_{2.5}$ guideline values changed from 25 μg m^{-3} to 15 μg m^{-3} and 10 μg m^{-3} to 5 μg m^{-3} for 24-hour mean and annual mean, respectively [12–13].

According to the air quality guidelines suggested by the WHO for 24-hour mean concentration of $PM_{2.5}$, the ambient $PM_{2.5}$ status of 45 megacities in the world were revealed by comparing between the measurement data and 24-hour air quality guideline (AQG) value. None of the 45 megacities failed to meet the former daily air quality guideline value (25 μg m^{-3}) [14]. A related study in rural South India reported that high $PM_{2.5}$ concentrations were presented during winter. Daily average concentrations were compared with the former 24-hour AQG, and found that they exceeded the guideline 76 to 98% of days and $PM_{2.5}$ episodes existed 7 to 19% of hours [15]. Air quality in coastal areas located in Pattaya, Thailand showed maximum of 24-hour moving average$PM_{2.5}$ concentrations close to the former 24-hour guideline suggested by the WHO [16]. In Chile, the $PM_{2.5}$ level was defined to be good when the 24-hour moving average concentration fell below 50 μg m^{-3} [17], equaling the $PM_{2.5}$ Interim target 2 of recent and previous WHO guidelines [12–13]. Development of $PM_{2.5}$ forecast in Chile used the neural network model successes to provide forecast hourly concentration. Results compared between the 24-hour moving average maxima from observed and forecasted data shown reproduced well [17]. The National Environmental Agency (NEA) of Singapore warned residents regarding the 24-hour moving average $PM_{2.5}$ concentration of 310 μg m^{-3} June 20, 2013. The maximum 24-hour moving average $PM_{2.5}$ concentration was reduced to 302 μg m^{-3} and greatly increased to 382 μg m^{-3} June 22, 2013 and June 22, 2013, respectively. The extreme concentration of 24-hour moving average was reported by the agency a day later that the actual extreme concentration was affecting residents [18]. These examples show many applications and calculation methods used to compare the 24-hour average AQG suggested by the WHO, e.g., daily average and 24-hour moving average. In addition, many countries regulated 24-hour $PM_{2.5}$ concentrations using the National Ambient Air Quality Standards (NAAQS).

$PM_{2.5}$ levels in megacities are monitored using different measurement techniques including gravimetric, beta attenuation, tapered element oscillating microbalance (TEOM) and TEOM with a filter dynamics measurement system [14]. The measurement results given by the gravimetric method are the average concentration representing for sampling period, i.e., 24 hours. On the other hand, continuous $PM_{2.5}$ measurements such as beta attenuation and TEOM can provide continuous hourly concentrations. To calculate the 24-hour concentration average of $PM_{2.5}$ from continuous monitoring equipment, the USEPA requires available data at least 75% of the 24-hour period use in calculating. The 24-hour average is stored at the first, that is 0.00 for using hourly monitoring data from 0.00 to 23.00 in calculation [19]. Therefore, averaging every block of hourly 24 data points (0.00–23.00) through the end of the time series will create a new time series of daily average concentrations. In addition to storing the average value at the starting point of the period, the average can be stored at the center or tail time period [20]. Disadvantages of the simple block average method in using hourly data generating daily time series is the loss of intra-day concentration fluctuation. The moving average method is the simple block average computation over 24 hours. e.g., 0.00 to 23.00 to obtain its mean value and then calculate the average value for the next period (1.00–0.00). Time series results given by the 24-hour moving average method can reveal variations of hourly $PM_{2.5}$ concentrations, and be evaluated using the WHO 24-hour AQG and NAAQs. The moving average technique is often used to smooth data and depict trends in a set of time series data. The trailing moving average method can be used to predict future values, whereas the central moving average technique is perhaps more appropriate to represent the actual fluctuation in time series [20]. However, it would be better to examine three recorded positions (left, center and right) of the 24-hour moving average, in which one represents the fluctuation and captures the high concentration event of hourly $PM_{2.5}$ concentration time series better than the others.

The similarity can indicate analogous characteristics of one time series with another. Euclidean distance is one of the various time series similarity measures that has been widely used. Its principle lies in the concept of point to point distance measured between two time series data [21–23]. Another index to assess similarity of time series data is cross-correlation function (CCF) or Pearson's correlation function [23]. CCF compares similarity in time series fluctuation shift (shape), whereas Euclidean distance employs comparing similarity tests in terms of different distances between two time series (magnitude). The correlation coefficient was used in a related study to measure similarity of stock prices, air temperatures, sea temperatures, wind speeds and electroencephalograms [22]. Euclidean distance was employed to reveal similarity of wind speed variation among many monitoring sites at each identical wind direction [24], The square of Euclidean distance and correlation were carried out in air quality data analysis to measure the similarity between the samples and the reference [25]. Therefore, comparing between 24-hour moving average $PM_{2.5}$ concentration data and the AQG or NAAQs would be examined in which 24-hour moving average $PM_{2.5}$ concentration time series (left, center and right) is more similar to its hourly time series and captured hourly fluctuation. In this paper, the use of Euclidean distance and correlation will demonstrate this. This provides reasons to support or negate data selection for air quality data and other analyzes.

2 Method

2.1 Air Quality Data

The data set used in this study constitutes hourly $PM_{2.5}$ observed data at various places in Bangkok, Thailand. The data sets given by 12 air quality monitoring stations belong to the Pollution Control Department (PCD), Thailand. Five air quality monitoring stations are located near the road within a distance of 5 m. The pollution control department classifies them to be roadside stations, other seven stations are located in residential areas. Information for each station is available in Table 1 and their locations are presented in Fig. 1. The continuous $PM_{2.5}$ monitoring equipment used by the PCD has to meet the methods recognized by the Pollution Control Department, Thailand. The PCD notification also mentioned that methods would be designated by the USEPA.

