Climatic factors and their availability in estimating long-term variations of fine particle distributions over East China

Yanyu Wang¹, Junyan Duan¹, Xin Xie¹, Qianshan He²,³ *, Tiantao Cheng⁴,¹,⁵ *, Haizhen Mu², Wei Gao²,³, Xiang Li¹

¹ Shanghai Key Laboratory of Atmospheric Particle Pollution and Prevention (LAP³), Department of Environmental Science and Engineering, Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China.
² Shanghai Meteorological Service, Shanghai 20030, China.
³ Shanghai Key Laboratory of Meteorology and Health, Shanghai 20030, China.
⁴ Department of Atmospheric and Oceanic Sciences, Institute of Atmospheric Sciences, Fudan University, Shanghai 200438, China
⁵ Shanghai Institute of Eco-Chongming (SIEC), Shanghai 200062, China.

Corresponding author: Qianshan He (oxeye75@163.com) ; Tiantao Cheng(ttcheng@fudan.edu.cn);

Key Points:
- The climatic factors are closely related with seasonal PM<sub>2.5</sub> distributions over East China during 2000-2015.
- It is feasible to use these factors as predictors to estimate PM<sub>2.5</sub> concentrations.
- The contributions of climatic factors to PM<sub>2.5</sub> variability is 5%-30%.
Abstract

Concerns about air pollutions in China have made fine particulate matter (PM$_{2.5}$) estimation a hot research topic in recent years. Using satellite-based PM$_{2.5}$ data obtained over East China in 2000–2015, climatic factors affecting the PM$_{2.5}$ variability were revealed in different seasons. Relevant climatic factors were used to estimate seasonal PM$_{2.5}$ distributions by combining Empirical Orthogonal Function (EOF) with stepwise regression. The assessment results indicate a greater impact by climatic factors on PM$_{2.5}$ variability in western East China than over coastal areas, and more significant influence by climatic factors on PM$_{2.5}$ variability also occurs in summer and autumn than spring and winter. Our favorable results for PM$_{2.5}$ estimation in 2016 further validate the capability of the estimation method for PM$_{2.5}$ distributions based on climatic factors. Individual contributions of dominant climatic factors to PM$_{2.5}$ variability over East China during the past 16 years ranges from 5% to 30%. In conclusion, climatic factors play non-negligible roles in modulating PM$_{2.5}$ concentrations, and they can feasibly be considered as predictors in the estimation of PM$_{2.5}$ concentrations in East China. Our work underscores the importance of climate change impacts on air pollutions and improves air quality forecasts over East China.

1 Introduction

Aerosols can influence global climate change directly, by absorbing and scattering solar radiation, and indirectly, through participation in cloud formation by acting as cloud condensation nuclei or ice nuclei (Ramanathan et al., 2001; Rosenfeld et al., 2007; Twomey, 1974). Near the surface, aerosols, especially fine particulate matter (PM$_{2.5}$; aerodynamic diameter ≤ 2.5 μm), often result in visibility impairment and air pollutions (Fu & Chen, 2017; Gao et al., 2017). Aerosols also impose adverse effects on human health, particularly respiratory and circulatory health (Chen et al., 2017).

