Assessment of 24-h moving average PM$_{2.5}$ concentrations in Bangkok, Thailand against WHO guidelines

Sirapong Sooktawee[1], Suwimon Kanchanasuta[2,3]* and Natthaya Bunplod[2]

Abstract

Particulate matter less than 2.5 μm in diameter (PM$_{2.5}$) pollution has gained more concern, as can be seen from the World Health Organization (WHO) revised air quality guideline (AQG) value. To compare with the 24-h AQG value, the 24-h mean (daily average) of hourly monitoring data has been widely used, but the nature of data in hourly variation has been lost. The 24-h moving average can be an alternative approach to preserving hourly data behavior, while various positions of the hour (leftmost, center, or rightmost hour) can be used to store the moving average. In this study, similarity testing by cross-correlation and Euclidean distance was performed to present a suitable 24-h moving average time series for hourly data, and then presented proportions of PM$_{2.5}$ concentrations within each interim target and AQG ranges. The PM$_{2.5}$ time series of the leftmost and rightmost 24-h moving average revealed lags of -13 to -10 h and 10 to 13 h for the observed hourly data, respectively. The lags of the center 24-h moving average time series were of -2 to 1 h, which showed more similar events to the hourly PM$_{2.5}$ fluctuation than the leftmost and rightmost time series. The center 24-h moving average concentrations measured in Bangkok were categorized into interim target and AQG ranges. The results revealed that the proportion of concentrations lower than AQG during nighttime and daytime was approximately 40 and 28%, respectively. However, the results given from the leftmost and rightmost time series showed time shifts and differences in diurnal variation caused by the position used to store the average value. To keep hourly PM$_{2.5}$ variation that can be compared with the 24-h WHO guidelines, this should be performed with careful and precautionary consideration of misinterpretation over a time scale, especially for implementing countermeasures and policies.

Keywords Bangkok, PM$_{2.5}$, Euclidean distance, Cross-correlation, Moving average

1 Introduction

Air pollution, especially that caused by particulate matter less than 2.5 μm in diameter (PM$_{2.5}$), remains a concern in many countries such as Australia, Brazil, China, India, Iran, Japan, Portugal, Thailand, and the United States [1–8]. Addressing PM$_{2.5}$ pollution is essential for sustainable development. Its annual mean levels have been included as a target to be met by 2030 under the United Nations Sustainable Development Goals (SDGs), thereby reducing the adverse per capita environmental impact in cities [9]. PM$_{2.5}$ pollution has affected health in many areas [10], and it includes carcinogenic substances such as polycyclic aromatic hydrocarbons (PAHs), whose concentration has been reported to be higher in urban areas than in semiurban and rural areas [11]. The WHO revised air quality guidelines in 2021 to challenge

*Correspondence:
Suwimon Kanchanasuta
suwimon.kan@mahidol.ac.th
[1] Environmental Research and Training Center, Ministry of Natural Resources and Environment, Pathum Thani 12120, Thailand
[2] Faculty of Public Health, Mahidol University, Bangkok 10400, Thailand
[3] Center of Excellence on Environmental Health and Toxicology, Bangkok 10400, Thailand

Sooktawee *et al. Sustainable Environment Research*   (2023) 33:3

Page 2 of 12

the global community to enhance air quality and reduce the health burden. Compared with the 2006 guidelines, in the 2021 guidelines, the recommended 24-h mean and annual mean $PM_{2.5}$ values were revised from 25 to $15\,\mu g\,m^{-3}$ and from 10 to $5\,\mu g\,m^{-3}$, respectively [12, 13].

A previous study compared the air quality guideline (AQG) value for a 24-h mean $PM_{2.5}$ concentration with the actual ambient $PM_{2.5}$ levels in 45 megacities in the world and found none to meet even the 2006 AQG value ($25\,\mu g\,m^{-3}$) [14]. A study from rural South India reported high $PM_{2.5}$ concentrations during winter, with daily average concentrations exceeding the former 24-h AQG most (76–98%) of days, and $PM_{2.5}$ pollution episodes existing 7–19% of the total hours [15]. The coastal area of Pattaya, Thailand showed a maximum of 24-h moving average $PM_{2.5}$ concentrations close to the former 24-h AQG value [16]. In Chile, the $PM_{2.5}$ level was defined as "good" when the 24-h moving average concentration was $<50\,\mu g\,m^{-3}$ [17], meeting the $PM_{2.5}$ Interim Target-2 of both 2006 and 2021 WHO guidelines [12, 13]. The study used a neural network for $PM_{2.5}$ forecasting in Chile and successfully provide forecast hourly concentration; the 24-h moving average maxima between the observed and forecasted data were comparable [17]. The National Environmental Agency of Singapore warned residents regarding the 24-h moving average $PM_{2.5}$ concentration of $310\,\mu g\,m^{-3}$ on June 20, 2013. The maximum 24-h moving average $PM_{2.5}$ concentration decreased to $302\,\mu g\,m^{-3}$ and greatly increased to $382\,\mu g\,m^{-3}$ on June 22, 2013, and June 22, 2013, respectively. This excessive $PM_{2.5}$ concentration was reported to affect the residents [18]. These examples demonstrated the various applications of methods used to calculate and compare the 24-h average AQG, including the daily average and 24-h moving average. Many countries have adopted the U.S. National Ambient Air Quality Standards (NAAQS) for 24-h $PM_{2.5}$ concentrations.