Table 1
Description of monitoring data

Station ID	Period	Description	Lat	Lon
02T	18 August 2019–31 December 2020	Residential area	13.7328 N	100.4877 E
03T	17 October 2018–31 December 2020	Roadside	13.6365 N	100.4143 E
05T	1 January 2018–31 December 2020	Residential area	13.6662 N	100.6057 E
10T	17 October 2018–31December 2020	Residential area	13.7799 N	100.6460 E
11T	17 October 2018–31 December 2020	Residential area	13.7755 N	100.5692 E
12T	18 August 2019–31 December 2020	Residential area	13.7081 N	100.5473 E
50T	1 January 2018–31 December 2020	Roadside	13.7299 N	100.5365 E
52T	1 January 2018–31 December 2020	Roadside	13.7276 N	100.4866 E
53T	1 January 2018–31 December 2020	Roadside	13.7954 N	100.5930 E
54T	25 January 2018–31 December 2020	Roadside	13.7925 N	100.5502 E
59T	1 January 2018–31 December 2020	Residential area	13.7832 N	100.5405 E
61T	1 January 2018–31 December 2020	Residential area	13.7697 N	100.6146 E

2.2 Statistical Analysis

To analyze and compare observed $PM_{2.5}$ data with the 24-hour average guidelines of WHO, the hourly $PM_{2.5}$ concentrations were computed as 24-hour average using the moving average method. In the case of hourly time series data such as $PM_{2.5}$ monitoring, the 24-hour average values computed by moving average could be stored at any hour within an input period of 24 hours. Generally, three places consisting of the first, middle and the last hour were used to store the 24-hour moving average value, e.g., the OpenAir package of R for air quality analysis has these three options to store the values. Thus, similar assessments of the 24-hour moving average $PM_{2.5}$ time series were recorded at the first, middle and the last hour with hourly observed $PM_{2.5}$ time series were examined. Hereafter, time series recorded at the first, middle and the last hour were entitled to be the leftmost, center and rightmost moving average $PM_{2.5}$ time series, respectively. The CCF and Euclidean distance were used to analyze the similarity between the leftmost, center and rightmost moving average $PM_{2.5}$ time series to hourly $PM_{2.5}$ time series. Similarities in terms of fluctuation shape and distance between the two time series were analyzed using the CCF and Euclidean distance, respectively.

Research using the CCF analysis was conducted such as studying the association between confirmed cases of COVID-19 and meteorologic variation [26] and using it to examine the relationship between the El Nino-Southern Oscillation (ENSO) variability represented by the Southern Oscillation Index (SOI) and associated time series of the number of new fish [27]. The CCF has been used to investigate the lead-lag relationship between the two time series in different time points, and can be used to determine the optimal time shift between of the two time series [27–28]. The correlation coefficients of 1 and − 1 indicate perfect relationships in the same and opposite directions, respectively. The strength of a relationship can be roughly explained using a verbal description without positive/negative direction by considering the coefficient value as follows: almost negligible relationship (< 0.2), small relationship (0.2–0.4), substantial relationship (0.4–0.7), marked relationship (0.7–0.9) and very dependable relationship (0.9-1.0), respectively [29]. The CCF described by Shumway and Stoffer [27] is calculated using the equations as shown below.

$$\gamma_{XY}\left(k\right)=C_{XY}\left(k\right)/S_XS_Y$$

1

,

where C_{XY} is cross-covariance function, S_X is sample standard deviations of time series *X*, and S_Y is sample standard deviations of time series *Y*. The equations used to determine cross-covariance function are

$C_{XY}=\sum_{i=1}^{N-k}\left(x_t-\bar{x}\right)\left(y_{t+k}-\bar{y}\right)/N$, N = 0, 1, …, N − 1 (2)

and

$C_{XY}=\sum_{i=1-k}^{N}\left(x_t-\bar{x}\right)\left(y_{t+k}-\bar{y}\right)/N$, N = − 1, − 2, …, −(N − 1). (3)

where lag time point is *k*, which is usually much less than the number of time points along sample time series (*N*).

When the two data sets have very positive dependable relationships, their temporal variation is quite similar to each other. We examined each relationship between the 24-hour moving average $PM_{2.5}$ time series and its hourly time series to reveal the lead-lag correlations of 72 time points (hours). A time point position showing the highest positive correlation coefficient, means the best shape similarity of both time series occurring at this time point. A good representation of the 24-hour time series for the hourly time series would have a high correlation coefficient and short lead or lag time length of the time point. The highest correlation presenting at a time point zero means no lead or lag time. This is a similarity that is shape-preserving, but represents the difference in magnitude between two time series, but a difference in magnitude (vertical shift) exists between the two time series. Comparison between the two time series based on the concept of distance measures can be performed using time series similarity measures, e.g., Euclidean distance, dynamic time warping (DTW), and others [21, 23, 30–32]. Euclidean distance is based on point to point measurement concept whereas DTW is based on the concept of point to many measurements. Both concepts are visualized in graphic form in the studies of Serra` and Arcos [32] and

Cassisi et al. [23]. This study used the point to point distance concept because we considered the coincident events between the 24-hour moving average $PM_{2.5}$ time series and hourly observed $PM_{2.5}$ time series. The calculation of similarity represented by the Euclidean distance [21, 33] can be determined using the equation

$$E_D = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2}$$

4

.

Less distance resulting in less vertical shift is more similar between both time series. Therefore, Euclidean distance and CCF analyses are performed to evaluate three types of the 24-hour moving average $PM_{2.5}$ time series in representing hourly $PM_{2.5}$ variation. Next, we analyzed the 24-hour moving average $PM_{2.5}$ time series against the 24-hour $PM_{2.5}$ average values suggested by WHO air quality guidelines. The 24-hour moving average $PM_{2.5}$ data were binned into each hour, 0.00, 1.00, ..., 23.00. Frequencies of concentrations falling in AQG, interim target (IT) 1, 2, 3, 4, and above were calculated for each hour as equation as shown below

$$F_{Th,Hr} = \frac{n_{Th,Hr}}{N_{Hr}} \times 100$$

5

.

where $F_{Th,Hr}$is the frequency of concentrations falling in each threshold (*Th*) ranges (AQG ≤ 15 μg m^{-3}, 15 μg m^{-3} <IT4 ≤ 25 μg m^{-3}, 25 μg m^{-3} <IT3 ≤ 37.5 μg m^{-3}, 37.5 μg m^{-3} <IT2 ≤ 50 μg m^{-3}, 50 μg m^{-3} <IT1 ≤ 75 μg m^{-3}, and > 75 μg m^{-3}) of an hour (*Hr*), 0.00, 1.00, ...,23.00. *N* is the total number of concentration values in an hour (*Hr*), and *n* is the number of concentration values in the threshold (*Th*) in an hour (*Hr*). The summation of $F_{Th,Hr}$on a particular hour equals 100. Visualization all of $F_{Th,Hr}$reveals the diurnal variation of each contribution of AQG and interim targets. All analyzes mentioned above were used by R statistical software and related packages.