Great efforts have been devoted to investigating the mechanism of air pollution in China (Huang et al., 2014; Zhang et al., 2014; Leng et al., 2016; Sun et al., 2016; Wang et al., 2018). It is well known that increase of emissions is considered as the primary reason for the increase of polluted days in China (Chen & Wang, 2015). Apart from emissions, meteorological conditions and climate also play vital roles in the air pollution. China is located in the monsoon climate region, where atmospheric circulations, temperature and precipitation have the distinct seasonal variations. Thus, the monsoon can influence the generation, transportation and deposition of aerosols (Bollasina et al., 2011; Li et al., 2016; Wu et al., 2016). In recent years, the changes of East Asian monsoon have exerted much influence on the long-term distributions of aerosols under the climate warming (Liu et al., 2016). For example, Li et al. (2016) found that more wintertime haze days are associated with weak East Asian Winter Monsoon (EAWM) in the interannual time scale. Cai et al. (2017) demonstrated that circulation variations induced by climate changes could increase the severity of hazes in Beijing, such as the upward trend in the Arctic Oscillation, the weakening East Asian monsoon, and the fast warming in the lower troposphere. Moreover, Arctic sea ice, boreal Eurasia snowfall, El Niño-Southern Oscillation (ENSO), Pacific–North America teleconnection (PNA), Siberian High and sea surface temperature (SST) anomalies also contribute somewhat to aerosol pollution (Chen & Wang, 2015; Hui & Xiang, 2015; Xiao et al., 2015; Feng et al., 2016; Wang & Chen, 2016; Zhang et al., 2016; Zhou et al., 2017; Zhao et al., 2018). The large-scale atmospheric circulations can impact on the spatial pattern of PM$_{2.5}$ mainly through changes in synoptic systems and meteorological conditions. Specific synoptic systems, such as the presence of high-pressure systems with subsiding airflow and uniform surface pressure fields, are favorable for the enhancement of PM$_{2.5}$ concentrations (Tai et al., 2011; Zheng et al., 2015; Miao et al., 2016; Leung et al., 2017). As for
meteorological conditions, many studies investigated that, high relative humidity, calm wind, shallow planetary boundary layer, and reduced precipitation can be typically responsible for instantaneous aerosol accumulations at the lower atmosphere (Wang et al., 2016; Yang et al., 2016; Ye et al., 2016; Tie et al., 2017).

A better understanding of the relationships between climatic factors and PM$_{2.5}$ is helpful for PM$_{2.5}$ prediction. Numerical simulations have been extensively used to predict PM$_{2.5}$ concentrations. However, because of the large uncertainties in pollutant emission inventories and imperfect model parametrization, PM$_{2.5}$ estimation by numerical experiments remains challenging. In addition to model simulations, statistical methods are also applied extensively for PM$_{2.5}$ estimation. For example, Leceur et al. (2014) proposed a novel statistical algorithm involving regression and analog steps to estimate PM$_{2.5}$ concentrations in Europe based on weather modeling. Using 2 years of aerosol optical depth (AOD) data from the Moderate Resolution Imaging Spectroradiometer (MODIS) and meteorological factors at five sites in China, Guo et al. (2013) developed a back-propagation neural work model to estimate PM$_{2.5}$ concentrations. By establishing a forecast meteorological index, Yin et al. (2016) used multiple linear regression (MLR) and generalized additive model to predict winter haze days over the central North China Plain (NCP). All these methods were obtained favorable prediction skills and can also provide useful bias corrections to improve numerical models. As an important statistical method, Empirical Orthogonal Function (EOF) decomposition has an outstanding advantage of combining statistical analysis with physical meanings (Wilks, 2006). It has been applied widely in climate predictions and socioeconomic researches, and highly praised for its favorable performance (Wang et al., 2013; Wu et al., 2013; Wang et al., 2015; Grunseich & Wang, 2016). EOF provides new insight for the development of methods for PM$_{2.5}$ prediction.

East China is one of the most populated and polluted regions in China. With rapid urbanization and economic development, air pollution has caused severe environmental and climate problems (Hou et al., 2010; Yang et al., 2018). However, the impact of climatic factors on PM$_{2.5}$ concentrations under climate change remains unclear over East China. An analysis of PM$_{2.5}$ variability from a perspective of seasonal scale also has been lacking. With the advantages of continuous spatial coverage and high resolution, satellite remote sensing datasets can provide long-term aerosol products and have been gradually retrieved surface-level PM$_{2.5}$ concentrations (Lin et al., 2015; Zhang et al., 2015). In this study, using seasonal satellite-based PM$_{2.5}$ data over East China from 2000 to 2015, the relationships between climatic factors and PM$_{2.5}$ variability were examined and analyzed in different seasons. The objective of this study is to use relevant climatic factors to estimate long-term variations of fine particle distributions in this region and to evaluate the contribution of these factors to PM$_{2.5}$ variability. Section 2 describes the data acquisition and methods. Section 3 discusses seasonal climatic factors and presents our scheme for the forecast of seasonal PM$_{2.5}$ anomalies. Section 4 evaluates the suitability of the method for PM$_{2.5}$ prediction, and quantifies the contributions of these factors to long-term PM$_{2.5}$ variations. The conclusion is presented in section 5.

