Atmospheric Environment 192 (2018) 116–127

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

Day-of-week and seasonal patterns of $PM_{2.5}$ concentrations over the United States: Time-series analyses using the Prophet procedure

Naizhuo Zhao[a,b,*], Ying Liu[a,b], Jennifer K. Vanos[c,d], Guofeng Cao[a,b]

[a] Center for Geospatial Technology, Texas Tech University, Lubbock, TX, 79409, USA
[b] Department of Geosciences, Texas Tech University, Lubbock, TX, 79409, USA
[c] Scripps Institution of Oceanography and School of Medicine, University of California San Diego, La Jolla, CA, 92093, USA
[d] School of Sustainability, Arizona State University, Tempe, AZ, 85281, USA

ARTICLE INFO

Keywords:
$PM_{2.5}$ concentrations
Time series
Prophet
Seasonality

ABSTRACT

Fluctuations of ambient fine particulate matter ($PM_{2.5}$) concentrations show clear yearly and weekly patterns, which has been revealed by previous studies. However, reliability of those studies may be affected by their small research areas, short observation periods, and/or the lack of using specialized statistical approaches for time series. The current study applies a recently developed time-series analysis procedure, Prophet, to investigate seasonality of daily $PM_{2.5}$ concentrations over nine years (2007–2015) measured at 220 monitoring stations across the United States. Prophet is a new tool for producing high quality forecasts from time series data that have characteristics of multiple temporal patterns with either linear or non-linear growth/decline. Through decomposing each $PM_{2.5}$ time series into three major components (i.e., trend, seasonality, and holidays), we observed periodically changing patterns of $PM_{2.5}$ concentrations weekly and yearly consistent with previous findings. Specifically, relatively high $PM_{2.5}$ concentrations tend to appear in the month of January and on Fridays, and $PM_{2.5}$ concentrations on Sunday are generally lower than those on most other days of the week. However, we discovered that high $PM_{2.5}$ concentrations are also likely to appear in July. Additionally, compared to Fridays in most studies, the highest $PM_{2.5}$ concentrations are found to more likely occur on Saturdays, while the lowest concentrations are found on Monday as universally as on Sunday. Beyond understanding the seasonality of $PM_{2.5}$ concentrations, this study revealed the potential use of Prophet, originally designed for business time series, for detecting periodicities of environmental phenomena.

1. Introduction

Ground-level ambient fine particulate matter ($PM_{2.5}$) concentrations are sensitive to meteorological factors (e.g. air temperature and relative humidity) (Dawson et al., 2007; Wang and Ogawa, 2015) and human activities (e.g. industrial production and transportation) (Bao et al., 2016; Song et al., 2006), and tend to demonstrate day-of-week and seasonal variations (Russell et al., 2004; Lough et al., 2006). Understanding the temporal trends is of critical importance to accurately forecast $PM_{2.5}$ concentrations and to mitigate human $PM_{2.5}$ exposure. Many studies have explored the seasonal patterns of $PM_{2.5}$ (e.g. Gehrig and Buchmann, 2003; Motallebi et al., 2003; DeGaetano, and Doherty, 2004; Eiguren-Fernandez et al., 2004; Russell et al., 2004; Vecchi et al., 2004; Zheng et al., 2005; Lough et al., 2006; Zhao et al., 2009), yet most have apparent shortcomings due to either limited spatial or temporal factors.

First, although a few studies measure $PM_{2.5}$ concentrations at higher temporal resolution (e.g. 5 min intervals in Russell et al. (2004) and Zhao et al., 2009), the full temporal spans of the studies are not sufficiently long to reveal the long-term seasonality (mostly shorter than three years) (e.g. Eiguren-Fernandez et al., 2004; Russell et al., 2004; Vecchi et al., 2004; Zheng et al., 2005; Lough et al., 2006; Zhao et al., 2009; Wang et al., 2015; Wang et al., 2016; Zhang et al., 2015). Second, most of the studies often focus on a small geographic area, e.g. one city (Vecchi et al., 2004; Zheng et al., 2005; Song et al., 2006; Zhao et al., 2009; Wang et al., 2015, 2016; Zhang et al., 2015) or one state (Motallebi et al., 2003; Eiguren-Fernandez et al., 2004) and use *in situ* $PM_{2.5}$ concentrations from a limited number of monitoring sites in the given cities or states. However, it has been demonstrated that the influence of meteorological factors on $PM_{2.5}$ concentrations varies across different regions, either showing inconsistent or contradictory relationships by region (Tai et al., 2010; Liu et al., 2017). Thus,

* Corresponding author. Department of Geosciences, Texas Tech University, MS 1053, Science Building 125, Lubbock, TX, 79409-1053, USA.
E-mail address: zhao.naizhuo@gmail.com (N. Zhao).

https://doi.org/10.1016/j.atmosenv.2018.08.050
Received 8 April 2018; Received in revised form 23 August 2018; Accepted 27 August 2018
Available online 29 August 2018
1352-2310/

conclusions obtained from studies on small areas may not be representative of general patterns over larger geographic areas yet provide important information for the given city or state.

A few studies have been conducted to explore seasonal variations in $PM_{2.5}$ concentrations across large areas (e.g., in 187 counties of the United States (U.S.) by Bell et al. (2007) and in 31 Chinese cities by Xie et al. (2015)), and seasonal patterns in those studies were obtained by averaging $PM_{2.5}$ concentrations measured within the same season (or month). Thus, the last but perhaps most crucial shortcoming in many $PM_{2.5}$-seasonality studies is the lack of using advanced statistical methods to handle the time series $PM_{2.5}$ data. With the use of averaging procedures to process insufficiently long time-series data, important outliers such as those generated by fires, extreme weather, and gross errors are very likely to radically affect seasonal patterns of $PM_{2.5}$ concentrations (Chen and Liu, 1993).

The conventional time-series analysis methods (e.g., the commonly used ARIMA (Autoregressive Integrated Moving Average) model), usually work well under regularized trends but are prone to generate large errors when changes happen in the trends (Brockwell and Davis, 2016). Niu et al. (2016) and Wang et al. (2017) pointed out that such conventional linear time series approaches are insufficient to model complex changes of $PM_{2.5}$ concentrations. Most recently, Facebook developed a new time-series forecasting model, namely Prophet, and synchronously released a cognominal open source package (available both for R and Python) to facilitate implementation of the model (Taylor and Letham, 2017). Differing from traditional exponential smoothing models such as Holt Winters and ARIMA, the Prophet adopts a generalized additive model to fit the smoothing and forecasting functions. The fitting procedure of the Prophet is fast, allowing analysts to interactively and flexibly explore different model specifications. Besides the robust forecasting function, Prophet also excels at processing daily periodicity data with large outliers and shifts in trends and can model multiple periods of seasonality simultaneously.

The major objective of this study is therefore to explore the weekly and yearly seasonality of $PM_{2.5}$ concentrations over the U.S. using the Prophet time-series forecasting model. A daily $PM_{2.5}$ concentration dataset spanning nine years (2007–2015) was used to ensure sufficient cycles of seasonality in a time series analysis. These $PM_{2.5}$ concentrations were collected by *in situ* measurements from 220 monitoring stations that cover all geographic and climate zones in the U.S. (see Figs. 1 and 2 for station locations).

2. Data and methods

2.1. $PM_{2.5}$ data and preprocessing

The daily *in situ* $PM_{2.5}$ concentration records for 2007–2015 were retrieved from the U.S. Environmental Protection Agency (EPA) (http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/download_files.html, last access April 5, 2018). Each daily record is the average of all subdaily measurements taken at one station (United States Environmental Protection Agency, 2015). The records were collected by 2431 monitoring stations, yet a considerable portion of the stations do not provide continuous daily measurements for the nine years. Although the Prophet is robust at handling missing data (Taylor and Letham, 2017), lacking numerous daily records in one station is still likely to reduce accuracy of the time series analysis. Additionally, the $PM_{2.5}$ record contains an attribute of "event type", indicating whether data were measured when exceptional events (typically wildfires) were occurring (United States Environmental Protection Agency, 2015). Occurrence of the exceptional events usually results in abnormal $PM_{2.5}$ concentrations (Wegesser et al., 2009). Thus, any individual record flagged as "events included" was removed from the $PM_{2.5}$ time series and the slot corresponding to the removed record was re-labelled as "data missing". Eventually, we utilized only the time-series $PM_{2.5}$ records with at least five complete yearly cycles or data missing rates less than 11.11% (1/9). Accordingly, 220 of the 2431 monitoring stations throughout 45 states as well as the District of Columbia, were used for the analysis (See Figs. 1 and 2). Following Liu et al. (2017), we divided the daily records from the 220 monitoring stations into four seasons (i.e., spring: March, April, May; summer: June, July, August; fall: September, October, November; and winter: December, January, February) and averaged the daily $PM_{2.5}$ concentrations for each day of the week. Fig. 3 clearly illustrates that $PM_{2.5}$ concentrations are higher in winter and summer while lower in spring and fall. Additionally, $PM_{2.5}$ concentrations are the highest on Saturday and Friday and the lowest on Sunday and Monday in each of the seasons except fall. This preliminary assessment reveals the possible day-of-week and seasonal effects in $PM_{2.5}$ concentrations in the U.S., yet more detailed and reliable analyses are warranted using a statistical time-series analysis procedure.