$PM_{2.5}$ levels in megacities are monitored using different measurement techniques including gravimetric, beta attenuation, tapered element oscillating microbalance (TEOM) and TEOM fitted with a filter dynamics measurement system [14]. The gravimetric method provides the average concentration representing for sampling period, i.e., 24 h, whereas continuous $PM_{2.5}$ measurement methods, such as beta attenuation and TEOM can provide continuous hourly concentrations. To calculate the 24-h concentration average of $PM_{2.5}$ from continuous monitoring equipment, the US EPA requires the data of at least 75% of the 24-h period used in calculating. The 24-h average is stored in the first hour, which is 0:00 [19]. Therefore, averaging every block of 24-h data points (0:00-23:00) through the end of the time series yields a new time series of daily

average concentrations. In addition to storing the average value at the start, it can be stored in the middle or at the end [20]. The disadvantages of using the simple block average method to generate daily time series is the loss of intraday concentration fluctuation. The moving average method is the simple block average computation over 24 h, including 0:00 to 23:00, to obtain a mean value; then, the average value for the subsequent period (1:00-0:00) is calculated. The time series results provided using the 24-h moving average method can reveal variations in hourly $PM_{2.5}$ concentrations and be evaluated using the WHO 24-h AQG and NAAQS. The moving average technique is often used to smooth data and depict trends. The trailing moving average method can be used to predict future values, whereas the central moving average technique is perhaps more appropriate for representing the actual fluctuation in time series [20]. However, examining the three recorded positions (left, center and right) of the 24-h moving average is crucial to determine whether one of them represents time series fluctuation and captures the high concentration event of hourly $PM_{2.5}$ concentration better than the others.

The similarity can indicate the analogous characteristics between two time series data. Euclidean distance is a widely used time series similarity measures [21–23]. Another index is the cross-correlation function (CCF) or Pearson's correlation function [23]. The CCF evaluates similarity in time series fluctuation shift (shape), whereas Euclidean distance evaluates similarity tests in terms of different distances between two time series (magnitude). The correlation coefficient was used in a study to measure the similarity of stock prices, air temperatures, sea temperatures, wind speeds and electroencephalograms [22]. Euclidean distance was employed to compare wind speed variation between many monitoring sites for the same wind direction [24]. A study analyzing air quality data used the square of Euclidean distance and correlation to compare the samples and the reference [25]. Given that variations in the three moving average time series (left, center, or right) differ from the hourly observed data behavior, when comparing 24-h moving average $PM_{2.5}$ concentration data and the AQG or NAAQS, it should be examined which moving average series is the most similar to the hourly time series and captured hourly fluctuation.

In this study, we performed similarity testing using cross-correlation and Euclidean distance to provide an acceptable 24-h period and then compared the three types of the 24-h moving average $PM_{2.5}$ concentration data with the observed hourly data and the AQG or NAAQS to offer an acceptable 24-h period. Our study is expected to maintain the data characteristics, and

proposes arguments for or against the choice of data analysis for further studies on air quality and others.

2 Methods

2.1 Air quality data

We obtained hourly $PM_{2.5}$ observed data from 12 air quality monitoring stations at various places in Bangkok, Thailand; the data sets were provided by Thailand's Pollution Control Department (PCD). Of the 12 stations, five are located within 5 m of the road (the PCD classifies them as "roadside stations"), and the other seven are located in residential areas (Table 1 and Fig. 1). The continuous $PM_{2.5}$ monitoring equipment used met the methods recognized by the PCD and also met the standards of the US EPA.

2.2 Statistical analysis

To compare the observed $PM_{2.5}$ data with the 24-h average AQG value, the hourly $PM_{2.5}$ concentrations were computed as the 24-h average using the moving average

Table 1 Description of monitoring data

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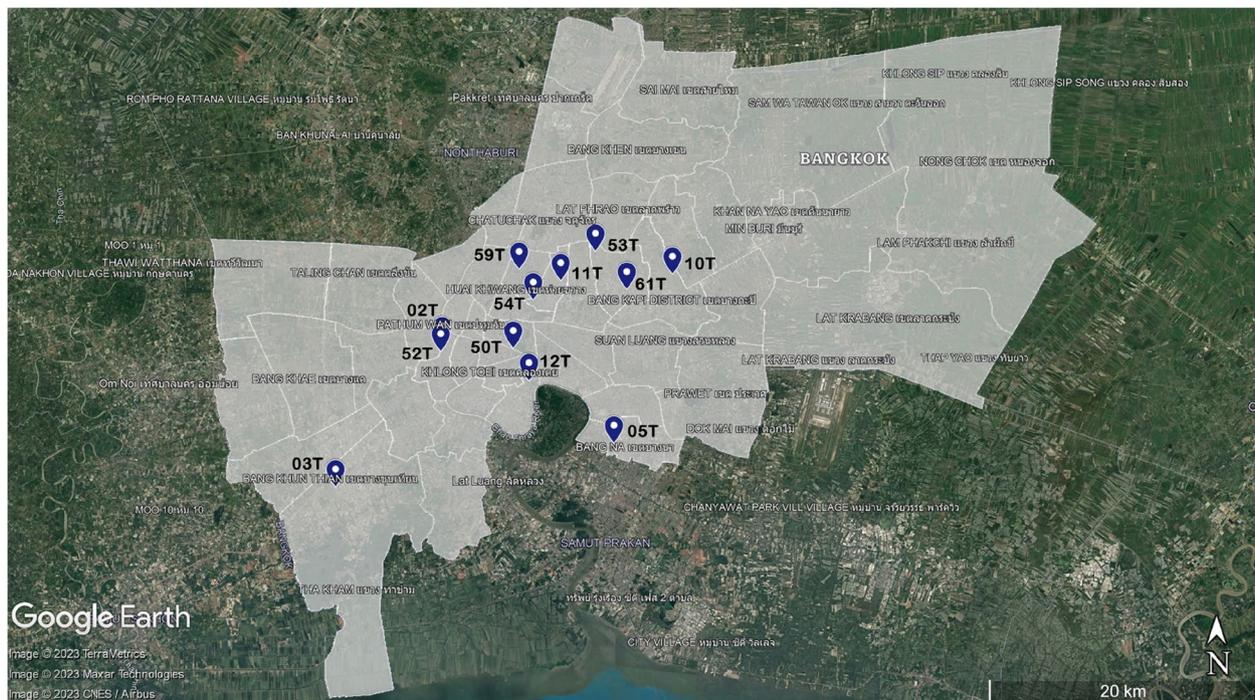
Fig. 1 Locations of air quality monitoring sites

method. This moving average was stored in the first, middle, or last hour (or leftmost, center, and rightmost hour) by using the OpenAir package for air quality analysis, and the similarities between the three types of values were compared. The CCF and Euclidean distance were used to compare the leftmost, center, and rightmost moving average $PM_{2.5}$ time series with hourly $PM_{2.5}$ time series. Similarities in terms of fluctuation shape and distance between the two time series were analyzed using CCF and Euclidean distance, respectively.