2 Data and Methods

2.1 Data

MODIS-derived AOD and the relative humidity (RH) data were used to calculate the surface-level PM$_{2.5}$ concentrations, and the detailed algorithm will be described in the following Methodology. The time length of AOD data (Collection 6) of MODIS Level 2 products over land from the Terra satellite were from February 25, 2000 to December 31,
2016. He et al. (2010) demonstrated that MODIS/AOD at 550 nm highly correlated with sun photometer (CE318) measurements at 7 sites in the Yangtze River Delta (YRD) region, with a correlation coefficient of 0.85 and 90% of cases falling in the range of ΔAOD = ± 0.05 ± 0.20 AOD (Remer et al., 2005). RH data in over 400 stations provided by the China Meteorological Administration (CMA) operation ranged from 2000 to 2016. According to “Quality control of surface meteorological observational data (QX/T 118–2010)”, the quality control of this dataset was performed by the National Meteorological Information Center. 130 monthly climatic indices (88 atmospheric circulation indices, 26 ocean temperature indices, and 16 other indices) were considered as reference factors to examine associations with PM$_{2.5}$ variability. Unlike meteorological index defined empirically, these climatic indices have clear definitions from the National Climate Center of China and have been widely used in climate diagnosis and predictions (Cantelaube et al., 2004; Hallett et al., 2004). It is acknowledged that climatic factors play vital roles in the long-term and large-scale distributions of aerosols, whereas the variations of meteorological parameters, such as a specific synoptic process, influence the instantaneous perturbations of PM$_{2.5}$ concentrations in the limited region (Zheng et al., 2015; Leung et al., 2017). Moreover, the variations of climatic indices can include all the changes in atmospheric circulation patterns and interactions among meteorological factors. The geopotential height field and wind field (1° × 1° resolution), which were used to characterize the climate conditions, were obtained from global daily atmospheric reanalysis data from the National Centers for Environmental Prediction–Final Operational Global Analysis (NCEP-FNL) (Kalnay et al., 1996). Monthly precipitation data (0.5° × 0.5° resolution) were obtained from CMA. In this study, the focus areas in East China were Shandong, Anhui, Jiangsu, Zhejiang, Jiangxi, and Fujian provinces and Shanghai. Seasons are defined as follows: spring, March–May; summer, June–August; autumn, September–November; and winter, December–February. Note that the winter between 2000 and 2001 is defined as the winter of 2001; the same convention was used for all years.

2.2 Methodology

2.2.1 the Satellite-based PM$_{2.5}$ retrieval

To obtain satellite-based PM$_{2.5}$ concentrations in East China, we firstly applied the vertical correction algorithm from He et al. (2016a) to calculate the surface-level aerosol extinction coefficient from the MODIS/AOD product. Then the humidity correction algorithm from He et al. (2016b) was used to retrieve PM$_{2.5}$ concentrations in spatial resolution of 10 km × 10 km, and the resolution is consistent with MODIS/AOD product. Brief descriptions of this process go as follows.

First, aerosol extinction coefficient was assumed to decrease exponentially with altitude above the top of Planet Boundary Layer (PBL) height. Based on this vertical distribution model of aerosols, two crucial parameters of retrieving the surface-level aerosol extinction coefficient from AOD are PBL height and aerosol layer height (ALH). The PBL heights were simulated using the WRF Model (version 3.2.1). According to the vertical distribution of meteorological parameters (temperature, relative humidity, wind speed and wind direction) in each grid from NCEP-FNL data, an automated workflow algorithm was proposed to calculate the ALH. The long-term spatial comparison between satellite estimations and the surface measurements displays that correlation coefficients of 90% samples are greater than 0.6, and 68% of the samples have coefficients higher than 0.7 (He et al., 2016a).