2.2. Seasonal decomposition

Prophet was originally designed to smooth and forecast daily business data and fully considers three basic features of business time series, i.e. piecewise trends, multiple seasonality, and floating holidays. Thus, a time series can be decomposed by the Prophet into three major components of trend, seasonality, and holidays, as well as an error term (Taylor and Letham, 2017). Previous studies have demonstrated that $PM_{2.5}$ concentrations significantly increase during holidays (e.g. Thanksgiving, Christmas, and New Year) (Motallebi, 1999; Gorin et al., 2006; Feng et al., 2012). Hence, it is necessary to set up an individual component for holidays when studying seasonal patterns of $PM_{2.5}$ concentrations. In the current study, Thanksgiving, Christmas, and New Year's Day were set as holidays, as well as the two days before and after each holiday, which provide predictable 'spikes' in $PM_{2.5}$ concentrations that do not follow either of the common periodic (day-of-week or seasonal) patterns (Motallebi, 1999; Gorin et al., 2006; Feng et al., 2012).

Prophet adopts a Bayesian-based curve fitting method to smooth and forecast time-series data, which is the most distinct feature of Prophet as compared to the traditional time-series forecasting models such as Holt Winters, ARIMA, among others. In other words, the Prophet utilizes several linear and/or non-linear functions to fit the decomposed components, with time as the only regressor. Compared to traditional exponential smoothing models, Prophet can more easily handle temporal patterns with multiple periods and has no requirements on regularly spaced measurements (Taylor and Letham, 2017). Given that missing $PM_{2.5}$ values are prevalent in the U.S. EPA's dataset, these features make Prophet particularly suitable to reveal the weekly and yearly seasonality simultaneously from the daily time series $PM_{2.5}$ concentration records.

The periodic effects in Prophet are modeled by Fourier series (Harvey and Shephard, 1993):

$$s(t) = \sum_{n=1}^{N}\left[a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right)\right] \tag{1}$$

where P represents the regular period that the time series is expected to have (e.g. 365.25 and 7 for yearly and weekly periods, respectively (Taylor and Letham, 2017)) and t denotes a date. Thus, the basis of fitting the seasonal model is to determine the value of N and then determine the $2 \times N$ parameters, i.e. $\beta = [a_1, b_1; \ldots; a_N, b_N]^{\mathrm{T}}$. Shortening the time series at N can be deemed as a low-pass filter. Hence, large values of N are appropriate to fit seasonal patterns with quick changes, yet an over-large N is prone to lead to overfitting (Taylor and Letham, 2017). In this study, we utilized the default values of the Prophet for N (i.e. $N = 10$ and $N = 3$ for modelling yearly and weekly components, respectively) that were empirically demonstrated to work well for most practical issues that follow yearly and weekly patterns (Taylor and Letham, 2017). Equation (1) can then be adapted to form Equations (2) and (3), and the $2 \times N$ parameters (i.e., β) can be fitted by L-BFGS

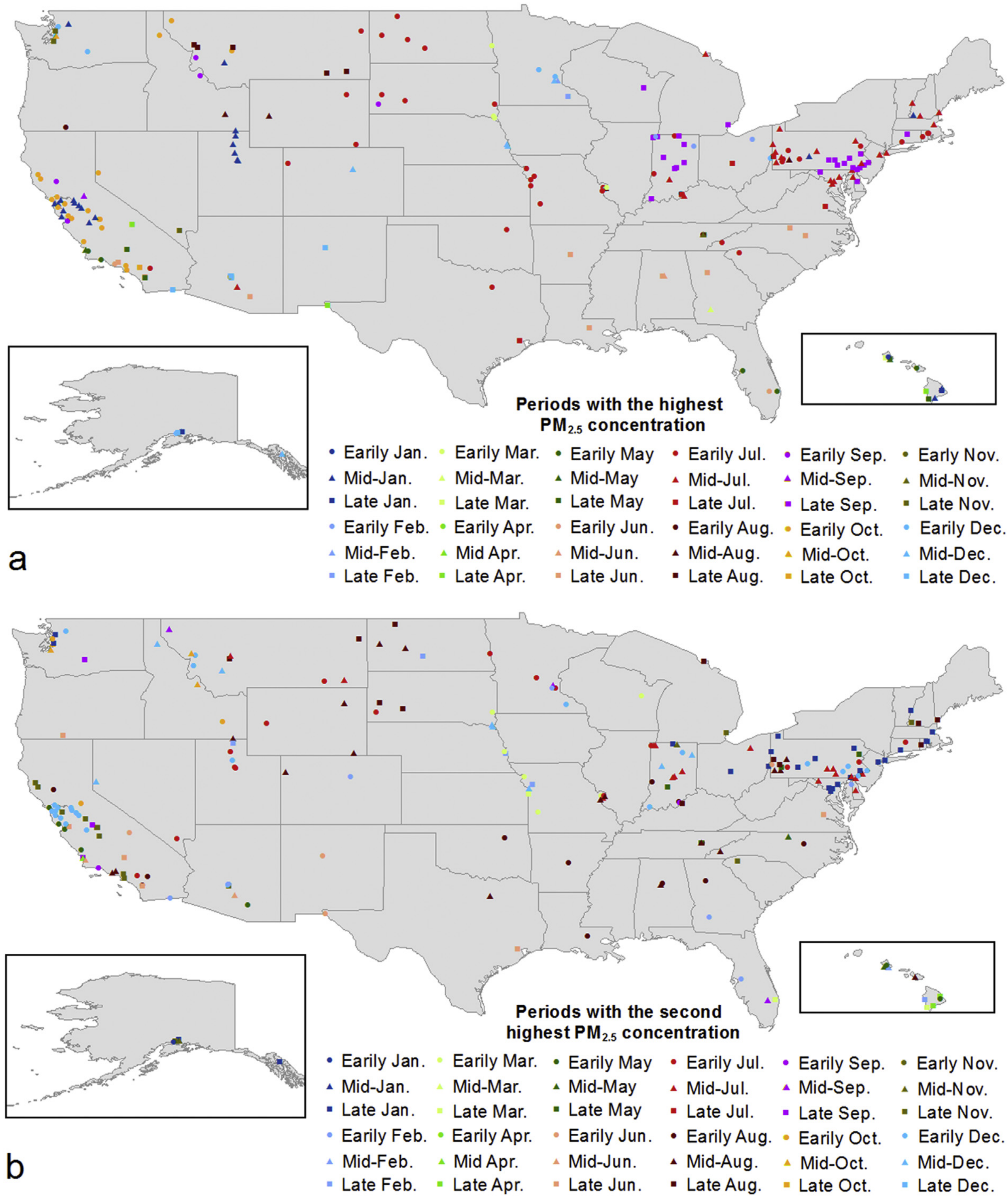

Fig. 1. The periods with the highest (*a*), the second highest (*b*), the lowest (*c*), and the second lowest (*d*) $PM_{2.5}$ concentrations at 220 monitoring stations. Time of season for each concentration maxima or minima mapped for each station are listed in Fig. 4.

(Limited-memory Broyden-Fletcher-Goldfarb-Shanno) (Byrd et al., 1995):

$$s(t) = X(t)\beta \tag{2}$$

$$X(t) = \begin{cases} \left[\cos\left(\frac{2\pi(1)t}{365.25}\right),\ \sin\left(\frac{2\pi(1)t}{365.25}\right),\ \ldots,\cos\left(\frac{2\pi(10)t}{365.25}\right),\ \sin\left(\frac{2\pi(10)t}{365.25}\right)\right] & \textit{for yearly} \\ \left[\cos\left(\frac{2\pi(1)t}{7}\right),\ \sin\left(\frac{2\pi(1)t}{7}\right),\ \ldots,\cos\left(\frac{2\pi(3)t}{7}\right),\ \sin\left(\frac{2\pi(3)t}{7}\right)\right] & \textit{for weekly} \end{cases} \tag{3}$$

In the current study, we focused on investigating yearly and weekly variation in $PM_{2.5}$ concentrations and thus skipped formulation details on the trend and holiday components. Those details have been elaborately exhibited in the paper of Taylor and Letham (2017).

2.3. Analyses on the decomposed seasonal component

The yearly and weekly seasonality of $PM_{2.5}$ concentrations at each of the 220 monitoring stations are described by two curves. We divided each month into three phases, i.e. early (the 1st day to the 10th day),