3 Results And Discussion

3.1 Investigation of representativeness on method recording moving average value

As mentioned above, observed $PM_{2.5}$ time series data are on a scale of hours but the guideline values provided by WHO are on a 24-hour average time scale. Data recording of 24-hour average calculations can be done by storing the calculated value at any desired hour. Usually, a position considered to store an average value is leftmost, center and rightmost, e.g., when the average value of a period of 0.00 to 23.00, leftmost is 0.00 and rightmost is 23.00. If we want to compare hourly $PM_{2.5}$ time series to the 24-hour standard value, the calculation of hourly data to be the 24-hour average data is required before comparing. However, three types of 24-hour moving average recording method resulted in time shifting of high concentration peaks to the peak of hourly time series data differently as shown in Fig. 2. For a long time series period, the difference between line graphs of three 24-hour moving average $PM_{2.5}$ concentration time series and hourly monitoring data are difficult to investigate. The question is what type of recording position of 24-hour moving time series is appropriate to capture fluctuations of hourly $PM_{2.5}$ time series? Examinations using correlation analysis between hourly time series data and each of 24-hour moving average time series (leftmost, center rightmost) were performed by CCF. It provides a measure of similarity between the two time series when a shift of a curve is found to another. Correlations between both signals at lead/lag of 72 time points (time steps) reveal the temporal relationship between them. A high CCF value indicates a strong relationship representing high similarity [34]. In this case, three types of 24-hour average fluctuations perhaps signal a time shift to the hourly signal as shown in Fig. 2a. A shorter period of time shift between a 24-hour average data set and the hourly data set means a greater possibility of representing the hourly data set. Figure 3 shows the results given by CCF analysis for station 02T. The highest correlation coefficient is 0.89 at lag times from − 10 to -12 meaning the peak of the leftmost 24-hour average $PM_{2.5}$ time series occurring before the peak of hourly $PM_{2.5}$ time series is around 10 to 12 hours. For the 24-hour moving average recording at center, the highest correlation coefficient is 0.89 at a lag time from − 1 to 0 revealing coincident peaks occurring in both time series (Fig. 3b). The last one has the highest correlation of the rightmost 24-hour moving average of $PM_{2.5}$ time series to hourly time series, namely, 0.89 at lags from 10 to 13. The high $PM_{2.5}$ peak of the 24-hour time series will come later than the peak of hourly $PM_{2.5}$ time series at about 10 to 13 hours, as shown in Fig. 3c. The results provided by CCF analysis for other monitoring stations exhibit similar results as shown in the Supplementary. A summary of lag times and correlation coefficients of all stations in this study is shown in Table 2. Station 11T shows the highest correlation of 0.893 for center and rightmost 24-hour moving average with lag from − 1 to 0 and 11, respectively. The lowest correlation coefficient presenting at station 03T is 0.819 for the leftmost and rightmost 24-hour moving averages with time lags from − 12 to − 10 and 10 to 13, respectively. Overall, they present highly marked relationships. For lead and lag time between them, the leftmost, center and rightmost 24-hour moving average are lags from − 13 to − 10, − 2 to 1 and 10 to 13, respectively. The center 24-hour moving average produces time series peaks coinciding with high concentration peaks of the hourly time series more than others. Figure 2b shows the time variation of $PM_{2.5}$ for hourly, leftmost, center and rightmost 24-hour moving average time series from 1 to 31 January 2020. The 24-hour moving average time series exhibits less fluctuation than that of hourly time series data because the moving average method smooths the data but still captures the concentration fluctuation.

Table 2
Highest correlation and its corresponding lag (hr) of all stations

Station ID	Rightmost		Leftmost		Center	
	Lag	R	Lag	R	Lag	R
02T	10 to 13	0.890	-12 to -10	0.890	-1 to 0	0.891
03T	10 to 13	0.819	-12 to -10	0.819	-1 to 0	0.820
05T	11 to 13	0.881	-12 to -10	0.881	0	0.882
10T	10 to 13	0.868	-12 to -10	0.868	-1 to 0	0.869
11T	11	0.893	-13 to -10	0.892	-1 to 0	0.893
12T	11	0.861	-13 to -10	0.860	-2 to 1	0.861
50T	11 to 12	0.880	-12 to -10	0.880	-2 to 1	0.880
52T	11 to 12	0.891	-12 to -11	0.891	-2 to 1	0.891
53T	10 to 13	0.872	-12 to -10	0.872	-2 to 1	0.872
54T	10 to 12	0.839	-13 to -10	0.839	-1 to 0	0.84
59T	10 to 12	0.858	-13 to -11	0.858	-1	0.859
61T	10 to 12	0.885	-13 to -10	0.885	-1 to 0	0.886

The 24-hour moving averages can show a tendency to change in hourly time series. Variations of leftmost, center and rightmost 24-hour moving averages $PM_{2.5}$ concentrations revealed associations with hourly $PM_{2.5}$ variation that occurs before, coincident and after to hourly variations, respectively, resulting from the CCF analysis. from 8 to 12 January 2020, the leftmost 24-hour moving average time series started 9 January 2020 to the highest concentration at 23.00, whereas the hourly concentration time series presented the highest concentration 10 January 2020 at 8.00. This means the leftmost 24-hour moving average time series presenting the peak event before it occurred (Fig. 4a). The 24-hour moving average concentrations recorded at the center was quite constant from 0.00 to 7.00 9 January 2020 and after that, concentration continued rising to a peak 10 January 2020 7.00 to 9.00, which was closest to a peak event of the hourly time series (Fig. 4b). The 24-hour moving averages recorded at the rightmost maintained a quite constant low concentration from 0.00 to 20.00 9 January 2020 and the highest concentration was observed at 21.00 10 January 2020 occurring later than the highest concentration of hourly time series 10 January 2020 at 8.00 (Fig. 4c). We conclude that the 24-hour moving average $PM_{2.5}$ concentrations recorded at the center were more similar to the fluctuation of hourly $PM_{2.5}$ time series than others. This constitutes a similarity of 24-hour moving average time series to hourly time series in terms of shape fluctuation.

Another measure is the similarity in terms of distance. Euclidean distance has been widely used to examine similarity and has been used to describe the terms of distance between two time series. The distance is determined by taking the square root of the sum of the squared differences between point to point of corresponding time series. The concept of point to point distance is shown in Fig. 4. The calculated distance between the leftmost 24-hour average $PM_{2.5}$ time series and hourly $PM_{2.5}$ time series resulting from Eq. 4 is 911.96 μg m^{-3}. Distances of the center 24-hour average $PM_{2.5}$ time series and the leftmost 24-hour average $PM_{2.5}$ time series to hourly $PM_{2.5}$ time series are 776.71 μg m^{-3} and 910.98 μg m^{-3}, respectively. The Euclidean distance presented in Eq. 4 is related to summation of point to point distance along the time series. A related study considered the number of points in determining the Euclidean distance between point to origin through the whole data length by dividing the summation by the number of points [30]. Thus, we calculated the square root of the sum of squared distances (Euclidean distance) divided by the number of points ($\sqrt{E_D^2/N}$) and hereafter referred to the averaged Euclidean distance. It would present distance in terms of the average distance between the two time series. The averaged Euclidean distances between the leftmost, center and right 24-hour average $PM_{2.5}$ time series and hourly time series were 8.374 μg m^{-3}, 7.125 μg m^{-3} and 8.365 μg m^{-3}, respectively. The center 24-hour average $PM_{2.5}$ time series showed the smallest value. According to the Euclidean distance is of 0 representing the perfect similarity in terms of distance, the increasing of Euclidean distance is related to reducing the similarity. Therefore, the center 24-hour average $PM_{2.5}$ time series was more similar to the hourly $PM_{2.5}$ time series than to the rest of the time series and reducing the similarity. Therefore, the center 24-hour average $PM_{2.5}$ time series was more similar to the hourly $PM_{2.5}$ time series than to the rest of the time series. We also calculated the mean value and mean absolute value of point to point distances along the time series. The mean values of the leftmost, center and rightmost 24-hour average $PM_{2.5}$time series to the hourly time series were zero because of moving average smoothing hourly data and canceling the upper and lower residuals. The absolute mean value of distances were 5.82 μg m^{-3}, 0 μg m^{-3} and 5.87 μg m^{-3} for the leftmost, center and rightmost 24-hour average $PM_{2.5}$ time series, respectively. The reason the three mean absolute values were not zero was the mean values are the absolute mean value of distance does not account for the positive and negative directions of each distance. The mean value and mean absolute value of distances are less suitable to describe the similarity in terms of distance than the Euclidean distance.