Subsequently, an observation-based spatial indicator for integrated humidity effect in East China was introduced to calculate PM$_{2.5}$ mass concentrations. Specifically, mass
extinction efficiency and hydroscopic growth coefficient were calculated by linear regression using aerosol extinction coefficient, RH and PM$_{2.5}$ in situ observation (Lin et al., 2015) during 2014. The calculated PM$_{2.5}$ mass concentrations were validated to have a good agreement with those in situ measurements, with correlation coefficients of over 0.85 (He et al., 2016b). Due to a lack of long-term observation data, these two factors (mass extinction efficiency and hygroscopic growth coefficient) were treated as the values calculated in 2014 in the following retrieval (2000-2016). The uncertainty in the variation of these two factors, however, needs concerns. These factors depend closely on the aerosol species and there are various emission sources in different years over East China (Titos et al., 2016). The uncertainty can be estimated by the discrepancy between the extremum with according specific aerosols and the mean value. For example, according to former studies about humidity effects on aerosol light scattering, the hygroscopic growth coefficient of aerosols in East China was measured from the lowest 0.5 (hydrophobic aerosols like urban or polluted aerosols) to the highest 0.83 (hygroscopic aerosols like marine aerosols) (Liu et al., 2008; Xu et al., 2012; Zhang et al., 2015; Zhuang et al., 2015). The uncertainty of hygroscopic growth coefficient can be evaluated within the range from 20% to 32%. Similarly, the uncertainty of aerosol extinction efficiency is estimated 26%-33%, which is calculated from Hand and Malm (2007) and Shen et al. (2017). Zhang et al. (2015) also estimated that, for the PM$_{2.5}$ retrieval, around 24% uncertainty on the hygroscopic growth item and 34% on all terms. In short, these uncertainties of humidity correction algorithm limit the accuracy of surface-level PM$_{2.5}$ retrieval with acceptance error.

To eliminate small-scale local PM$_{2.5}$ fluctuations, seasonal PM$_{2.5}$ data was averaged in the spatial resolution of 30 km × 30 km.

2.2.2 EOF analysis

EOF analysis can decompose three-dimensional data (e.g., a geophysical field) into two components, representing spatial and temporal variations, as follows:

$$X_{m \times n} = V_{m \times m}T_{m \times n}$$  \hspace{1cm} (1)

where $X_{m \times n}$ is an $m \times n$ matrix of observed PM$_{2.5}$ data, $V_{m \times m}$ is an $m \times m$ matrix of spatial modes, and $T_{m \times n}$ is an $m \times n$ matrix of principal components (PCs). The variance proportion associated with the modes is in a descending order. Therefore, the first several EOF modes, which represent the sum of spatial patterns multiplied by the PCs, generally capture most of the PM$_{2.5}$ variability (Wilks, 2006). In this study, seasonal satellite-based PM$_{2.5}$ data were decomposed by EOF analysis, the only two leading modes of PM$_{2.5}$ variability were chosen because they were well separated from the remaining modes according to criteria proposed by North et al. (1982). The spatial patterns of the two leading modes includes the first and second EOF spatial patterns (EOF1 and EOF2, respectively), corresponding to the first and second principal components (PC1 and PC2, respectively).

2.2.3 Key climatic factors for PC estimation

Prior to PC estimation, all climatic indices and PCs were standardized by subtracting the mean and dividing by the standard deviation (SD) as follows:

$$\bar{x}_i(t) = \frac{x_i(t) - \bar{x}_i}{s_{x_i}}$$  \hspace{1cm} (2)

$$\bar{T}_i(t) = \frac{T_i(t) - \bar{T}_i}{s_{T_i}}$$  \hspace{1cm} (3)
where \( \hat{x}_i(t) \) is the time series of normalized climatic indices, \( \bar{x}_i \) and \( s_{x_i} \) are the mean and SD of the climatic index time series, \( \hat{T}_i(t) \) is the normalized PC time series, \( \bar{T}_i \) and \( s_{T_i} \) are the mean and SD of the PC time series.

PC prediction involves two steps: (1) screening of potential predictors and (2) use of stepwise regression to establish the equation for PC prediction based on these predictors. We first performed correlation analysis between climatic indices and PCs, and the indices with passing at the 95% significance level in the F-test are selected. Then, these indices, which have no discernible physical linkage with PM\(_{2.5}\) variations in East China, were disregarded, such as North African Subtropical High Area Index, North American Subtropical High Intensity Index and so on. About ten indices were retained and can be considered as potential predictors after predictor screening. Finally, in stepwise regression, these potential predictors can be removed or added in order to ensure high correlation with predictant and eliminate the collinearity among the selected climatic indices. Specifically, in the stepwise regression model, the statistical significance of each variable was assessed at a level of \( P < 0.05 \) in standard F test (Panofsky & Brier, 1965). Once a predictor was added to the model, it could be removed only if its significance level fell below \( P < 0.05 \) after the addition or removal of another variable. This screening process was performed once for every PC for all seasons. Thus, the fitted PC equation was obtained as:

\[
\hat{T}(t) = \sum_{j=1}^{N} b_j \hat{x}_j(t) + e
\]  

where \( b_j \) is the regression coefficient and \( N \) is the number of climatic factors selected for the regression equations. \( e \) denotes the noise term, which is the difference between the observed PC and the predicted PC. In this manner, next-year climatic indices (\( \hat{x}_j(t+1) \)) are predicted from literatures and input into their respective equations to yield the predicted next-year PCs.