Several studies have used CCF for various reasons—for example, to examine the association between confirmed cases of COVID-19 and meteorologic variation [26] to evaluate the relationship between the El Nino-Southern Oscillation variability represented by the Southern Oscillation Index and the associated time series of the number of new fish [27], and to investigate the lead–lag relationship between the two time series at different time points; CCF can also be used to determine the optimal time shift between the two time series [27, 28]. The correlation coefficients of 1 and − 1 indicate perfect relationships in the same and opposite directions, respectively, with other positive or negative values implying the following: no relationship (0), 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 Chatfield and Xing [28] was calculated as

$$\gamma_{XY}(k) = \frac{C_{XY}(k)}{S_X S_Y} \tag{1}$$

where $\gamma_{XY}(k)$ is cross-correlation coefficient at lag *k*. 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 the cross-covariance function are

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

and

$$C_{XY} = \sum_{i=1-k}^{N} (x_t - \bar{x})(y_{t+k} - \bar{y})/N, \ N = -1, -2, \ldots, -(N-1) \tag{3}$$

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

When the two data sets have very positive dependable relationships, their temporal variations are similar. We examined the relationships between the 24-h moving average $PM_{2.5}$ time series and its hourly time series to reveal the lead–lag correlations of 72 time points (h). A time point showing the highest positive correlation coefficient, indicates the best shape similarity of both time

series at this time point. A good representation of the 24-h time series for the hourly time series would have a high correlation coefficient and a short lead or lag time length of the time point. The highest correlation presenting at a time point of zero indicates no lead or lag time. This similarity is shape-preserving, but has different magnitudes (vertical shifts) between two time series. Comparison between the two time series based on the concept of distance measures can be performed using time series similarity measures, including Euclidean distance and dynamic time warping (DTW) [21, 23, 30–32]. Euclidean distance is based on the point-to-point measurement concept whereas DTW is based on the concept of one-to-many to obtain minimum distance. Both concepts are visualized in graphic form in the studies of Serra and Arcos [32] and Cassisi et al. [23]. We used the point-to-point distance concept because we considered the coincident events between the 24-h moving average $PM_{2.5}$ and hourly $PM_{2.5}$ time series. The calculation of similarity represented by the Euclidean distance [21, 33] can be determined using the following equation:

$$E_D = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \tag{4}$$

Less distance resulting in less vertical shift is more similar between both time series. Therefore, Euclidean distance and CCF analyses were calculated to evaluate the three types of the 24-h moving average. Next, we analyzed the 24-h moving average $PM_{2.5}$ time series against the 24-h $PM_{2.5}$ average values suggested by the WHO AQG. The 24-h 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 follows:

$$F_{Th,h} = \frac{n_{Th,h}}{N_h} \tag{5}$$

where $F_{Th,\,h}$ is the frequency of concentrations falling in each threshold (*Th*) ranges (AQG: $\leq 15\,\mu g\,m^{-3}$, IT4: $15–25\,\mu g\,m^{-3}$, IT3: $25–37.5\,\mu g\,m^{-3}$, IT2: $37.5–50\,\mu g\,m^{-3}$, IT1: $50–75\,\mu g\,m^{-3}$, and $> 75\,\mu g\,m^{-3}$) of an hour (*h*). *N* is the total number of concentration values in an hour (*h*), and *n* is the number of concentration values in the threshold (*Th*) in an hour (*h*). The summation of $F_{Th,\,h}$ on a particular hour equals 100. Visualization all of $F_{Th,\,h}$ reveals the diurnal variation of each contribution of AQG and ITs.

3 Results and discussion
3.1 Investigation of representativeness on method recording moving average value
The three types of 24-h moving average recording method resulted in the time shifting of high

concentration peaks to the peak of hourly time series data (Fig. 2). For a long time series period, the difference between line graphs of three 24-h moving average $PM_{2.5}$ concentration time series and hourly monitoring data is difficult to investigate. What type of recording position of 24-h 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 24-h moving average time series (leftmost, center rightmost) were performed by the CCF. This provides a measure of the similarity between the two time series when a curve shift is found to the original time series. Correlations between both signals at lead/lag of 72 time points (time steps) reveal their temporal relationship. A high CCF value indicates a strong relationship representing high similarity [34]. In this case, three types of 24-h average fluctuations may signal a time shift to the hourly signal (Fig. 2a). A shorter period shift between a 24-h average data set and the hourly data set means a greater possibility of representing the hourly data set. Figure 3 presents the results given by the CCF analysis

for station 02 T. The highest correlation coefficient is 0.89 at lag times from -10 to -12 h meaning the peak of the leftmost 24-h average $PM_{2.5}$ time series occurring before the peak of the hourly $PM_{2.5}$ time series is approximately 10 to 12 h. For the 24-h moving average recording at the center, the highest correlation coefficient was 0.89 at a lag time from -1 to 0 h revealing coincident peaks occurring in both time series (Fig. 3b). The last one had the highest correlation (0.89) of the rightmost 24-h moving average of $PM_{2.5}$ time series to hourly time series at lags from 10 to 13 h. The high $PM_{2.5}$ peak of the 24-h time series arrived later than the peak of the hourly $PM_{2.5}$ time series at approximately $10-13$ h (Fig. 3c). The results provided by the CCF analysis for other monitoring stations exhibit similar results as shown in the Supplementary Materials. A summary of the lag times and correlation coefficients of all stations in this study is presented in Table 2. Station 11 T exhibited the highest correlation of 0.893 for the center and rightmost averages with lags from -1 to 0 h and 11 h, respectively. The lowest correlation coefficient presented at station 03 T, with 0.819 for the leftmost