3.2 State of $PM_{2.5}$ level associated with WHO guidelines

In 2021, WHO updated the air quality guidelines, with $PM_{2.5}$ level classification of 24-hour average concentration as five levels. The 1st, 2nd, 3rd, 41th interim targets and the guideline values were 75 μg m^{-3}, 50 μg m^{-3}, 37.5 μg m^{-3}, 25 μg m^{-3} and 15 μg m^{-3}, respectively [13]. Thailand has responded to a new version of the guidelines by revising the standard value (annual average) of $PM_{2.5}$ to 5 μg m^{-3}. For the 24-hour average standard value, the update is on a process revising the value of 50 μg m^{-3} to be 37.5 μg m^{-3}. The improved standard of 24-hour average value would affect the state of $PM_{2.5}$ level. The 24-

hour average concentrations of the station 02T was plotted by shading with $PM_{2.5}$ level classification of WHO guidelines as shown in Fig. 5. Concentration levels during the red shade and above were greater than the interim target 2 (50 μg m^{-3}), namely, the previous Thai standard value. The high concentration periods over 50 μg m^{-3} were late September 2019 to March 2020 and October 2020 to December 2020 (end of data). These periods occurred during the transition season (summer to winter monsoon) and winter. The climatic conditions that govern Thailand and neighboring countries during winter is the winter monsoon decreases temperature during this period [35]. Suppose that the $PM_{2.5}$ emission in the area is quite constant such as emissions from transportation and industrial sectors vary little throughout the year. This means the mass of $PM_{2.5}$ is also quite constant. The factor related to change in concentration would be the volume of air, which is the area at ground level multiplied by the height. The area does not change whereas the planetary boundary layer (PBL) height can vary. In the Northern Hemisphere, variability of PBL showed that the PBL height decreased during winter and increased during summer [36]. Therefore, reducing in PBL height during winter reduced the air volume that constitutes an important factor in enhancing $PM_{2.5}$ concentration in the atmosphere even when no emission increases. One half decrease of PBL height corresponds to one half decrease of air volume doubling the increasing concentration.

Another factor is $PM_{2.5}$ emissions from increasing particular emission sources. High $PM_{2.5}$ concentration during winter were mostly contributed by the traffic and transport sectors, and biomass and open burning sectors. The number of fire hotspots used to represent open burning greatly increased when compared with fire hotspot numbers during other seasons. Because the increase in registered vehicles in Bangkok and fuel consumption varies less than the intra-annual variation of biomass and open burning emissions, it possibly indicates that emissions of the transport sector were constantly suspended throughout the year. However, the time series of $PM_{2.5}$ shown in Fig. 5 exhibits most concentrations throughout the year over the interim target 2 level (25 μg m^{-3}). All in all, PBL height reduction and emissions from open burning are factors in enhancing severity of $PM_{2.5}$ concentration during the winter. Without them, the $PM_{2.5}$ level remains above the threshold, the interim target 2 level (25 μg m^{-3}), and some would exceed the interim target 3 level (37.5 μg m^{-3}) value. Accordingly, the 24-hour average $PM_{2.5}$ standard level of Thailand was strengthened from 50 μg m^{-3} (interim target 2) to 37.5 μg m^{-3} (interim target 3). Achieving this new standard is possible by reducing emissions, which equals the summation of the impacts of PBL height decreases concerning concentration increases, open burning and other sources during the winter. The contribution of reduced PBL height on increasing $PM_{2.5}$ concentration should be studied to determine the relevant increased $PM_{2.5}$ mass. Differences in open burning emissions and other sources during the winter to their emissions during the low concentration period would also be determined to reveal increasing emissions. This required that the reduced mass in $PM_{2.5}$ emission assumed that long range transport exhibited no influence. The required reduction in $PM_{2.5}$ mass amount should be assigned and distributed to various source sectors with the acceptance of stakeholders. This leads to success in achieving a lower $PM_{2.5}$ concentration level in Bangkok than the threshold level.

Another consideration is investigating the diurnal variation of 24-hour moving average $PM_{2.5}$ concentration pro-portion that is associated with each WHO guideline level. First, an investigation uses 24-hour moving average data recorded at the center. The results are shown in Fig. 6b. Blue represents percent of concentrations below the AQG value (15 μg m^{-3}), Green, yellow, orange, red and purple represent a range of 15 to 25 μg m^{-3}, 25-37.5 μg m^{-3}, 37.5–50 μg m^{-3}, 50–75 μg m^{-3} and above 75 μg m^{-3}, respectively. The proportion of concentration lower than 15 μg m^{-3} (blue) is about 40% on 0.00 and the portion reduced to 26% at 08.00. The contribution during daytime was quite constant after 08.00 and revealed few increases during the afternoon to around 30%. Then the proportion increases again until midnight. This shows that air quality during nighttime to morning exhibited a proportion of low concentration larger than during daytime, implying the air quality in terms of $PM_{2.5}$ in the night and early morning was safer for health than in late morning and afternoon. On the other hand, the proportion of concentration level above the previous Thai NAAQs (red and purple) equaling the interim target 2 (50 μg m^{-3}) is approximately 10%. The smallest proportion presented from 12.00 to 17.00 means less high concentrations accumulated during the afternoon. This corresponds to a related study reporting that high wind speed during the afternoon in Bangkok caused a greater advection process to reduce particulate matter in ambient air [7]. Moreover, Thailand changed the national standard from interim target 2 (50 μg m^{-3}) to interim target 3 (37.5 μg m^{-3}). From this result, the exceedances will increase from 10 to about 22% (orange, red and purple) but the state of air quality remains at a similar level. The possibility exists that residents may misunderstand and know the air quality becomes more severe. The government should spend more effort to reduce emissions and ambient air concentrations than earlier endeavors.