2.2.4 PM\(_{2.5}\) estimation and validation

The predicted PCs and spatial patterns of two leading modes (\( k = 2 \)) were used to estimate the next-year PM\(_{2.5}\) distribution (\( X_{m,n+1} \)):

\[
X_{m,n+1} \approx V_m k T_{k,n+1}
\]

where \( T_{k,n+1} \) represents next-year PCs predicted by stepwise regression. Briefly, PM\(_{2.5}\) estimation is an integral method combing EOF analysis and stepwise regression. EOF analysis detects the two leading EOF modes, which represent the most important PM\(_{2.5}\) variations, and stepwise regression is used to predict PCs based on climatic factors. Consequently, the predicted PCs and spatial patterns jointly contribute to PM\(_{2.5}\) estimation.

To evaluate the feasibility of climatic factors for PM\(_{2.5}\) estimation and the capability of the PM\(_{2.5}\) estimation method, a retrospective experiment was conducted to obtain PM\(_{2.5}\) data (\( X_{m,n} \)) approximately by using the sum of the two leading EOF spatial modes (\( k = 2 \)) multiplied by PCs as follows:

\[
X_{m,n} \approx V_m k T_{k,n}
\]

In the retrospective experiment, the predicted PM\(_{2.5}\) data (PPD) is obtained by the sum of observed two leading EOF spatial patterns multiplied by predicted PCs, which is calculated by inputting the climatic indices from the previous 16 years into the equation (4). For comparison, the reconstructed PM\(_{2.5}\) data (RPD) is also calculated by the sum of observed two leading EOF spatial patterns multiplied by observed PCs. That is to say, RPD depend on the observed PCs and PPD depend on the predicted PCs.
Besides above, all statistical correlation analyses were evaluated using two-tailed Student's t-tests.

3 Results

3.1 Climatic factors influencing PM$_{2.5}$ distributions

The dominant spatiotemporal variations of seasonal PM$_{2.5}$ distributions in East China were identified by EOF analysis. The two leading EOF modes accounted for about 38.7% of the total variance in spring, 65.9% in summer, 60.2% in autumn, and 44.7% in winter (Table 1). He et al. (2018) reported the impacts of climatic factors on PM$_{2.5}$ variability in East China. On the basis of the selected factors, collinearity among the selected climatic variables was eliminated using stepwise regression, and the links between large-scale atmospheric circulations and local PM$_{2.5}$ variations were examined in spring. During this period, the Western Pacific Warm Pool Strength (WPWPS) index exhibits a significant in-phase relationship with spring PC1. The 500 hPa geopotential height field (H500) is correlated positively with the WPWPS in East China (Fig. 1a), suggesting that the H500 responds to variation in the WPWPS through atmospheric teleconnection. The mean wind anomaly fields at 850 hPa during two different periods are presented in Figure 1b and c, where high-index (low-index) years are defined as those in which the WPWPS was higher (lower) than average during 2000–2015. Figure 1b shows an apparent anticyclone in the South China Sea (SCS), and southerly anticyclone winds over the northwest transporting water vapor to South China in high-index years. The opposite pattern occurs in the SCS in low-index years (Figure 1c). Figure 1d displays a positive correlation between the WPWPS and precipitation in the middle of East China, indicating that more rainfall is induced by enhanced WPWPS. These results indicate that the H500 responds significantly to increased WPWPS, and wet warm anticyclone airflow from the west in the SCS and dry cold airflow from northern regions converge on East China, often producing precipitation (He et al., 2015; Ren et al., 2017). Consequently, precipitation favors PM$_{2.5}$ scavenging over this region.