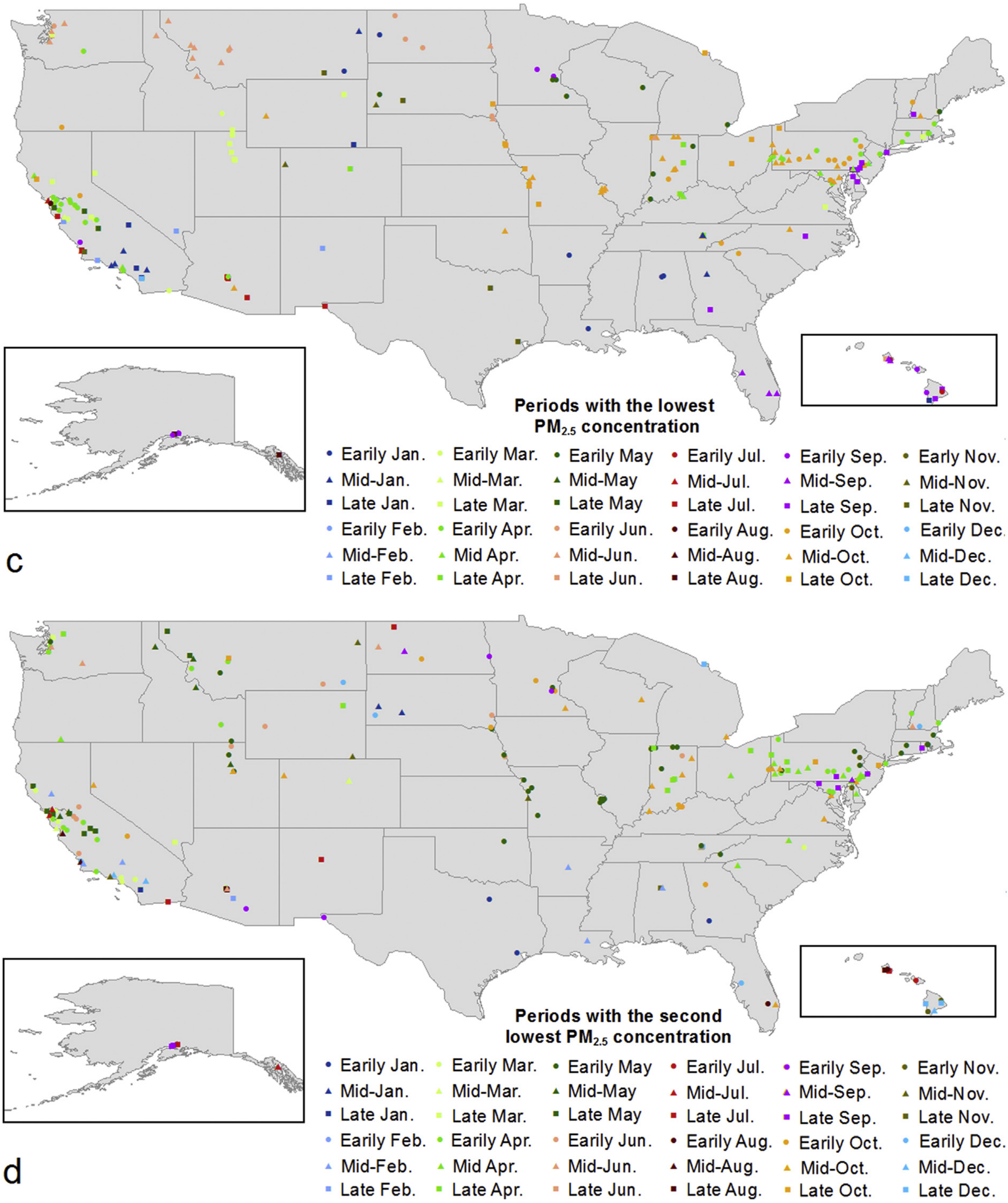

Fig. 1. (*continued*)

middle (the 11th day to the 20th day), and late (the 21st day to the last day of the month), and thus there are 36 (3 × 12) periods (e.g. early January, mid-June, and late November) for the one-year cycle. From each of the day-of-year curves, we labelled all dates with local maximum and minimum $PM_{2.5}$ concentrations. If two or more local maximums (minimums) are located in one period, we only retained the first maximum (minimum) and deleted the others. Next, for each of the monitoring stations we ordered the dates by their $PM_{2.5}$ concentrations and selected four periods, including the dates with the two highest and the two lowest $PM_{2.5}$ concentrations. Finally, we summed the number of monitoring stations with the highest and lowest $PM_{2.5}$ concentrations for each of the 36 periods (see Fig. 4). Similarly, based on each of the day-of-week curves, we ordered the days by their $PM_{2.5}$ concentrations and determined the four days with the two highest and the two lowest $PM_{2.5}$ concentrations for each of the monitoring stations. We then summed the number of monitoring stations with the highest and lowest $PM_{2.5}$ concentrations for each of the seven days (see Fig. 5).

Chi-Square Goodness-of-Fit tests were performed to investigate whether the highest and the lowest $PM_{2.5}$ concentrations unevenly appear within particular periods/seasons of the year and day of week. The specific null hypothesis is that the highest (or the lowest) $PM_{2.5}$ concentrations equally appear in different periods or days among the

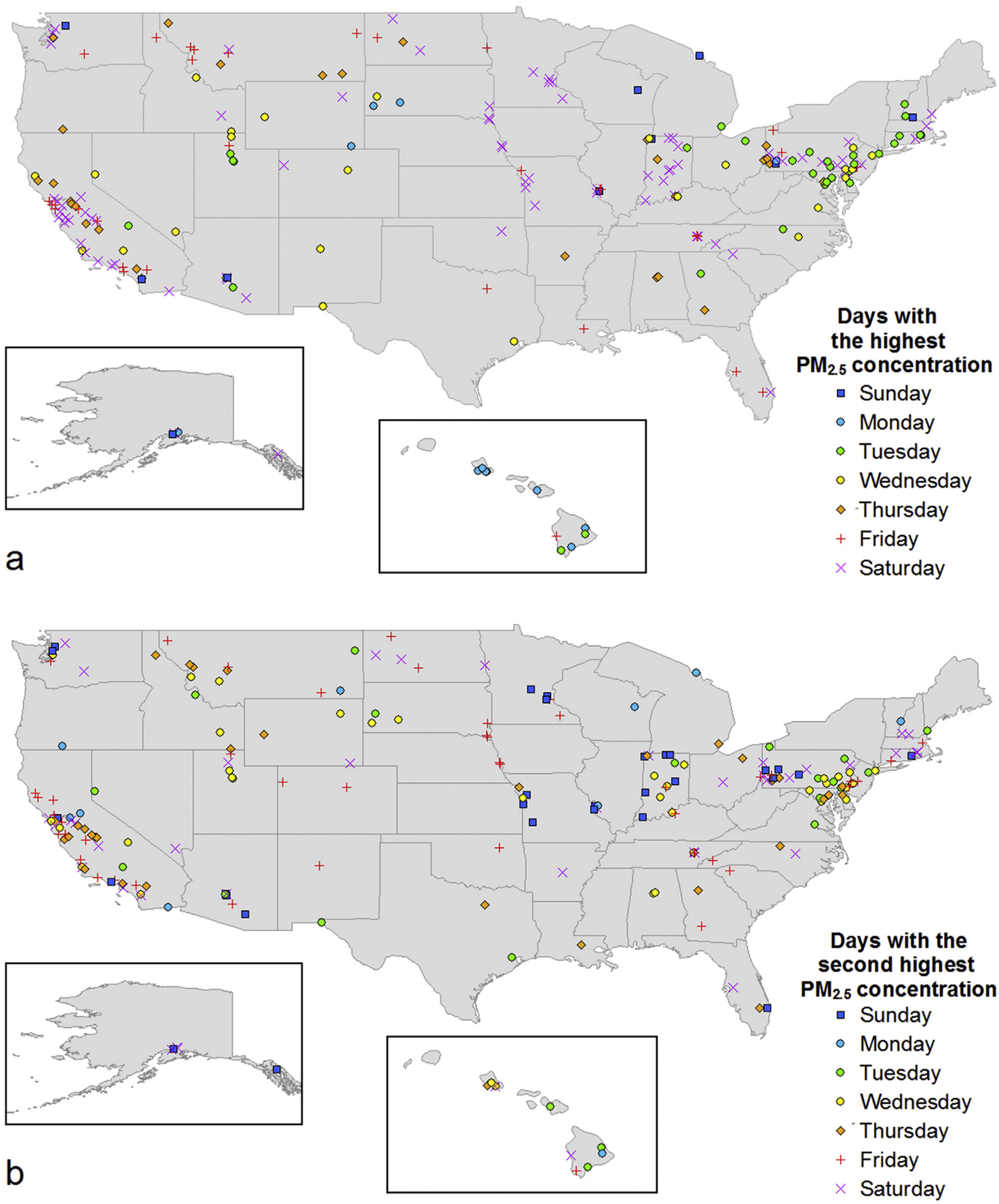

Fig. 2. The day-of-week with the highest (a), the second highest (b), the lowest (c), and the second lowest (d) $PM_{2.5}$ concentrations at 220 monitoring stations. The total number of stations at each concentration maxima or minima mapped are listed in Fig. 5.

220 monitoring locations.

A Durbin-Watson test was conducted to examine whether residuals (i.e. the error term) are auto-correlated for each of the 220 time series. The null hypothesis of the Durbin-Watson test is that the error term in one time stamp is independent with the preceding error term. Serially dependent residuals usually suggest an inadequate time series model (StatSoft, 2011). Thus, the Durbin-Watson test can inform whether the Prophet adequately decomposed the $PM_{2.5}$-concentration time series to obtain reliable seasonality.