(a) Aug 18, 2019 - Dec 31, 2020
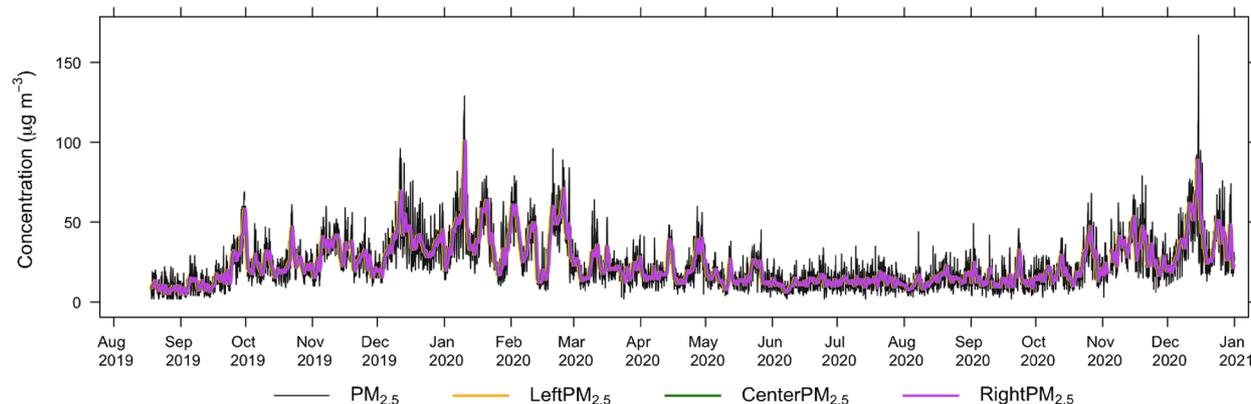
(b) Jan 1, 2020 - Jan 31, 2020
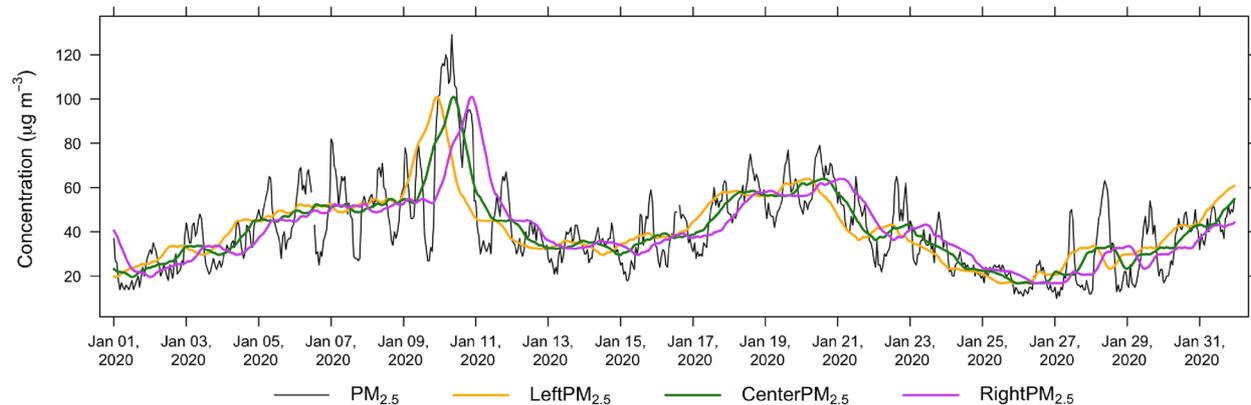
Fig. 2 Time series variation of hourly $PM_{2.5}$ concentrations (black line) measured at station 02 T and its 24-h average $PM_{2.5}$ concentrations recorded at leftmost (yellow line), center (green line), and rightmost (purple line), respectively
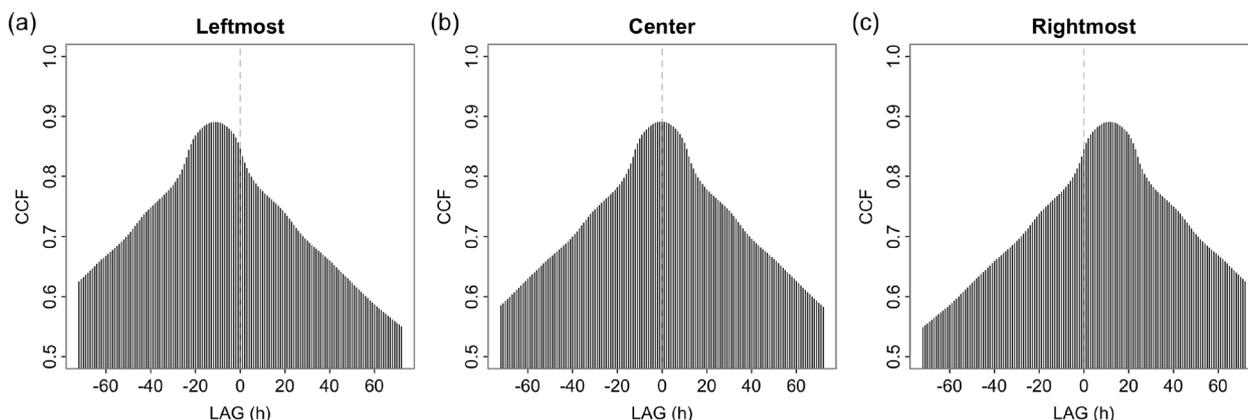
Fig. 3 Correlograms of cross-correlation values between hourly PM$_{2.5}$ concentrations measured at station 02T and its 24-h average PM$_{2.5}$ concentrations recorded at leftmost, center, and rightmost, respectively

and rightmost 24-h moving averages, with time lags from -12 to -10h and 10 to 13h, respectively. Overall, they presented highly marked relationships. For lead and lag time between them, the leftmost, center, and rightmost 24-h moving average are lags from -13 to -10h, -2 to 1h, and 10 to 13h, respectively. The center 24-h moving average produces time series peaks coinciding with the high concentration peaks of the hourly time series more than the others. Figure 2b presents the time variation of PM$_{2.5}$ for hourly, leftmost, center, and rightmost 24-h moving average time series from January 1 to 31, 2020. The 24-h moving average time series exhibited less fluctuation than that of hourly time series data because the moving average method smoothed the data but still captures concentration fluctuation.