Using the leftmost and rightmost 24-hour moving average $PM_{2.5}$ time series in analysis affected the time shift of the concentration proportions. The proportion of concentrations less than AQG presenting at 5.00 as shown in Fig. 6a shifted from that occurring at 8.00 (Fig. 6b), using the center 24-hour moving average $PM_{2.5}$ time series in analysis. This time shift revealed the events preceding the real occurrence may have resulted in misinterpretation of the analysis. However, it may be useful for some analyses aiming to warn against extreme events. On the other hand, the use of the rightmost 24-hour moving average $PM_{2.5}$ time series exhibits time shift lags (Fig. 6c). Presenting a proportion less than that of delayed AQG is the proportion resulting from using the center 24-hour moving average occurring at 8.00 moving to 12.00 and the analysis using the rightmost data. The analysis of other stations in Bangkok, presents a time shift as well (shown in the Supplementary). We suggest that in the analysis using 24-hour moving average $PM_{2.5}$ data, the position of the stored data should be addressed to avoid misinterpretations and misunderstandings.

4 Conclusions

The WHO has suggested air quality guideline values and interim targets for $PM_{2.5}$ in 24-hour average, but the continuous ambient air monitoring system provides hourly $PM_{2.5}$ time series. The hourly $PM_{2.5}$ monitoring data was converted to 24-hour average time series using the moving average technique for this study. Three places, the leftmost, center and rightmost positions of the 24-hour length, are used to store moving average values. An hourly $PM_{2.5}$ monitoring data produces three data sets of 24-hour moving time series. We performed a similarity test of three 24-hour moving time series to hourly $PM_{2.5}$ concentration time series in terms of shape and distance. The CCF analysis suggested that all 24-hour time series exhibited a marked relationship to hourly $PM_{2.5}$ monitoring data. The 24-hour moving average concentration recorded at the center was more similar to the hourly concentration time series than the recorded moving average value at the leftmost and rightmost positions. The leftmost and rightmost 24-hour moving average time series showed the peak of concentration presented before and after the hourly occurring peak with lags from − 13 to − 10 and 10 to 13, respectively. The center 24-hour moving average time series had lags from − 2 to 1 to the hourly time series meaning it showed more similar events to the hourly $PM_{2.5}$ fluctuation than the leftmost and rightmost time series. The Euclidean distance to hourly time series were 5.82 μg m^{-3}, 0 μg m^{-3} and 5.87 μg m^{-3} for the leftmost, center and rightmost time series, respectively. The center 24-hour moving average time series was more similar to the observed hourly $PM_{2.5}$ monitoring data in terms of shape and distance. Thus, comparing with WHO guideline values was more suitable than others.

Levels of $PM_{2.5}$ in Bangkok were exhibited by comparison between the center 24-hour moving average time series and the AQG guidelines suggested by the WHO. Observed concentrations were binned in four WHO interim targets and AQG for 24-hour average. The proportion of concentration lower than the AQG level of 15 μg m^{-3} (blue) was about 40% at 0.00 and the portion reduced to 26% at 08.00. The contribution during daytime was quite constant after 08.00 and the small increase during afternoon was around 30%. On the other hand, the proportion of concentration level above the previous Thai NAAQs (red and purple) equaling the interim target 2 (50 μg m^{-3}) was approximately 10%. The smallest proportion of high concentration was present from 12.00 to 17.00. This implied that the level of $PM_{2.5}$ at nighttime was mostly within the interim target 4 (low concentration level). For daytime, the high concentration level (above interim target 3) occurred less from 12.00 to 17.00 meaning less possibility to expose high concentration than that in the morning and late afternoon. Moreover, the Thai national air quality standard of 24-hour $PM_{2.5}$ was changed from the interim target 2 (50 μg m^{-3}) to interim target 3 (37.5 μg m^{-3}). The exceedances will increase from 10 to about 22% but the state of air quality remains similar. Possibly, residents may misunderstand and know the air quality becomes more severe. The government should spend more effort to reduce emissions and ambient air concentrations than earlier endeavors.

Declarations

The authors declare they have no competing interests.

Availability of data and materials

All data generated or analyzed during this study are included within the submitted manuscript and the supplementary material.

Competing interests

The authors declare they have no competing interests.

Funding

This work was supported by Thailand Center of Excellence for Life Science (Public Organization) (Grant no. TC-A (ERP) 9/2564).

Authors' contributions

Siripong Sooktawee analyzed the data, discussed the results, and writing-original draft preparation. Suwimon Kanchanasuta: project administration, acquisition of funding provided conceptualization, and final proofed the manuscript. Natthaya Bunplod analyzed the data and contributed to the final manuscript.

Acknowledgements

The authors acknowledge the Pollution Control Department, Thailand for providing the monitoring data. We also thank the R program and R package developers for producing useful tools to analyze air quality.

References

1. Lin Y, Zou J, Yang W, Li CQ. A review of recent advances in research on $PM_{2.5}$ in China. Int. J. Environ. Res. Public Health. 2018;15(3):438.
2. Squizzato S, Masiol M, Rich DQ, Hopke PK. $PM_{2.5}$ and gaseous pollutants in New York State during 2005–2016: Spatial variability, temporal trends, and economic influences. Atmospheric Environ. 2018;183:209-24.
3. Gama C, Monteiro A, Pio C, Miranda AI, Baldasano JM, Tchepel O. Temporal patterns and trends of particulate matter over Portugal: a long-term analysis of background concentrations. Air Qual Atmos Health. 2018;11(4):397-407.
4. Faridi S, Shamsipour M, Krzyzanowski M, Ku¨nzli N, Amini H, Azimi F et al. Long-term trends and health impact of $PM_{2.5}$ and O_3 in Tehran, Iran, 2006–2015. Environ Int. 2018;114:37-49.
5. Hammer MS, van Donkelaar A, Li C, Lyapustin A, Sayer AM, Hsu NC et al. Global estimates and long-term trends of fine particulate matter concentrations (1998–2018). Environ. Sci. Technol. 2020;54(13):7879–90.
6. Lorelei de Jesus A, Thompson H, Knibbs LD, Kowalski M, Cyrys J, Niemi JV et al. Long-term trends in $PM_{2.5}$ mass and particle number concentrations in urban air: The impacts of mitigation measures and extreme events due to changing climates. Environ. Pollut. 2020;263:114500.
7. Kanchanasuta S, Sooktawee S, Patpai A, Vatanasomboon P. Temporal variations and potential source areas of fine particulate matter in Bangkok, Thailand. Air, Soil Water Res. 2020;13:1-10.