A number of previous studies (e.g. Bagan, and Yamagata, 2015; Dobson et al., 2000; Hardin et al., 2018; Sutton et al., 2001; Sutton et al., 2010) have demonstrated that brightness of nighttime lights (NTL) observed by satellites is a good indicator of population density and can be used to delimit urban extents. In general, a region with brighter NTL usually has a larger population and is located closer to an urban center (Bagan and Yamagata, 2015). Thus, we extracted brightness of NTL at each of the 220 monitoring stations from the latest NTL image product, i.e. the version 1 Visible Infrared Imaging Radiometer Suite (VIIRS) Day/Night Band (BND) image composite for 2015 (Elvidge et al., 2017) (see Figure S1 for the NTL image product). Four

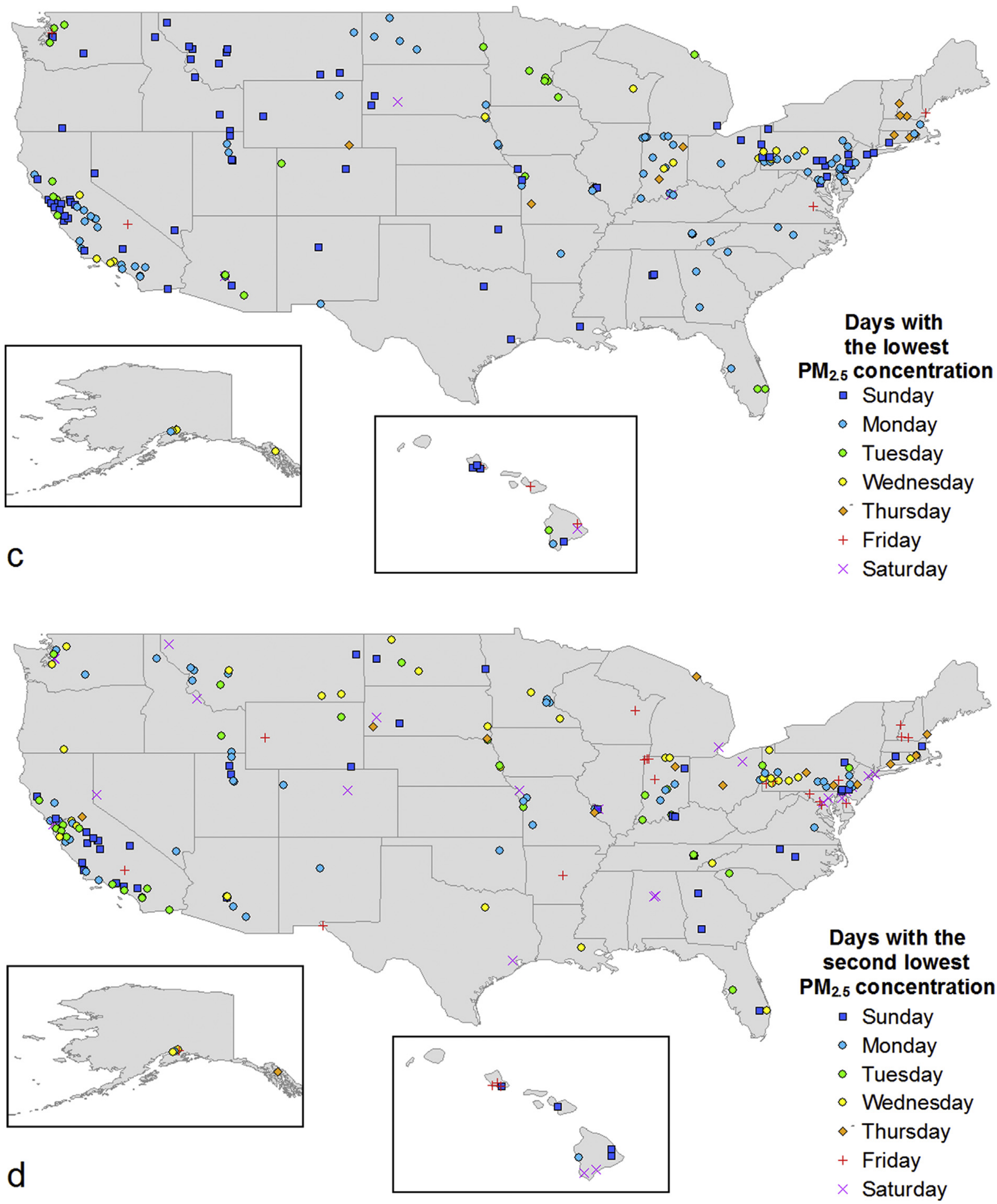

Fig. 2. (*continued*)

one-way ANOVA tests were conducted to examine if stations with the highest or the lowest $PM_{2.5}$ concentrations during a given period of the year or a day of week correspond to significantly different brightness of NTL. These differences in brightness of NTL imply that the seasonal patterns of the $PM_{2.5}$ concentration are influenced by geographic variation present between populated urban areas versus sparsely populated rural areas. Specifically, we divided a calendar year into five periods: 1) July and August and 2) late November to late January when the highest $PM_{2.5}$ concentrations are likely to appear, 3) late March to early May and 4) late September to late October when the lowest $PM_{2.5}$ concentrations tend to appear, and 5) the remaining dates (see Fig. 4). We then performed two ANOVA tests, with the null hypothesis that brightness of NTL is equal among the five groups of monitoring stations with the highest or lowest $PM_{2.5}$ concentrations appearing in the five periods. Similarly, another two ANOVA tests were performed with the null hypothesis that brightness of NTL is equal among the five groups of monitoring stations with the largest or lowest $PM_{2.5}$ concentrations appearing in seven different days of week.

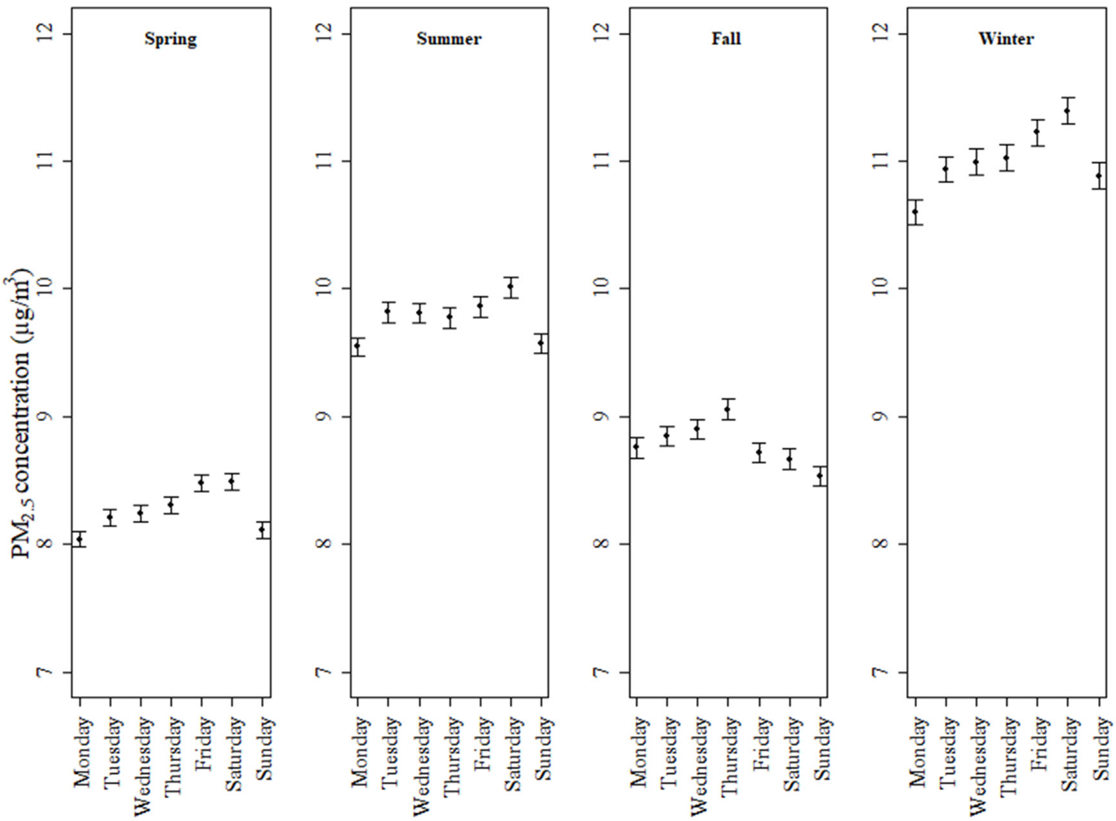

Fig. 3. Average $PM_{2.5}$ concentrations and confidence intervals (at the 95% confidence level) of the 220 monitoring stations for the same days of the week in four seasons.

3. Results

3.1. Decomposed components

Decomposed components of $PM_{2.5}$ time series for the 220 stations are available from GitHub (https://github.com/thestarlab/pm25-time-series). Of the 220 Durbin-Watson tests, zero was rejected at the 0.01 level, suggesting that the Prophet adequately decomposed the $PM_{2.5}$ concentration time series and the changing patterns observed from the decomposed seasonal components are reliable (StatSoft, 2011). Fig. 6 shows a $PM_{2.5}$ time series at a monitoring station in Tulsa, OK (N 36.2049°, W 95.9765°), which was randomly selected from the 220 stations for example purposes. From the original $PM_{2.5}$ concentration time series, large seasonal fluctuations and a general decreasing trend can be discerned even though the trend is not fully clear. Fig. 6 also exhibits the estimated $PM_{2.5}$ concentrations by the Prophet model. An estimated $PM_{2.5}$ concentration is the sum of values of the trend, holiday, weekly, and yearly components at a given time. These components are shown by Fig. 7, from which an apparently decreasing trend in $PM_{2.5}$ concentrations can be observed over the nine years. Moreover, a change happened in 2011, displaying that the decreasing rate from 2007 to 2011 is smaller than that from 2011 to 2016. The holiday component indicates that $PM_{2.5}$ concentrations at this site are exceptionally larger at Christmas and New Year's than usual. The day-of-week curve exhibits the $PM_{2.5}$ concentration to steadily increase from Monday to Saturday and sharply fall from Saturday to Sunday, followed by a slight decrease to reach the lowest levels on Monday. The day-of-year curve shows three relatively large local maximums in early April, early July, and mid-December and three apparent local minima in early January, early May, and mid-October. The above results for Tulsa, OK substantiate that within the varied trends, significant increases occur during the holidays, and multiple seasonality exist in the long-term $PM_{2.5}$ time series. Thus, the traditional exponential smoothing models (e.g. ARIMA) are not fairly suitable to tease out such complex and nonlinear time series data (Niu et al., 2016; Wang et al., 2017).