The 24-h moving averages tended to change in hourly time series. Variations of leftmost, center, and rightmost 24-h 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. On January 8–12, 2020, the leftmost 24-h moving average time series started January 9, 2020 to the highest concentration at 23:00, whereas the hourly concentration time series presented the highest concentration January 10, 2020 at 8:00, meaning that the leftmost 24-h moving average time series indicated the peak event before it occurred (Fig. 4a). The 24-h moving average concentrations recorded at the center was quite constant from 0.00 to 7:00 January 9, 2020, and thereafter, concentration continued rising to a peak January 10, 2020 7:00–9:00, which was closest to a peak event of the hourly time series (Fig. 4b). The 24-h moving averages recorded at the rightmost maintained a quite constant low concentration

Table 2 Highest correlation and its corresponding lag (h) 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

from 0:00–20:00 January 9, 2020, and the highest concentration was observed at 21:00 January 10, 2020 occurring later than the highest concentration of hourly time series January 10, 2020 at 8:00 (Fig. 4c). We conclude that the 24-h moving average $PM_{2.5}$ concentrations recorded at the center were more similar to the fluctuation of the hourly $PM_{2.5}$ time series than the others. This constitutes a similarity of 24-h 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 points to points of the corresponding time series. The concept of point-to-point distance is illustrated in Fig. 4. The calculated distance between the leftmost 24-h average $PM_{2.5}$ time series and the hourly $PM_{2.5}$ time series resulting from Eq. (4) is $912\,\mu g\,m^{-3}$. The distances of the center 24-h average $PM_{2.5}$ time series and the leftmost 24-h average $PM_{2.5}$ time series to the hourly $PM_{2.5}$ time series are 777 and $911\,\mu g\,m^{-3}$, respectively. Euclidean distance presented in Eq. (4) is related to the 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 presents the distance in terms of the average distance between the two time series. The averaged Euclidean distances between the leftmost, center, and rightmost 24-h average $PM_{2.5}$ time series and hourly time series were 8.4, 7.1, and $8.4\,\mu g\,m^{-3}$, respectively. The center 24-h average $PM_{2.5}$ time series showed the smallest value. According to the Euclidean distance of 0 representing the perfect similarity in terms of distance, increasing the Euclidean distance is related to reducing the similarity. Therefore, the center 24-h 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 reduced the similarity. Therefore, the center 24-h 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 the mean absolute value of point-to-point distances along the time series. The mean values of the leftmost, center, and rightmost 24-h 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 values of distances were 5.8, 0, and $5.9\,\mu g\,m^{-3}$ for the leftmost, center, and rightmost 24-h average $PM_{2.5}$ time series, respectively. The reason the three mean absolute values were not zero was the mean values, because 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 for describing the similarity in terms of distance than 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-h average concentration as five levels. The 1st, 2nd, 3rd, and 4th interim targets and the guideline values were 75, 50, 37.5, 25, and $15\,\mu g\,m^{-3}$, respectively [13]. Thailand has responded to a new version of the guidelines by revising the standard
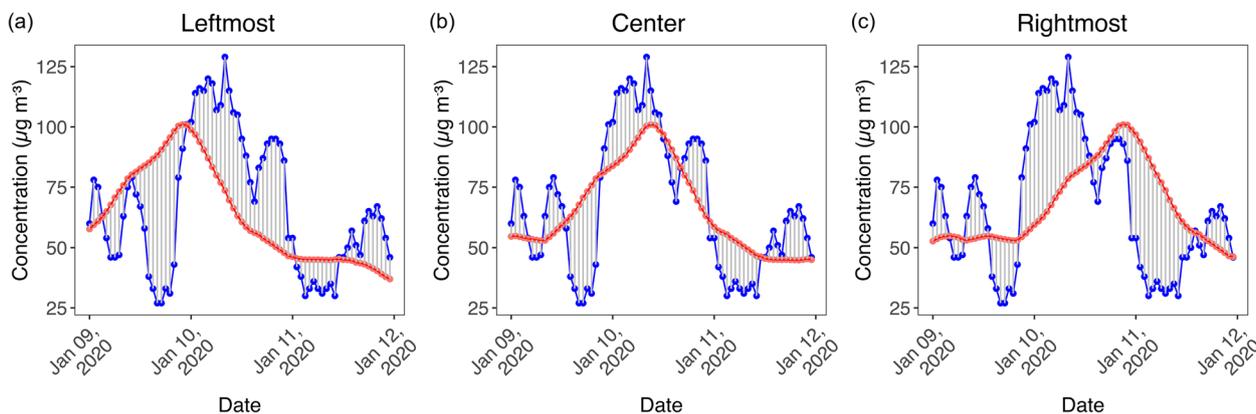
Fig. 4 Distance between hourly $PM_{2.5}$ concentrations measured at station 02T and its 24-h average $PM_{2.5}$ concentrations recorded at leftmost, center, and rightmost, respectively. Blue color represents hourly time series data, red color represents 24-h moving average time series data, and gray color is point to point distance

Sooktawee *et al. Sustainable Environment Research*   (2023) 33:3

Page 8 of 12

value (annual average) of $PM_{2.5}$ to $5\,\mu g\,m^{-3}$. For the 24-h average standard value, the update is on a process revising the value of 50 to be $37.5\,\mu g\,m^{-3}$. The improved standard of 24-h average value would affect the state of $PM_{2.5}$ level. The 24-h average concentrations of the station 02 T were plotted by shading with $PM_{2.5}$ level classification of WHO guidelines (Fig. 5). Concentration levels during the red shade and above were greater than the interim target 2 ($50\,\mu g\,m^{-3}$), namely, the previous Thai standard value. The high concentration periods over $50\,\mu g\,m^{-3}$ were late September 2019 to March 2020 and October 2020 to December 2020 (end of data) because atmospheric conditions do not favor pollutant dispersion. 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 that the winter monsoon decreases temperature during this period [35]. If $PM_{2.5}$ emissions in an area are constant, as emissions from transportation and industrial sectors are quite stable in terms of time variation throughout the year, the mass of $PM_{2.5}$ would also be constant. The factor related to change in concentration is 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, the variability of PBL indicated that PBL height decreased during winter and increased during summer [36]. Therefore, reducing PBL height during winter reduced the air volume, which constitutes an essential factor in enhancing $PM_{2.5}$ concentration in the atmosphere, even when no emission increases. A one-half decrease in PBL height corresponds to a one-half decrease in air volume doubling the increasing concentration.