Other sea temperature indices associated with PM$_{2.5}$ variability include the summer Kuroshio Current SST index (KCSST) and winter El Niño A SST anomaly (NINO A SSTA). The KCSST is linked closely to summer precipitation in China (Dong et al., 2004), and the winter NINO A SSTA is related to the ENSO. Using statistical and numerical methods, Zhao et al. (2018) demonstrated that haze days in southern China occurred less (more) frequently than normal when El Niño (La Niña) events occurred in winter. Overall, SST is an important external forcing factor that can adjust large-scale atmospheric circulations through atmospheric teleconnection. Subsequently it exerts considerable effects on local synoptic systems and meteorological conditions, indirectly modifying PM$_{2.5}$ concentrations. Yin and Wang (2015) showed that SST anomalies in the subtropical western Pacific were associated with haze because they can weaken EAWM, leading to favorable conditions for haze occurrence. These conditions include a weaker Siberian High, reduced wind speed, higher humidity, and abnormal southerly wind. By a similar mechanism, Atlantic Ocean SST anomalies were related closely to haze days on the NCP due to the weakened wind field induced by the Atlantic Multidecadal Oscillation (Xiao et al., 2015).

Apart from sea temperature indices, other indices are also found to modulate seasonal PM$_{2.5}$ distributions. The Asia Polar Vortex is a non-negligible contributor of PM$_{2.5}$ concentrations in spring, autumn, and winter. The intensity and location of the Asia Polar Vortex are coupled closely to outbreaks of cold air, which facilitate pollutants dispersion (He et al., 2018). In summer, the Northern Hemisphere Subtropical High Ridge and Western North Pacific Typhoon are the key factors influencing the East Asian summer monsoon
(EASM), causing temperature changes and rainbelt movement in summer. Many studies have demonstrated the close relationship between the EASM and summer aerosol concentrations (Li et al., 2016; Wang et al., 2018). For example, Zhu et al. (2012) used a simulation model to show that surface-layer PM$_{2.5}$ concentrations were 17.7% higher in the weakest monsoon years than in strong monsoon years. Autumn PM$_{2.5}$ variability also depends on the intensity of the East Asian Trough; a shallow East Asian Trough can lead to the reduction of horizontal and vertical advection, followed by increased haze frequency (Zhang et al., 2016). In winter, cold air breaks can significantly transport and scavenge pollutants. As illustrated by Zhang et al. (2016), a cold surge over northern China was found to contribute about half of the decrease in total PM$_{2.5}$ during the Asia-Pacific Economic Cooperation Summit (APEC) in 2014. Zhao et al. (2016) reported that the Pacific Decadal Oscillation (PDO) was associated with decadal variation in winter haze days in central eastern China, during the positive phase of the PDO, the Mongolia High became significantly stronger, accompanied by a strong sinking motion, and then increased atmospheric stability and aerosol accumulation. A deeper understanding of these climatic factors needed in the future research, especially through numerical simulations. Nonetheless, these discoveries have provided a basis for the prediction of PM$_{2.5}$ concentrations.

In this study, the climatic factors discussed above were used as predictors to conduct regression analysis and predict PCs in each season. Spring PC prediction was performed with two predictors: the WPWPS for PC1 and the Asia Polar Vortex Intensity Index (APVI) for PC2. For summer, three predictors were used: the Northern Hemisphere Subtropical High Ridge Position Index (NHSHRP) and Western North Pacific Typhoon number (WNPTN) for PC1, and the KCSST for PC2. Autumn PC prediction was performed with two predictors: the Asia Polar Vortex Area Index (APVA) for PC1 and the East Asian Trough Intensity Index (EATI) for PC2. For winter, four predictors were used: the Cold Air Activity Index (CAA) and the NINO A SSTA Index for PC1, and the PDO and Pacific Polar Vortex Area Index (PPVA) for PC2. All climatic factors and regression equations used for PC prediction are listed in Table 2, which also shows the correlation coefficients between climatic factors and PCs. The precise definitions of these predictors are provided by He et al. (2018).