3.2. Characteristics of the yearly and weekly seasonality

The null hypotheses of all the Chi-Square Goodness-of-Fit tests in this study were rejected at the 0.01 level, indicating that the highest and the lowest $PM_{2.5}$ concentrations unevenly appear within different periods and on different days. Figs. 1 and 3 shows that the highest $PM_{2.5}$ concentrations are more likely to appear in January and July, on average across the U.S. The second highest $PM_{2.5}$ concentrations also tend to appear in the same two months, as well as August, late November, and early December. In short, during a yearly cycle the highest $PM_{2.5}$ concentrations are more likely to appear in wintertime and summertime. On the contrary, the lowest $PM_{2.5}$ concentrations tend to appear in the spring (from late March to early May) and the fall (from late September to late October) (see Figs. 2 and 4).

California, New York, and Massachusetts have a relatively large number of $PM_{2.5}$ monitoring stations with relatively complete daily $PM_{2.5}$ measurements compared with the other states. It can be seen from Fig. 1*a* and b (1*c* and 1d) that the phenomenon of the highest (lowest) $PM_{2.5}$ concentrations in the winter or summer (fall or spring) exists not only in the three mentioned states, but in the majority of the remaining states as well. We observed that 33% of the monitoring stations from 18 states presented the highest $PM_{2.5}$ concentrations in January. Further, 40% of the stations from 25 states showed the second highest concentrations between late November and early February. We also found that the highest $PM_{2.5}$ concentrations appear in the summer period from June 21st to July 20th at 30% of the sites within 27 states. Additionally, 76 monitoring stations from 30 states presented the

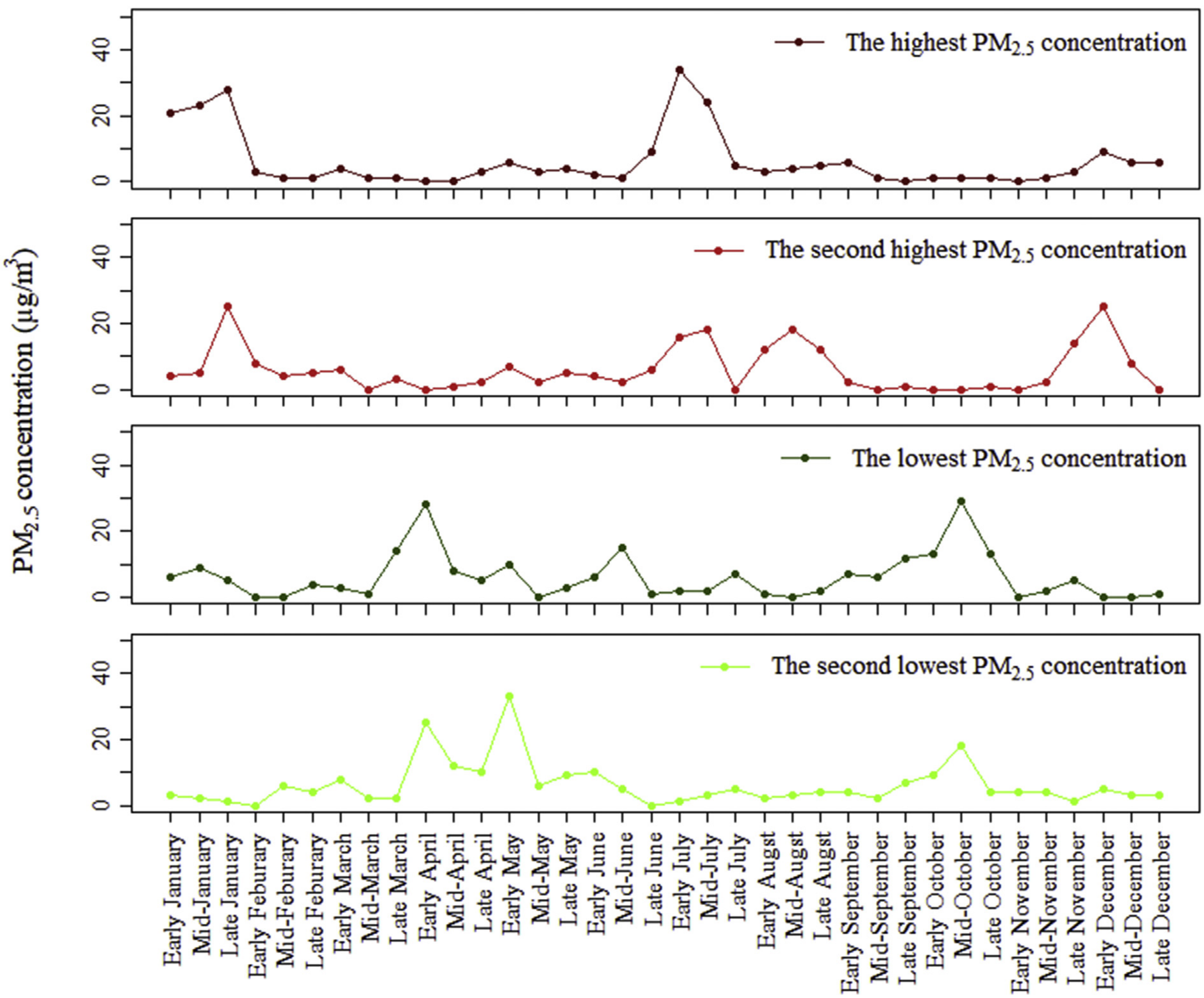

Fig. 4. The number of stations with the highest or the lowest $PM_{2.5}$ concentration appearing during a given period of the year.

second highest $PM_{2.5}$ concentrations in the hottest two months of the U.S. (i.e. July and August) (see Figs. 1 and 3). On the contrary, 37% (30%) of the monitoring stations from 18 (24) states presented the lowest (second lowest) $PM_{2.5}$ concentrations in the spring (i.e., from late March to early May) and 17% (30%) of the sites from 20 (18) states presented the lowest (second lowest) $PM_{2.5}$ concentrations in the fall (i.e., from late September to late October). Thus, the yearly seasonality (i.e., high $PM_{2.5}$ concentrations in winter and summer and low $PM_{2.5}$ concentrations in spring and fall) is a widespread pattern but not a regional phenomenon.

It can be seen from Fig. 4 that the highest $PM_{2.5}$ concentrations are most likely to appear on Saturday in a weekly cycle, and most unlikely to appear on Monday and Sunday. Conversely, the lowest $PM_{2.5}$ concentrations are most likely to appear on Monday and Sunday, and most unlikely to appear in Saturday and Friday. Similar with the yearly seasonality, the above day-of-week changing patterns extensively exist across different geographic regions of the U.S. (see Fig. 2). Specifically, we observed that 35% of the monitoring stations from 23 states presented the highest $PM_{2.5}$ concentrations on Saturday while only 16% of the monitoring stations from 14 states presented the highest $PM_{2.5}$ concentrations on Friday. We also found that the highest $PM_{2.5}$ concentrations appear on Monday at 38% of the sites from 23 states while occur on Sunday at nearly the same percentage (35%) of the sites from 26 states.

4. Discussion

4.1. Comparison with the previous findings

The highest $PM_{2.5}$ concentrations are observed during wintertime in many previous studies (e.g. Gehrig and Buchmann, 2003; Eiguren-Fernandez et al., 2004; Zheng et al., 2009; Gu et al., 2011; Xu et al., 2012). The large increase in wood and coal combustion for heating is the major reason leading to the appearance of the highest $PM_{2.5}$ concentrations in wintertime in many regions (Gehrig and Buchmann, 2003; Gorin et al., 2006; Zheng et al., 2009; Gu et al., 2011). In the current study, day-of-year curves for the 220 monitoring stations showed that 81% of the 220 curves have, at least, two peak values, one in the wintertime (i.e. December and January) and the second in the summertime (i.e. July and August). Differing from the major reason resulting in the high $PM_{2.5}$ concentrations in wintertime, the high $PM_{2.5}$ concentrations in summertime are often associated with days or periods of enhanced temperature — as shown by (Tai et al., 2010; Liu et al., 2017; Vanos et al., 2015) — that often pair with high pressure, low winds, and stagnant air (e.g., hot dry air masses), thus trapping in particulate matter and other pollutants (Greene et al., 1999; Dixon et al., 2016). Additionally, the large increases in tourists and consequently uses of motor vehicles in the summertime are also likely to contribute to the high $PM_{2.5}$ concentrations (Zheng et al., 2005; Martellini et al., 2012).