Another factor is $PM_{2.5}$ emissions during winter, mainly due to traffic, biomass and open burning. The number of fire hotspots used to represent open burning was considerably greater in winter than that in the other seasons. Because the increase in registered vehicles in Bangkok and fuel consumption varies less than the intra-annual variation in biomass and open-burning emissions, the transport sector emission may be more or less constant throughout the year. However, the time series of $PM_{2.5}$ illustrated in Fig. 5 exhibits high concentrations throughout the year over the interim target 2 level ($25\,\mu g\,m^{-3}$). Overall, PBL height reduction and emissions from open burning were factors that enhanced the severity of $PM_{2.5}$ concentration during the winter. With them, the $PM_{2.5}$ level remains above the threshold, the interim target 2 ($25\,\mu g\,m^{-3}$), and some would exceed the interim target 3 ($37.5\,\mu g\,m^{-3}$) value. Accordingly, the 24-h average $PM_{2.5}$ standard level of Thailand was strengthened from $50\,\mu g\,m^{-3}$ (interim target 2) to $37.5\,\mu g\,m^{-3}$ (interim target 3). Achieving the new standards is possible by reducing emissions, which equals the summation of the impacts of PBL height decreases, 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 from 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-h moving average $PM_{2.5}$ concentration proportion that is associated with each WHO guideline level. First, we investigated the 24-h moving average data recorded at the center. The results are illustrated in Fig. 6b. Blue, green, yellow, orange,
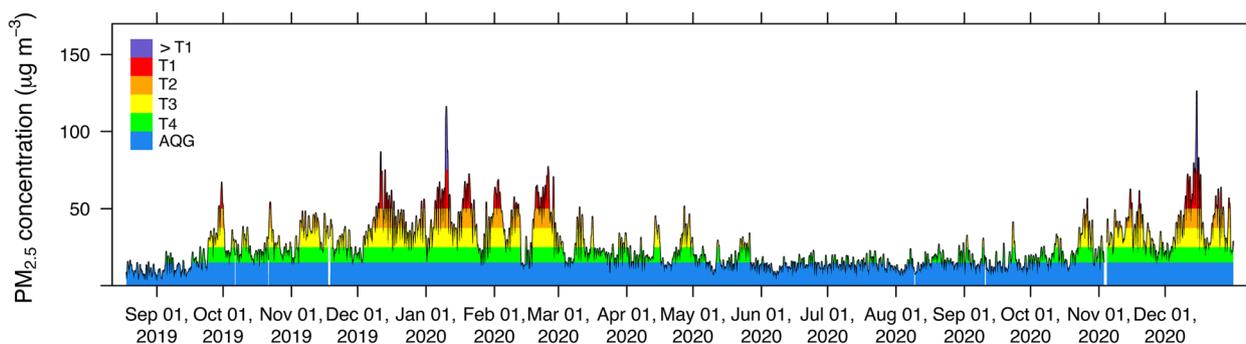
Fig. 5 Time series of 24-h average (center) $PM_{2.5}$ concentration measured at station 02 T

Sooktawee *et al. Sustainable Environment Research*　　(2023) 33:3

Page 9 of 12

red and purple represent percentage of concentrations within AQG value < 15, $15-25$, $25-37.5$, $37.5-50$, $50-75$, and $\geq 75\,\mu g\,m^{-3}$, respectively. The proportion of concentrations lower than $15\,\mu g\,m^{-3}$ (blue) was approximately 40% on 0:00 and 26% at 08:00. The contribution during the daytime was quite constant after 08:00 and revealed quite increases during the afternoon to approximately 30%. Then the proportion of concentrations lower than AQG increased again until midnight. These finding imply that air quality in terms of $PM_{2.5}$ was safer for health in the night and early morning than in the 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 $(50\,\mu g\,m^{-3})$ was approximately 10%. The smallest proportion presented from 12:00 to 17:00 means that less high concentrations accumulated during the afternoon. This corresponds to a study reporting that high wind speed during the afternoon in Bangkok caused a greater advection process to reduce ambient air $PM_{2.5}$ [37]. Moreover, Thailand changed the national standard from interim target 2 $(50\,\mu g\,m^{-3})$ to interim target 3 $(37.5\,\mu g\,m^{-3})$. From this result, the exceedances will increase from 10% to approximately 22% (orange, red and purple) but the state of air quality remains at a similar level. Residents may misunderstand and know that the air quality becomes more severe. The government should spend more effort reducing emissions and ambient air concentrations than earlier endeavors.

Using the leftmost and rightmost 24-h moving average $PM_{2.5}$ time series in the analysis affected the time shift of the concentration proportions. The proportion of concentrations less than AQG presenting at 5:00 (Fig. 6a), shifted from that occurring at 8:00 (Fig. 6b), using the center 24-h moving average $PM_{2.5}$ time series in analysis. This time shift revealed that 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-h moving average $PM_{2.5}$ time series exhibited time-shift lags (Fig. 6c). Presenting a proportion less than that of delayed AQG is the proportion resulting from using the center 24-h moving average occurring at 8:00 moving to 12:00 and the analysis using the rightmost data. The analysis of other stations in Bangkok also presents a time shift (shown in the Supplementary Materials). We suggest that in the analysis using 24-h moving average $PM_{2.5}$ data, the position of the stored data should be addressed to avoid misinterpretations and misunderstandings.