8. Yamagami M, Ikemori F, Nakashima H, Hisatsune K, Ueda K, Wakamatsu S et al. Trends in $PM_{2.5}$ concentration in Nagoya, Japan, from 2003 to 2018 and impacts of $PM_{2.5}$ countermeasures. Atmosphere. 2021;12(5):590.
9. Jia B, Gao M, Zhang X, Xiao X, Zhang S, Lam Yung KK. Rapid increase in mortality attributable to $PM_{2.5}$ exposure in India over 1998–2015. Chemosphere. 2021;269:128715.
10. Yu P, Xu R, Coelho MS, Saldiva PH, Li S, Zhao Q et al. The impacts of long-term exposure to $PM_{2.5}$ on cancer hospitalizations in Brazil. Environ Int. 2021;154:106671.
11. Franco IB, Chatterji T, Derbyshire E, Tracey J. Actioning the global goals for local impact: Towards sustainability science, policy, education and practice, Science for sustainable societies. Singapore: Springer; 2020.
12. World Health Organization. WHO Air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment. Geneva: WHO; 2006.
13. World Health Organization. WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: WHO; 2021.
14. Cheng Z, Luo L, Wang S, Wang Y, Sharma S, Shimadera H et al. Status and characteristics of ambient $PM_{2.5}$ pollution in global megacities. Environ Int. 2016;89-90:212-21.
15. Kumar MK, Sreekanth V, Salmon M, Tonne C, Marshall JD. Use of spatiotemporal characteristics of ambient $PM_{2.5}$ in rural South India to infer local versus regional contributions. Environ. Pollut. 2018;239:803-11.
16. Kanchanasuta S, Sooktawee S, Bunplod N, Patpai A, Piemyai N, Ketwang R. Analysis of short-term air quality monitoring data in a coastal area. AIMS Environ. Sci. 2021;8(6):517-31.
17. Perez P, Menares C. Forecasting of hourly $PM_{2.5}$ in South-West Zone in Santiago de Chile. Aerosol Air Qual Res. 2018;18(10):2666-79.
18. Velasco E, Rastan S. Air quality in Singapore during the 2013 smoke-haze episode over the Strait of Malacca: Lessons learned. Sustain. Cities Soc. 2015;17:122-131.
19. USEPA. 2006 National Ambient Air Quality Standards for Particulate Matter; Final Rule. Environmental Protection Agency; 2006.
20. Warner RA. Moving averages for identifying trends and changes in the data. Amsterdam: Elsevier; 2016.
21. Goldin DQ, Kanellakis PC. On similarity queries for time-series data: Constraint specification and implementation. In: 1st International Conference on Principles and Practice of Constraint Programming. Cassis; 1995. Sep 19-22
22. Fink E, Pratt K. Indexing of compressed time series. In: Last M, Kandel A, Bunke H, editors. Data mining in time series databases, Machine perception and artificial intelligence. Singapore: World Scientific; 2004. p. 43-65.
23. Cassisi C, Montalto P, Aliotta M, Cannata A, Pulvirenti A. Similarity measures and dimensionality reduction techniques for time series data mining. Rijeka: IntechOpen; 2012.

24. Ratto G, Maronna R, Repossi P, Videla F, Nico A, Almandos JR. Analysis of winds affecting air pollutant transport at La Plata, Argentina. Atmospheric and Climate Sciences. 2012;2(1):60-75.
25. Guo J, Zhao M, Xue P, Liang X, Fan G, Ding B et al. New indicators for air quality and distribution characteristics of pollutants in China. Build Environ. 2020;172:106723.
26. Abraham J, Turville C, Dowling K, Florentine S. Does Climate play any role in COVID-19 spreading?— An Australian perspective. Int. J. Environ. Res. Public Health. 2021;18(17):9086.
27. Shumway RH, Stoffer DS. Time series analysis and its applications, Springer texts in statistics. 4th ed. New York: Springer International Publishing; 2017.
28. Chatfield C, Xing H. The analysis of time series: an introduction with R, Texts in statistical science. 7th ed. Florida: CRC Press; 2019.
29. GUILFORD JP. Fundamental statistics in psychology and education. New York and London: McGRAW-HILL BOOK COMPANY, Inc; 1942.
30. Zhao F, Gao Y, Li X, An Z, Ge S, Zhang C. A similarity measurement for time series and its application to the stock market. Expert Syst. Appl. 2021;182:115217.
31. Kljun M, Tersˇek M. A review and comparison of time series similarity measures. In: 29th International Electrotechnical and Computer Science Conference (ERK 2020). Portorozˇ; 2020, Sep 21-22.
32. Serra`J, Arcos JL. An empirical evaluation of similarity measures for time series classification. Knowl Based Syst. 2014;67:305-14.
33. Anton H, Rorres C. Elementary linear algebra: Applications version. 11th ed. Hoboken: Wiley & Sons; 2005.
34. Derrick TR, Thomas JM. Time-series analysis: The cross-correlation function. Champaign: Human Kinetics Publishers; 2004.
35. Sooktawee S, Humphries U, Limsakul A, Wongwises P. Spatio-temporal variability of winter monsoon over the Indochina Peninsula. Atmosphere. 2014;5(1):101-21.
36. McGrath-Spangler EL, Denning AS. Global seasonal variations of midday planetary boundary layer depth from CALIPSO space-borne LIDAR: VARIATIONS OF PBL DEPTH FROM CALIPSO. J. Geophys. Res. Atmos. 2013;118(3):1226-33.