Previous studies nearly consistently reported that the weekly maximal $PM_{2.5}$ concentrations appear in the late of workweek (i.e. Friday)

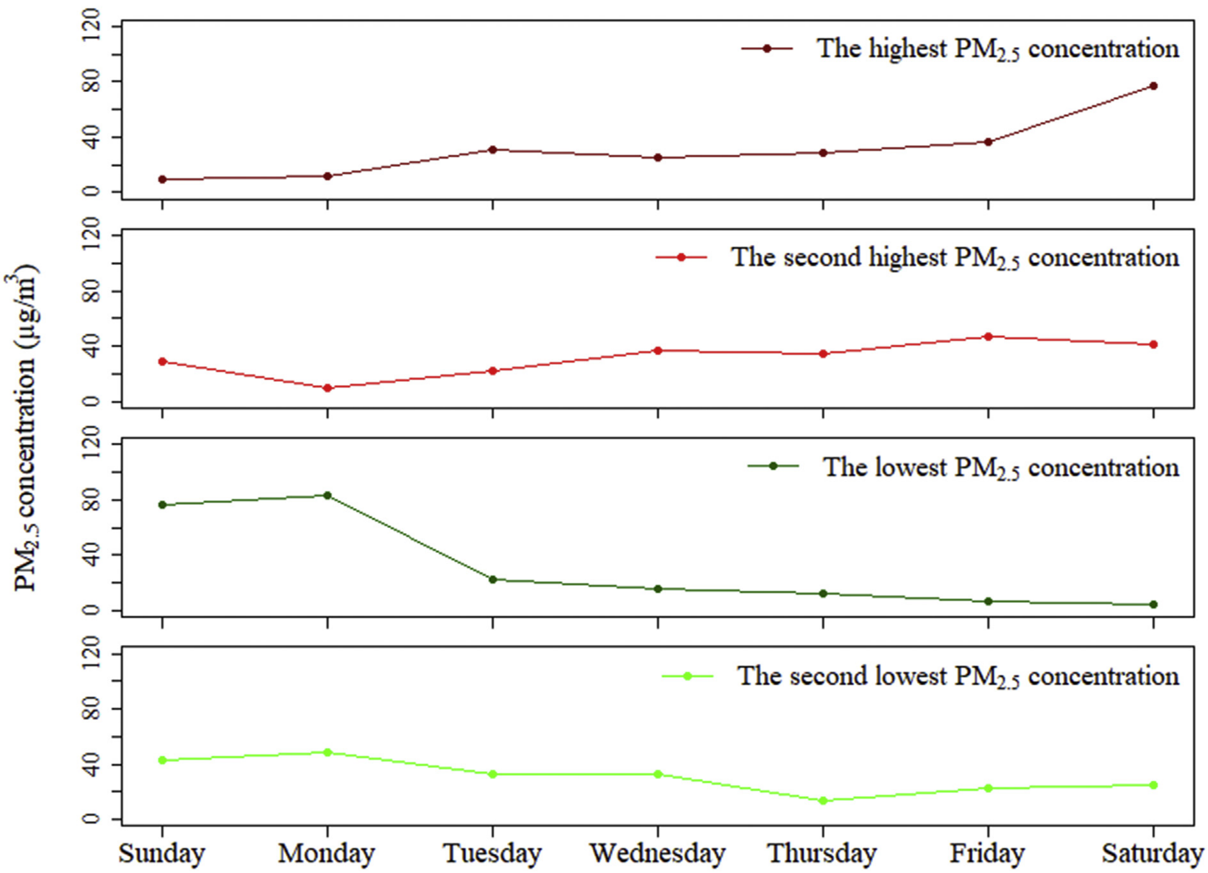

Fig. 5. The number of stations with the highest or the lowest $PM_{2.5}$ concentration appearing on a day.

and the minimum occurs on weekends (especially Sunday) (Motallebi et al., 2003; Lough et al., 2006). In urban areas, motor vehicles are the primary emission sources of $PM_{2.5}$ (Pongpiachan et al., 2015). The decrease in the use of motor vehicle and especially heavy-duty trucks is likely the major reason leading to the relatively low $PM_{2.5}$ levels on weekends in the current study as well as others (Motallebi et al., 2003). However, the current study shows that at more selected stations the highest $PM_{2.5}$ concentrations occur on Saturday and the lowest $PM_{2.5}$ concentrations can be found on both Monday and Sunday (see Fig. 4). Previous studies (e.g. Huang et al., 2014; Gentner et al., 2017) have demonstrated that contributions of the secondary organic and inorganic aerosol to total $PM_{2.5}$ concentration are larger than those of primary particulate emissions. Formation and accumulation/diffusion of the secondary $PM_{2.5}$ pollutants are complex chemical reactions in the atmosphere and are affected by many meteorological factors (e.g. wind speed and relative humidity) respectively, thus need certain time period to form (Gao et al., 2015). The previous studies did not demonstrate a clear and consistent time of forming the secondary $PM_{2.5}$ pollutants. The new findings in this study suggest that a one-day lag extensively exists between effects of human activities and changes in total $PM_{2.5}$ concentrations in the U.S. During a weekly cycle, accumulation of the primary particulate emissions reaches maximum on Friday, yet the

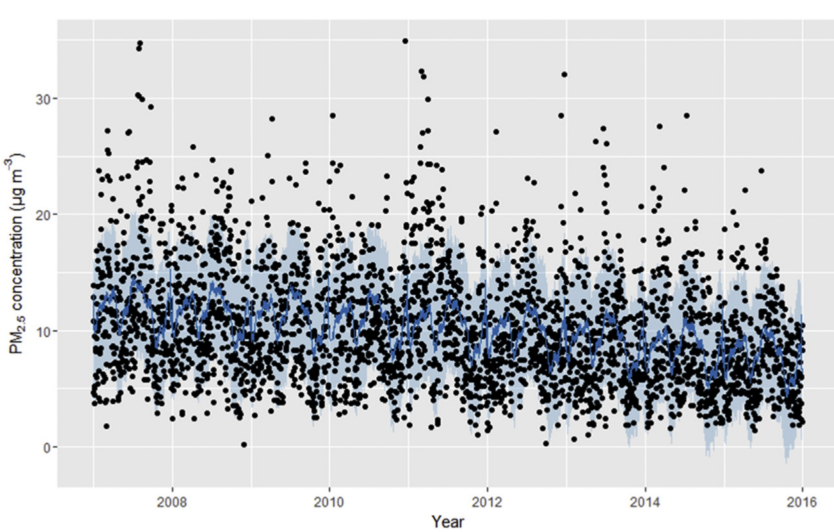

Fig. 6. The original $PM_{2.5}$-concentration time series (black points) collected from a monitoring station in Tulsa, OK and the means of estimated $PM_{2.5}$-concentration time series (dark blue line) at the 95% confidence interval (light blue area). The estimated $PM_{2.5}$ concentration is the sum of values of the trend, holiday, weekly, and yearly components at a given time. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

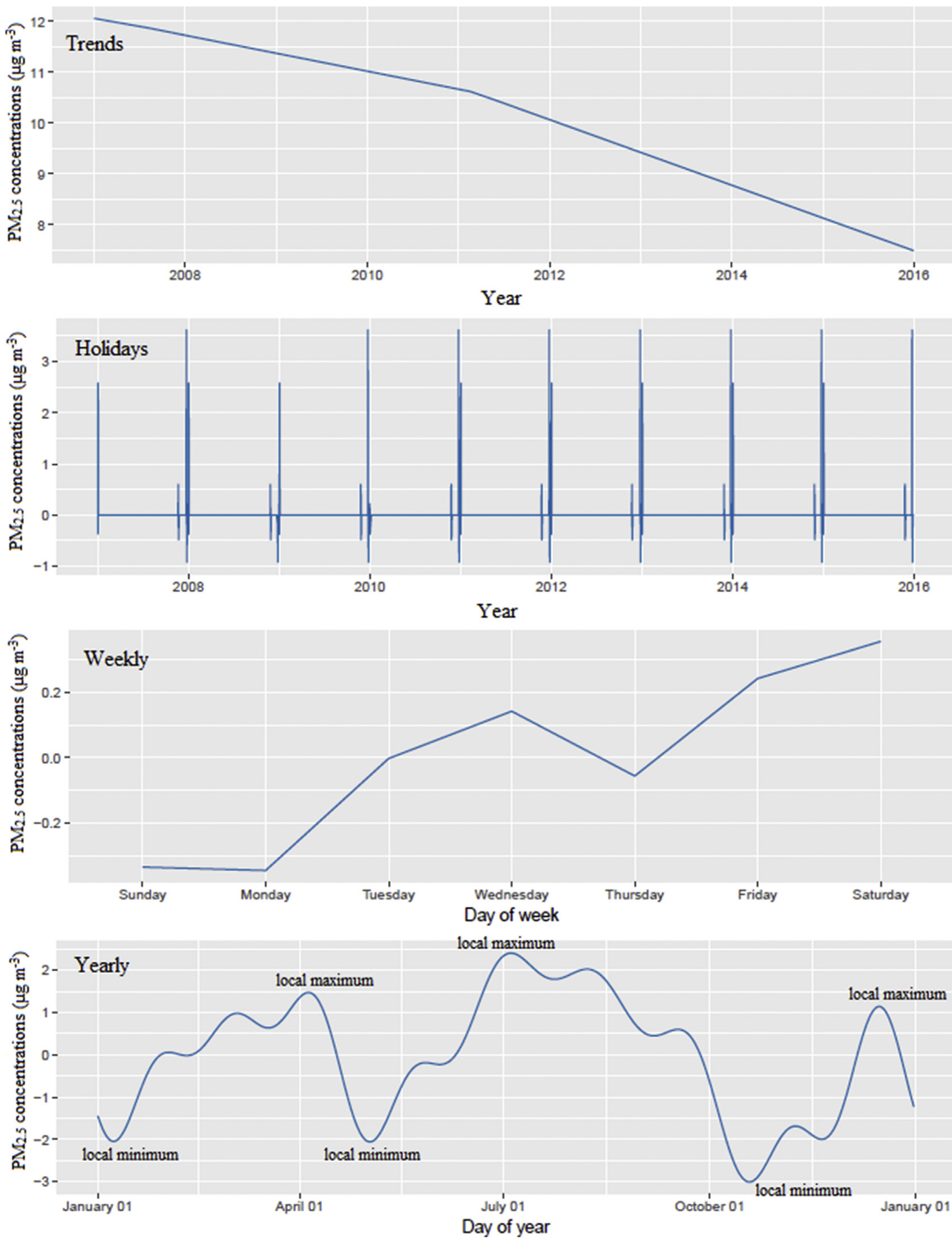

Fig. 7. The decomposed components of a randomly selected $PM_{2.5}$-concentration time series for Tulsa, OK. Note that for the day-of-year curve, we only labelled the most prominent (but not all) local maximums and minimum for esthetical reasons.

secondary $PM_{2.5}$ pollutants contribute to the highest levels on Saturday. On Sunday, the number of human's industrial production and travel activities reaches the lowest (Motallebi et al., 2003). After a night's diffusion and absorption, the concentration of $PM_{2.5}$ in the early morning of Monday are lower than that in the day of Sunday (DeGaetano and Doherty, 2004). As stated in the Data and Methods section, each daily record used for the time-series analysis is the average of all sub-daily measurements taken at one station. Therefore, even though human's industrial production and travel are restarted in the daytime of Monday, most secondary $PM_{2.5}$ pollutants are not formed immediately which leads to the average $PM_{2.5}$ concentration of Monday is still lower than that of Sunday at a considerable number of monitoring stations.