4 Conclusions

The WHO's AQG value and interim targets for $PM_{2.5}$ in 24-h average, but the continuous ambient air monitoring system provides hourly $PM_{2.5}$ time series. Therefore, we converted the hourly $PM_{2.5}$ monitoring data to 24-h average time series using the moving average technique, storing the moving average values at the leftmost, center, and rightmost positions. We compared the three 24-h moving time series with hourly $PM_{2.5}$ concentration time series in terms of shape (CCF) and magnitude (Euclidean distance). The CCF analysis suggested that all 24-h time series exhibited a marked relationship with hourly $PM_{2.5}$ monitoring data. The 24-h 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-h moving average time series exhibited the peak of concentration presented before and after the hourly occurring peak with lags from -13 to $-10\,h$ and 10 to 13h, respectively. The center 24-h moving average time series had lags from -2 to 1h 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, 0, and $5.87\,\mu g\,m^{-3}$ for the leftmost, center, and rightmost time series, respectively. The center 24-h 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 guidelines, values were more suitable than others.

$PM_{2.5}$ levels in Bangkok were compared between the center 24-h moving average time series and the AQG values. The observed concentrations were binned in four WHO interim targets and AQG for a 24-h average. The proportion of concentration lower than the AQG level of $15\,\mu g\,m^{-3}$ (blue) was approximately 40% at 0:00 and the portion reduced to 26% at 08:00. The contribution during daytime was more or less contrast after 08:00 increasing slightly to 30% in the afternoon. On the other hand, the proportion of concentration level above the previous Thai NAAQS (red and purple) equaling the interim target 2 $(50\,\mu g\,m^{-3})$ was approximately 10%. The smallest proportion of high concentration was observed 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-h $PM_{2.5}$ was revised from interim target 2 $(50\,\mu g\,m^{-3})$

(a) Leftmost

(b) Center

(c) Rightmost
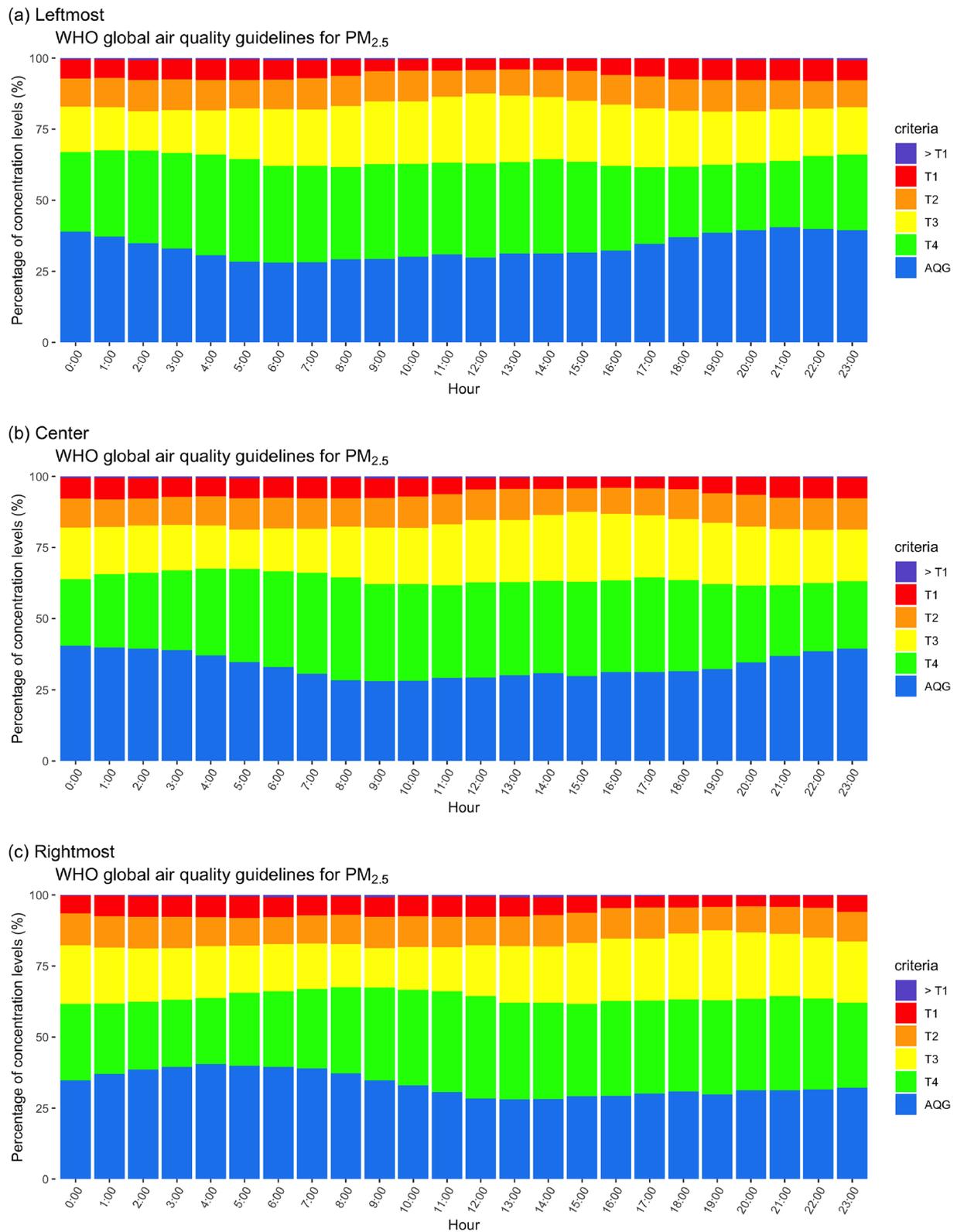
Fig. 6 Time variation of PM$_{2.5}$ concentration measured against WHO guideline at station 02T

to interim target 3 ($37.5\,\mu g\,m^{-3}$). The exceedances will increase from 10 to approximately 22% but the state of air quality remains similar. This may cause residents to misunderstand the information and know the air quality is becoming more severe. The government should spend more effort reducing emissions and ambient air concentrations.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s42834-023-00165-y.