Figures

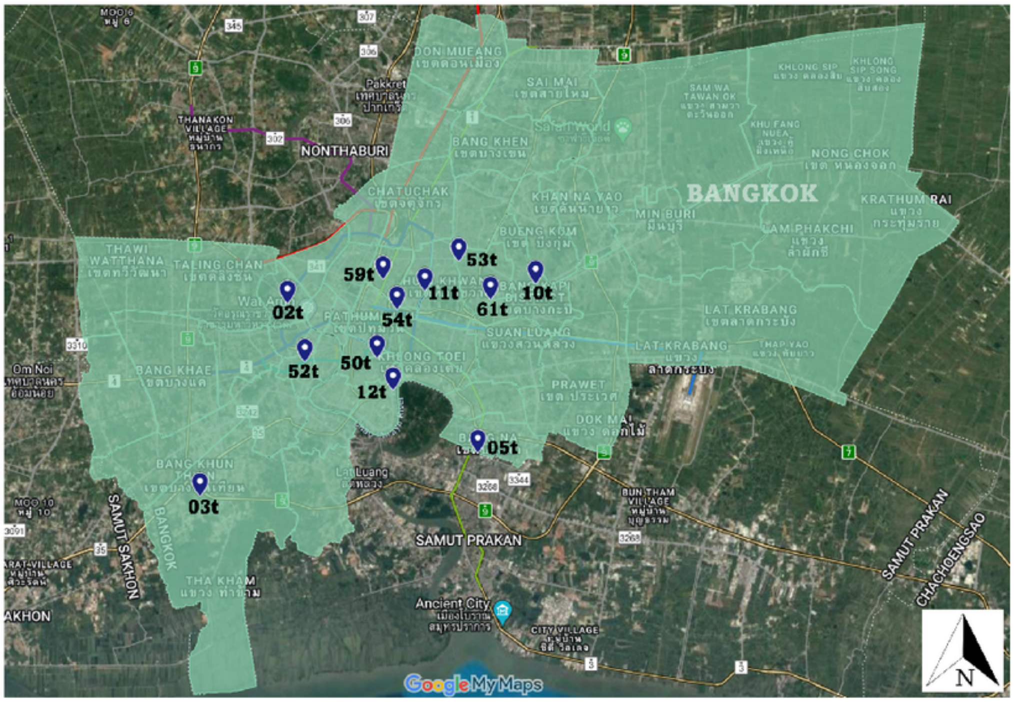


Figure 1

Locations of air quality monitoring sites

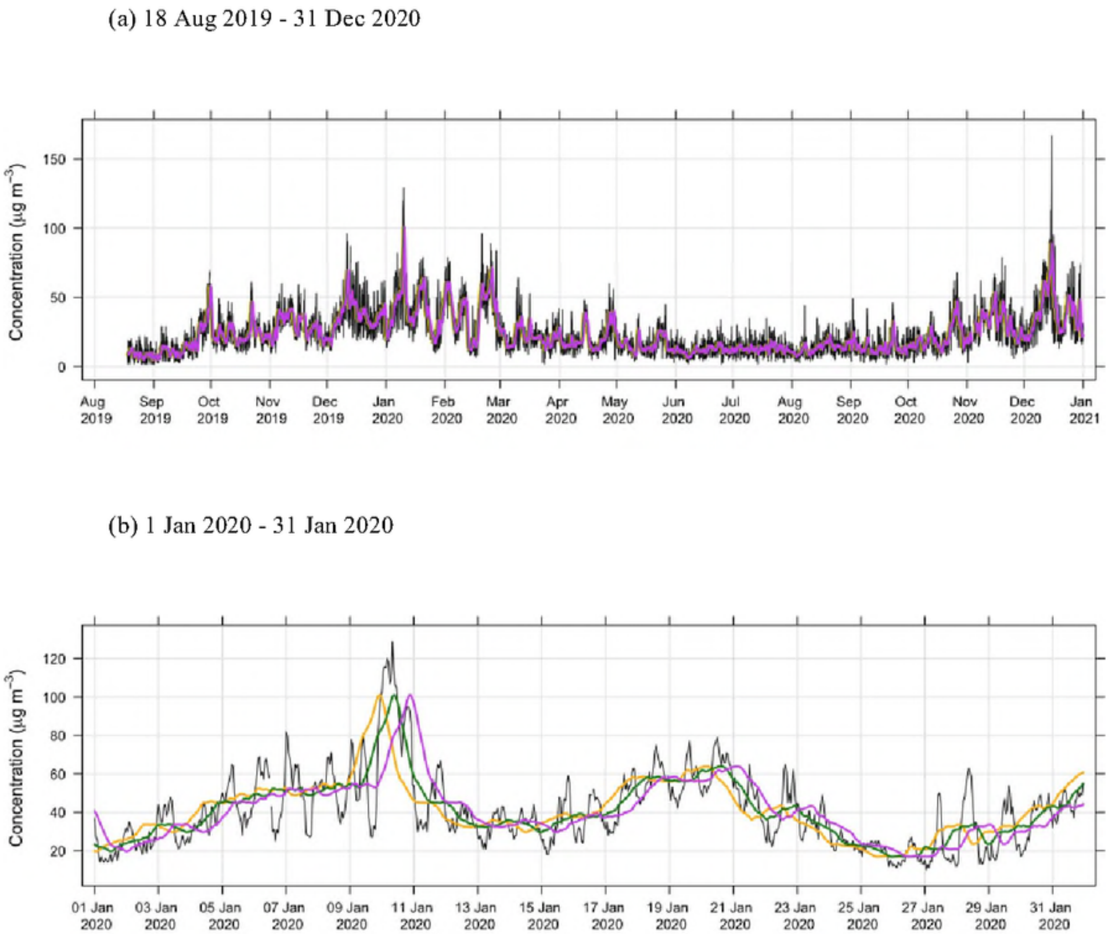
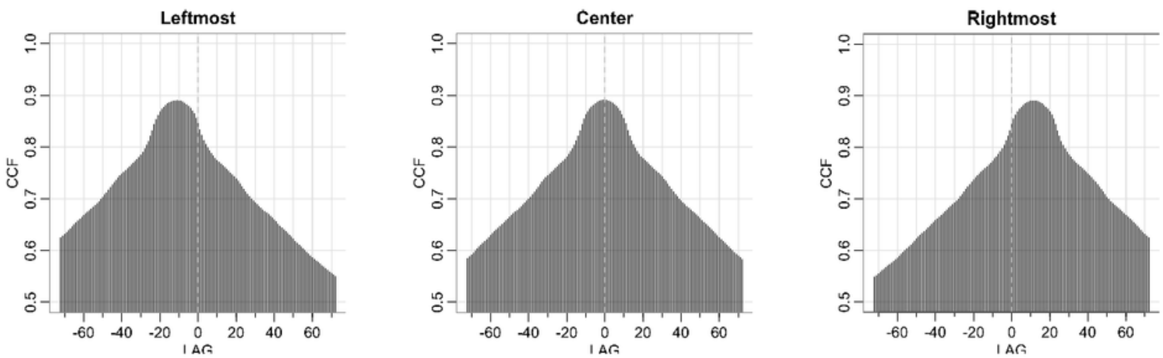


Figure 2

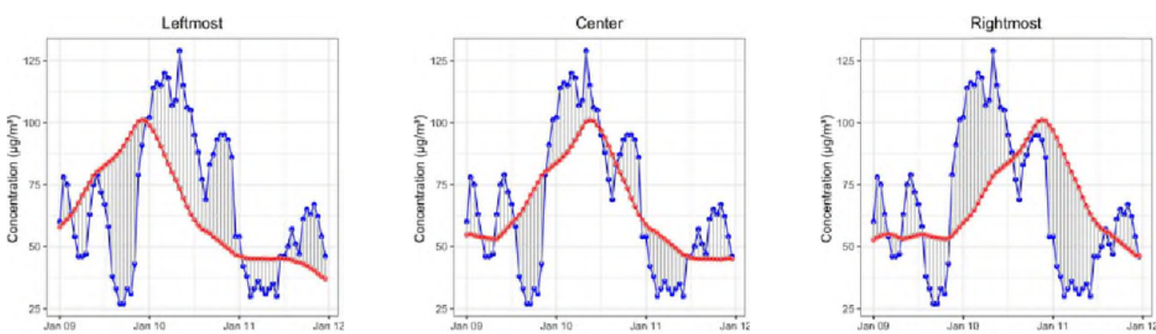
Time series variation of hourly $PM_{2.5}$ concentrations (black line) measured at station 02T and its 24-hour average $PM_{2.5}$ concentrations recorded at leftmost (yellow line), center (green line) and rightmost (purple line), respectively.