4.2. Spatial patterns of the $PM_{2.5}$ seasonality

Figures S1 and S2 shows that the 220 selected monitoring stations are scattered in rural (brightness of NTL: $< 1\ \mathrm{nW\,cm^{-2}\,sr^{-1}}$), suburban (brightness of NTL: $\geq 1\ \mathrm{nW\,cm^{-2}\,sr^{-1}}$ and $\leq 30\ \mathrm{nW\,cm^{-2}\,sr^{-1}}$), urban (brightness of NTL: $\geq 31\ \mathrm{nW\,cm^{-2}\,sr^{-1}}$ and $\leq 100\ \mathrm{nW\,cm^{-2}\,sr^{-1}}$), and urban core (brightness of NTL: $> 100\ \mathrm{nW\,cm^{-2}\,sr^{-1}}$) areas of the U.S. (Sutton et al., 2010). However, the varied locations of the monitoring stations at the metropolitan scale (i.e., from rural to urban core areas) do not generate significant influences on the seasonal or day-of-week changing patterns of the $PM_{2.5}$ concentration as no ANOVA test was rejected at the 0.01 level.

The weekly changes of $PM_{2.5}$ concentrations are mainly affected by the human activities (e.g. industrial production and traffic) (Motallebi et al., 2003; Pongpiachan et al., 2015), and human activities (e.g. working in workdays and resting in weekends) are nearly the same across the U.S. By contrast, the yearly fluctuations of $PM_{2.5}$ concentrations are considerably impacted by meteorological factors (Tai et al., 2010; Liu et al., 2017). Hence, at the larger geographic scale, varied climatic situations are likely to lead to the highest $PM_{2.5}$ concentrations occurring in different periods across the U.S.

In this study, the 220 monitoring stations throughout 45 states as well as the District of Columbia. Within the 45 states, California has the most (39) monitoring stations. The influence of the Pacific Ocean moderates temperature extremes, resulting in warmer winters and relatively cooler summers in the coastal areas of California. The relatively small increases in temperature lead to relatively small increases in $PM_{2.5}$ concentrations in summertime (Liu et al., 2017). Thus, it can be seen from Fig. 1 that at most of the stations located in California, especially in northern California, the highest $PM_{2.5}$ concentrations occur in January and the second highest $PM_{2.5}$ concentrations also appear in the cold periods (e.g. December or late November), but the highest $PM_{2.5}$ concentrations are rare to appear in summertime. However, at a few monitoring stations located in Los Angeles or in National Parks of central and southern California (e.g. Joshua Tree National Park and Death Valley National Park), the highest and/or the second highest $PM_{2.5}$ concentrations appear in the periods from June to August (see Fig. 1*a* and b). The large increases in tourists and consequently motor vehicles in summer are likely to be a reason leading to the occurrence of the highest $PM_{2.5}$ concentrations in the summertime. Additionally, southern California's winters are warm and thus it does not need to burn much wood for heating in winters. This may be another reason for the highest $PM_{2.5}$ concentrations in some of the southern California stations appearing in the summertime but not wintertime. As stated in the methods, records with exceptional events were removed from the time series; thus, wildfires and dust mainly brought by the Santa Ana winds (Viswanathan et al., 2006; Wu et al., 2006) cannot significantly affect the $PM_{2.5}$ seasonality revealed in the current study. In brief, the above spatial patterns suggest that the yearly seasonality of $PM_{2.5}$ concentrations is considerably affected by both natural forces and anthropologic activities.

5. Conclusions

The simultaneous analyses on the day-of-week and seasonal patterns of $PM_{2.5}$ concentrations across a larger spatiotemporal (nine years and 220 sites over the U.S.) scope are some distinct advantages of the current study as compared to previous similar studies. In particular, the use of the Prophet mitigates adverse impacts of common problems in time-series data (e.g. data missing, changed trends, and unexpected outliers) on analyses of periodicities. However, the large study scale limits the detailed investigation of periodic changes of individual $PM_{2.5}$ components (e.g. organic carbon, elemental carbon, and oxides of nitrogen) at each site. This is because of the 2431 $PM_{2.5}$ monitoring stations of the U.S. EPA, only less than 10% of the stations can provide continued daily measurements on total $PM_{2.5}$ concentrations for our given study period (i.e. 2007–2015). Even fewer stations can provide complete and continued daily measurements of individual $PM_{2.5}$ components. Thus, if we further advance investigation on seasonality of individual $PM_{2.5}$ components, many geographic or climate zones will not have any qualified $PM_{2.5}$ time-series data.

Through analyzing the $PM_{2.5}$ time series at 220 stations over the U.S., we obtained novel findings, whereby in addition to high wintertime concentrations, high $PM_{2.5}$ concentrations were shown to also reach yearly maximums in summertime across the U.S. Additionally, compared to Friday, the weekly $PM_{2.5}$ maximum is more likely to appear on Saturday, while the minimum is likely to appear on not only Sunday but also Monday. Beyond the findings, this study substantiates the feasibility and advantages of the Prophet, which was originally developed for business time series, in detecting periodicities of environmental phenomena. In the future, we will further use the Prophet to forecast concentrations of air pollutants and compare forecasting accuracy with currently prevalent deep-learning methods such as the long short-term memory recurrent neural network (e.g., Sak et al., 2014).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.atmosenv.2018.08.050.

References

Bagan, H., Yamagata, Y., 2015. Analysis of urban growth and estimating population density using satellite images of nighttime lights and land-use and population data. GIScience Remote Sens. 52, 765–780.

Bao, C., Chai, P., Lin, H., Zhang, Z., Ye, Z., Gu, M., Lu, H., Shen, P., Jin, M., Wang, J., Chen, L., 2016. Association of PM2.5 pollution with the pattern of human activity: a case study of a developed city in eastern China. J. Air Waste Manag. Assoc. 66, 1202–1213.

Bell, M.L., Dominici, F., Ebisu, K., Zeger, S.L., Samet, J.M., 2007. Spatial and temporal variation in $PM_{2.5}$ chemical composition in the United States for health effects studies. Environ. Health Perspect. 115 (7), 989–995.

Brockwell, P.J., Davis, R.A., 2016. Introduction to Time Series and Forecasting, third ed. Springer Texts in Statistics.

Byrd, R.H., Lu, P., Nocedal, J., 1995. A limited memory algorithm for bound constrained optimization. SIAM J. Sci. Stat. Comput. 16, 1190–1208.

Chen, C., Liu, L.-M., 1993. Joint estimation of model parameters and outlier effects in time series. J. Am. Stat. Assoc. 88, 284–297.

Dawson, J.P., Adams, P.J., Pandis, S.N., 2007. Sensitivity of PM2.5 to climate in the Eastern US: a modeling case study. Atmos. Chem. Phys. 7, 4295–4309.

DeGaetano, A.T., Doherty, O.M., 2004. Temporal, spatial and meteorological variations in hourly $PM_{2.5}$ concentration extremes in New York City. Atmos. Environ. 38, 1547–1558.

Dixon, P.G., Allen, M., Gosling, S.N., Hondula, D.M., Ingole, V., Lucas, R., Vanos, J., 2016. Perspectives on the synoptic climate classification and its role in interdisciplinary research. Geography Compass 10, 147–164.

Dobson, J.E., Bright, E.A., Coleman, P.R., Durfee, R.C., Worley, B.A., 2000. LandScan: a global population database for estimating populations at risk. Photogramm. Eng. Rem. Sens. 66, 849–857.

Eiguren-Fernandez, A., Miguel, A.H., Froines, J.R., Thurairatnam, S., Avol, E.L., 2004. Seasonal and spatial variation of polycyclic aromatic hydrocarbons in vapor-phase and PM2.5 in Southern California urban and rural communities. Aerosol. Sci. Technol. 38, 447–455.

Elvidge, C.D., Baugh, K., Zhizhin, M., Hsu, F.C., Ghosh, T., 2017. VIIRS night-time lights. Int. J. Rem. Sens. 38, 5860–5879.

Feng, J., Sun, P., Hu, X., Zhao, W., Wu, M., Fu, J., 2012. The chemical composition and sources of $PM_{2.5}$ during the 2009 Chinese New Year's holiday in Shanghai. Atmos. Res. 118, 435–444.

Gao, J., Tian, H., Cheng, K., Lu, L., Zheng, M., Wang, S., Hao, J., Wang, K., Hua, S., Zhu, C., Wang, Y., 2015. The variation of chemical characteristics of $PM_{2.5}$ and PM_{10} and formation causes during two haze pollution events in urban Beijing, China. Atmos. Environ. 107, 1–8.

Gehrig, R., Buchmann, B., 2003. Characterising seasonal variations and spatial distribution of ambient PM10 and PM2.5 concentrations based on long-term Swiss monitoring data. Atmos. Environ. 37, 2571–2580.