Additional file 1: Fig. S1. Correlograms of cross-correlation values between hourly $PM_{2.5}$ concentrations and its 24-h average $PM_{2.5}$ concentrations recorded at leftmost, center, and rightmost, respectively. **Fig. S2.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 03T station. **Fig. S3.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 05T station. **Fig. S4.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 10T station. **Fig. S5.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 11T station. **Fig. S6.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 12T station. **Fig. S7.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 50T station. **Fig. S8.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 52T station. **Fig. S9.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 53T station. **Fig. S10.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 54T station. **Fig. S11.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 59T station. **Fig. S12.** Diurnal hourly $PM_{2.5}$ proportion against WHO levels for the 61T station.

Authors' contributions
Siripong Sooktawee analyzed the data, discussed the results, and writing-original draft preparation. Suwimon Kanchanasuta: project administration, acquisition of funding, provided conceptualization, discussed the results and final proofed the manuscript. Natthaya Bunplod analyzed the data and contributed to the final manuscript. The author(s) read and approved the final manuscript.

Availability of data and materials
All data generated or analyzed during this study are included within the submitted manuscript and the Supplementary Materials.

Declarations

Competing interests
The authors declare they have no competing interests.

References
1. Lin YL, Zou JL, Yang W, Li CQ. A review of recent advances in research on $PM_{2.5}$ in China. Int J Env Res Pub He. 2018;15:438.
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. Atmos 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 Hlth. 2018;11:397–407.
4. Faridi S, Shamsipour M, Krzyzanowski M, Kunzli 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;54:7879–90.
6. de Jesus AL, 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. 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-Basel. 2021;12:590.
8. Jia BX, Gao M, Zhang XR, Xiao X, Zhang SQ, Yung KKL. Rapid increase in mortality attributable to $PM_{2.5}$ exposure in India over 1998–2015. Chemosphere. 2021;269:128715
9. Butz C, Liechti J, Bodin J, Cornell SE. Towards defining an environmental investment universe within planetary boundaries. Sustain Sci. 2018;13:1031–44.
10. Yu P, Xu RB, Coelho MSZS, Saldiva PHN, Li SS, 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. Kumar A, Ambade B, Sankar TK, Sethi SS, Kurwadkar S. Source identification and health risk assessment of atmospheric $PM_{2.5}$-bound polycyclic aromatic hydrocarbons in Jamshedpur, India. Sustain Cities Soc. 2020;52:101801.
12. WHO. WHO air quality guidelines for particulate matter, ozone, nitrogen dioxide and sulfur dioxide: global update 2005: summary of risk assessment. Geneva: World Health Organization; 2006.
13. WHO. WHO global air quality guidelines: particulate matter ($PM_{2.5}$ and PM_{10}), ozone, nitrogen dioxide, sulfur dioxide and carbon monoxide. Geneva: World Health Organization; 2021.
14. Cheng Z, Luo L, Wang SX, Wang YG, 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;8:517–31.
17. Perez P, Menares C. Forecasting of hourly $PM_{2.5}$ in south-west zone in Santiago de Chile. Aerosol Air Qual Res. 2018;18:2666–79.
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;17:122–31.
19. USEPA. 2006 National Ambient Air Quality Standards (NAAQS) for particulate matter ($PM_{2.5}$): final rule. Washington, DC: US Environmental Protection Agency; 2006.
20. Warner RA. Moving averages for identifying trends and changes in the data. In: Warner RA, editor. Optimizing the display and interpretation of data. Amsterdam: Elsevier; 2016. 53–73.
21. Goldin DQ, Kanellakis PC. On similarity queries for time-series data: Constraint specification and implementation. In: Montanari, U., Rossi, F. (eds) Principles and Practice of Constraint Programming — CP '95. CP 1995. Lecture Notes in Computer Science. Berlin: Springer; 1995;976. https://doi.org/10.1007/3-540-60299-2_9.
22. Fink E, Pratt KB. Indexing of compressed time series. Last M, Kandel A, Bunke H, editors. Data mining in time series databases, machine perception and artificial intelligence. Singapore: World Scientific; 2004. 43–65.
23. Cassisi C, Montalto P, Aliotta M, Cannata A, Pulvirenti A. Similarity measures and dimensionality reduction techniques for time series data mining. Karahoca A, editor. Advances in data mining knowledge discovery and applications. London: 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. Atmos Clim Sci. 2012;2:60–75.

25. Guo JM, Zhao MJ, Xue P, Liang X, Fan GT, Ding BH, 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 Env Res Pub He. 2021;18:9086.
27. Shumway RH, Stoffer DS. Time series analysis and its applications: with R examples. 4th ed. Berlin: Springer; 2017.
28. Chatfield C, Xing H. The analysis of time series: an introduction with R. 7th ed. Boca Raton: CRC Press; 2019.
29. Guilford JP. Fundamental statistics in psychology and education. New York: McGraw-Hill; 1965.
30. Zhao F, Gao YT, Li XN, An ZY, Ge SY, Zhang CM. A similarity measurement for time series and its application to the stock market. Expert Syst Appl. 2021;182:115217.
31. Kljun M, Tersek M, Strumbelj Erik. A review and comparison of time series similarity measures. In: 29th international electrotechnical and computer science conference. Portorose; 2020. https://link.springer.com/chapter/10.1007/3-540-60299-2_9#citeas.
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; 2013.
34. Dean RT, Dunsmuir WTM. Dangers and uses of cross-correlation in analyzing time series in perception, performance, movement, and neuroscience: the importance of constructing transfer function autoregressive models. Behav Res Methods. 2016;48:783–802.
35. Sooktawee S, Humphries U, Limsakul A, Wongwises P. Spatio-temporal variability of winter monsoon over the Indochina Peninsula. Atmosphere-Basel. 2014;5:101–21.
36. McGrath-Spangler EL, Denning AS. Global seasonal variations of midday planetary boundary layer depth from CALIPSO space-borne LIDAR. J Geophys Res-Atmos. 2013;118:1226–33.
37. 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.

Publisher's Note