(a) Leftmost (b) Center (c) Rightmost

Figure 3

Correlograms of cross-correlation values between hourly $PM_{2.5}$ concentrations measured at station 02T and its 24-hour average $PM_{2.5}$ concentrations recorded at leftmost, center and rightmost, respectively.



(a) Leftmost (b) Center (c) Rightmost

Figure 4

Distance between hourly $PM_{2.5}$ concentrations measured at station 02T and its 24-hour average $PM_{2.5}$ concentrations recorded at leftmost, center and rightmost, respectively. Blue represents hourly time series data, red shows 24-hour moving average time series data, and gray indicated point to point distance.

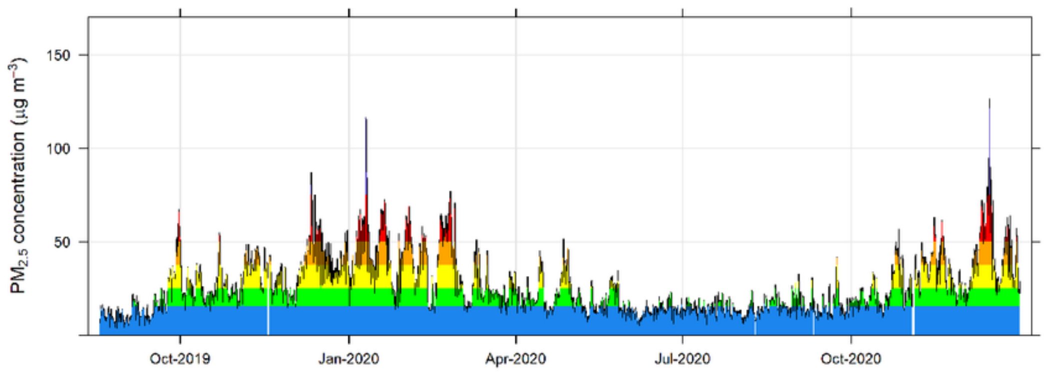


Figure 5

Time series of 24-hour average (center) $PM_{2.5}$ concentration measured at station 02T

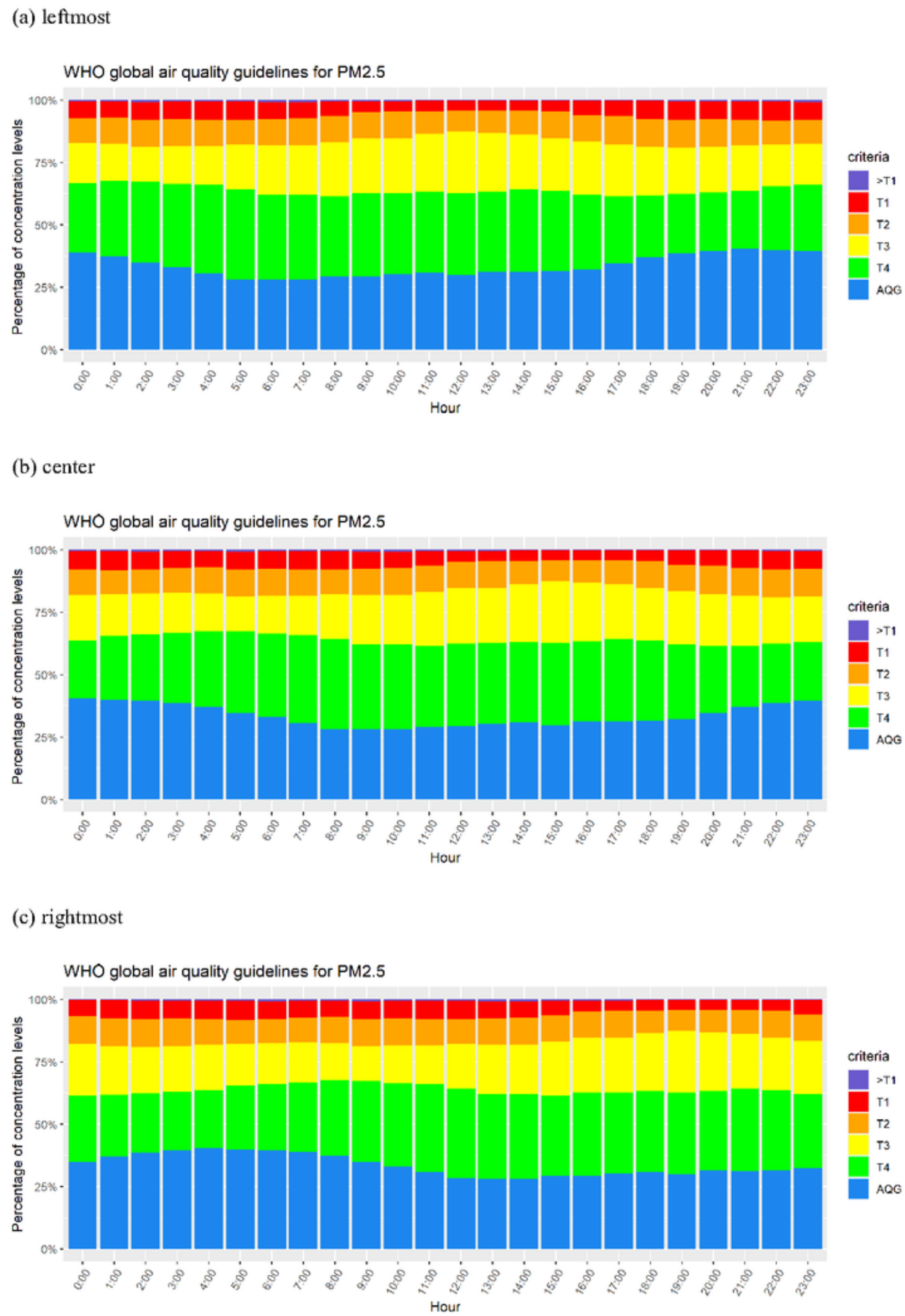


Figure 6

Time variation of $PM_{2.5}$ concentration measured against WHO guidelines at station 02T

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Suplementary.pdf