Gentner, D.R., Jathar, S.H., Gordon, T.D., Bahreini, R., Day, D.A., Haddad, I.E., Hayes, P.L., Pieber, S.M., Platt, S.M., Gouw, J., Goldstein, A.H., Harley, R.A., Jimenez, J.L., Prevot, A.S.H., Robinson, A.L., 2017. Review of urban secondary organic aerosol formation from gasoline and diesel motor vehicle emissions. ES T (Environ. Sci. Technol.) 51, 1074–1093.

Gorin, C.A., Collett, J.L., Herckes, P., 2006. Wood smoke contribution to winter aerosol in Fresno, CA. J. Air Waste Manag. Assoc. 56, 1584–1590.

Greene, J.S., Kalkstein, L.S., Ye, H., Smoyer, K., 1999. Relationships between synoptic climatology and atmospheric pollution at 4 US cities. Theor. Appl. Climatol. 62, 163–174.

Gu, J., Bai, Z., Li, W., Wu, L., Liu, A., Dong, H., Xie, Y., 2011. Chemical composition of $PM_{2.5}$ during winter in Tianjin, China. Particuology 9 (3), 215–221.

Hardin, A.W., Liu, Y., Cao, G., Vanos, J.K., 2018. Urban heat island intensity and spatial variability by synoptic weather type in the northeast U.S. Urban Climate 24, 747–762.

Harvey, A.C., Shephard, N., 1993. Structural time series models. In: In: Maddala, G., Rao, C., Vinod, H. (Eds.), Handbook of Statistics, vol 11. Elsvier, pp. 261–302 (Chapter 10).

Huang, R.-J., Zhang, Y., Bozzetti, C., Ho, K.-F., Cao, J.-J., Han, Y., Daellenbach, K.R., Slowik, J.G., Platt, S.M., Canonaco, F., Zotter, P., Wolf, R., Pieber, S.M., Bruns, E.A., Crippa, M., Ciarelli, G., Piazzalunga, A., Schwikowski, M., Abbaszade, G., Schnelle-Kreis, J., Zimmermann, R., Zhisheng, An, Szidat, S., Baltensperger, U., Haddad, I., Prevot, A.H., 2014. High secondary aerosol contribution to particulate pollution during haze events in China. Nature 514, 218–222.

Liu, Y., Zhao, N., Vanos, J.K., Cao, G., 2017. Effects of synoptic weather on ground-level $PM_{2.5}$ concentrations in the United States. Atmos. Environ. 148, 297–305.
Lough, G.C., Schauer, J.J., Lawson, D.R., 2006. Day-of-week trends in carbonaceous aerosol composition in the urban atmosphere. Atmos. Environ. 40, 4137–4149.
Martellini, T., Giannoni, M., Lepri, L., Katsoyiannis, A., Cincinelli, A., 2012. One year intensive $PM_{2.5}$ bound polycyclic aromatic hydrocarbons monitoring in the area of Tuscany, Italy. Concentrations, source understanding and implications. Environ. Pollut. 164, 252–258.
Motallebi, N., 1999. Wintertime $PM_{2.5}$ and PM_{10} source apportionment at Sacramento, California. J. Air Waste Manag. Assoc. 49, 25–34.
Motallebi, N., Tran, H., Croes, B.E., Larsen, L.C., 2003. Day-of-week patterns of particulate matter and its chemical components at selected sites in California. J. Air Waste Manag. Assoc. 53, 876–888.
Niu, M., Wang, Y., Sun, S., Li, Y., 2016. A novel hybrid decomposition-and-ensemble model based on CEEMD and GWO for short-term $PM_{2.5}$ concentration forecasting. Atmos. Environ. 134, 168–180.
Pongpiachan, S., Kositanont, C., Palakun, J., Liu, S., Ho, K.F., Cao, J., 2015. Effects of day-of-week trends and vehicle types on $PM_{2.5}$-bounded carbonaceous compositions. Sci. Total Environ. 532, 484–494.
Russell, M., Allen, D.T., Collins, D.R., Fraser, M.P., 2004. Daily, seasonal, and spatial trends in $PM_{2.5}$ mass and composition in Southeast Texas. Aerosol. Sci. Technol. 38, 14–26.
Sak, H., Senior, A., Beaufays, F., 2014. Long short-term memory recurrent neural network architectures for large scale acoustic modeling. In INTERSPEECH 2014, 338–342.
Song, Y., Xie, S., Zhang, Y., Zeng, L., Salmon, L.G., Zheng, M., 2006. Source apportionment of $PM_{2.5}$ in Beijing using principal component analysis/absolute principal component scores and UNMIX. Sci. Total Environ. 372, 278–286.
StatSoft, 2011. Time series analysis. In: Electronic Statistics Textbook, Available from. https://www.betterevaluation.org/en/evaluation-options/timeseriesanalysis, Accessed date: 9 August 2018.
Sutton, P.C., Roberts, D., Elvidge, C.D., Baugh, K., 2001. Census from Heaven: an estimate of the global human population using night-time satellite imagery. Int. J. Rem. Sens. 22, 3061–3076.
Sutton, P.C., Goetz, A.R., Fildes, S., Forster, C., Ghosh, T., 2010. Darkness on the edge of town: mapping urban and peri-urban Australia using nighttime satellite imagery. Prof. Geogr. 62, 119–133.
Tai, A.P., Mickley, L.J., Jacob, D.J., 2010. Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: implications for the sensitivity of $PM_{2.5}$ to climate change. Atmos. Environ. 44, 3976–3984.
Taylor, S.J., Letham, B., 2017. Forecasting at scale. Am. Statistician. https://doi.org/10.1080/00031305.2017.1380080.
United States Environmental Protection Agency, 2015. AirData download files documentation. Available from. http://aqsdr1.epa.gov/aqsweb/aqstmp/airdata/FileFormats.html, Accessed date: 8 April 2018.
Vanos, J.K., Cakmak, S., Kalkstein, L.S., Yagouti, A., 2015. Association of weather and air pollution interactions on daily mortality in 12 Canadian cities. Air Quality, Atmosphere & Health 8 (3), 307–320.
Vecchi, R., Marcazzan, G., Valli, G., Ceriani, M., Antoniazzi, C., 2004. The role of atmospheric dispersion in the seasonal variation of PM1 and PM2.5 concentration and composition in the urban area of Milan (Italy). Atmos. Environ. 38, 4437–4446.
Viswanathan, S., Eria, L., Diunugala, N., Johnson, J., McClean, C., 2006. An analysis of effects of San Diego wildfire on ambient air quality. J. Air Waste Manag. Assoc. 56 (1), 56–67.
Wang, P., Cao, J.-J., Shen, Z.-X., Han, Y.-M., Lee, S.-C., Huang, Y., Zhu, C.-S., Wang, Q.-Y., Xu, H.-M., Huang, R.-J., 2015. Spatial and seasonal variations of $PM_{2.5}$ mass and species during 2010 in Xi'an, China. Sci. Total Environ. 508, 477–487.
Wang, J., Ogawa, S., 2015. Effects of meteorological conditions on PM2.5 concentrations in Nagasaki, Japan. Int. J. Environ. Res. Publ. Health 12, 9089–9101.
Wang, P., Zhang, H., Qin, Z., Zhang, G., 2017. A novel hybrid-Garch model based on ARIMA and SVM for $PM_{2.5}$ concentrations forecasting. Atmospheric Pollution Research 8, 850–860.
Wang, F., Guo, Z., Lin, T., Rose, N.L., 2016. Seasonal variation of carbonaceous pollutants in PM2.5 at an urban 'supersite' in Shanghai, China. Chemosphere 146, 238–244.
Wegesser, T.C., Pinkerton, K.E., Last, J.A., 2009. California wildfires of 2008: coarse and fine particulate matter toxicity. Environ. Health Perspect. 117, 893–897.
Wu, J., Winer, A.M., Delfino, R., 2006. Exposure assessment of particulate matter air pollution before, during, and after the 2003 Southern California wildfires. Atmos. Environ. 40, 3333–3348.
Xie, Y., Zhao, B., Zhang, L., Luo, R., 2015. Spatiotemporal variations of $PM_{2.5}$ and PM_{10} concentrations between 31 Chinese cities and their relationships with SO_2, NO_2, CO and O_3. Particuology 20, 141–149.
Xu, L., Chen, X., Chen, J., Zhang, F., He, C., Zhao, J., Yin, L., 2012. Seasonal variations and chemical compositions of $PM_{2.5}$ aerosol in the urban area of Fuzhou, China. Atmos. Res. 104–105, 264–272.
Zhang, F., Wang, Z.-W., Cheng, H.-R., Lv, X.-P., Gong, W., Wang, X.-M., Zhang, G., 2015. Seasonal variations and chemical characteristics of $PM_{2.5}$ in Wuhan, central China. Sci. Total Environ. 518–519, 97–105.
Zhao, X., Zhang, X., Xu, X., Xu, J., Meng, W., Pu, W., 2009. Seasonal and diurnal variations of ambient $PM_{2.5}$ concentration in urban and rural environments in Beijing. Atmos. Environ. 43, 2893–2900.
Zheng, L., Porter, E.N., Sjodin, A., Needham, L.L., Lee, S., Russell, A.G., Mulholland, J.A., 2009. Characterization of $PM_{2.5}$-bound polycyclic aromatic hydrocarbons in Atlanta—seasonal variations at urban, suburban, and rural ambient air monitoring sites. Atmos. Environ. 43, 4187–4193.
Zheng, M., Salmon, L.G., Schauer, J.J., Zeng, L., Kiang, C.S., Zhang, Y., Cass, G.R., 2005. Seasonal trends in PM2.5 source contributions in Beijing, China. Atmos. Environ. 39, 3967–3976.