3.2 Availability of climatic indices for PM$_{2.5}$ estimation

3.2.1 Evaluation of climatic factors on the PM$_{2.5}$ long-terms variations

To verify the stability of the relationships between PCs and climatic factors, a leave-one-out cross-validation was performed for PCs (Michaelsen, 1987), in which one PC was removed every year and the predictors (climatic factors) for the remaining PCs were used to predict PC data for that year. This process was repeated 16 times (for 16 years). Time series of observed and predicted PCs by cross-validation, and their correlation coefficients are provided in Figure 2. Correlations between cross-validated predicted PCs and observed PCs are relatively high, exceeding the 90% significance level in spring, summer, and autumn. These results indicate that the relationships between climatic indices and PCs are stable during the past 16 years and that this method for PM$_{2.5}$ estimation may be reliable. However, in winter, the correlation coefficient dropped sharply, implying that the influence of climatic factors on PM$_{2.5}$ is changeable. Possible reasons for the poor result in winter could be attributed to two main aspects. First, the shorter time period (15 years) and less valid samples due to cloud mask in East China in winter, leading to worsen representation of the first two EOF modes compared with other seasons. Second, during recent years, the energy savings and emissions reduction in China significantly impact the variation of PM$_{2.5}$ distributions,
especially for winter when residents heating is a preponderant aerosol source. These anthropogenic activities perturb and even dominate the variations of winter PM$_{2.5}$, making the relatively low correlation between climatic factors and PCs in the cross-validation. This unstable relationship also indicates unexpected prediction skills of PM$_{2.5}$ may occur in winter, which will further discuss in the next section.

The pattern correlation coefficient (PCC, http://glossary.ametsoc.org/wiki/Pattern_correlation) and the temporal correlation coefficient (TCC) were introduced to evaluate the forecast skill and the feasibility of climatic factors for PM$_{2.5}$ estimation in the retrospective experiment. Briefly, PCC is the correlativity of the PPD/PRD to the observed PM$_{2.5}$ data for all grids at the same year, and TCC is the temporal correlation coefficient in each grid between PPD/PRD and the observed during the past 16 years. Time series of PCC values between the observed PM$_{2.5}$ data and RPD and PPD is shown in Figure 3. The mean PCCs between the RPD and observed PM$_{2.5}$ are 0.42 in spring, 0.48 in summer, 0.53 in autumn, and 0.48 in winter. The mean PCCs between the PPD and observed PM$_{2.5}$ are 0.12 in spring, 0.36 in summer, 0.39 in autumn, and 0.36 in winter. These results indicate the high degree of similarity between the observed PM$_{2.5}$ data using all EOF modes and PM$_{2.5}$ data constructed with the two leading modes. PCC values might be improved by the addition of EOF modes (Wilks, 2006). The discrepancy between PPD and RPD shows the difference between predicted and observed PCs. To narrow this gap and improve the accuracy of PC prediction, researchers have tried many approaches such as partial least squares (Xing, et al., 2016), MLR (Wang et al., 2017), and vector autoregressive models (Wang et al., 2016). The distinct high-low variation in correlation coefficients shows that forecast skills are time-variant. Taking the PCC of PPD in summer as an example, high TCC values (> 0.6) were obtained for 2001, 2003, and 2014, and low TCC values (< 0.2) were obtained for 2002, 2007, 2010, and 2015. This pattern is commonly attributed to the uncertainty inherent in complex physical mechanisms. It may also be affected by extreme climate events, such as serious haze episodes (Pang et al., 2014).

Figure 4 shows the distribution of TCCs between observed PM$_{2.5}$ data and RPD/PPD in four seasons. High correlations between observed PM$_{2.5}$ data and RPD are found in most regions in different seasons (Figure 4a, 4c, 4e, 4g). High areal-averaged TCC between observed PM$_{2.5}$ data and RPD in Table 3 further implies that use of the two leading EOF modes to reconstruct PM$_{2.5}$ data is feasible. Evidently, RPD field reconstruction produces a closer approximation to the observed PM$_{2.5}$ data than PPD. Among the PPD results, TCCs show significant spatial and seasonal differences in East China. Strong correlations between PPD and observed PM$_{2.5}$ data are found over western areas, indicating that the impact of climatic factors on PM$_{2.5}$ is significant in these regions, although atmospheric circulation differs between the northern and southern regions of East China. Thus, PM$_{2.5}$ concentrations in inland areas are more sensitive to circulation conditions than in coastal areas. Moreover, correlations exceeding the 90% significance level are detected over nearly all regions in summer and autumn. Weak correlation occurs over almost half of East China, particularly in the Yangtze River Delta. The area-averaged TCCs of PPD with observed PM$_{2.5}$ data are 0.41 in spring, 0.57 in summer, 0.48 in autumn, and 0.42 in winter (Table 3). TCCs are higher in summer and autumn than in spring and winter, indicating that our method of PM$_{2.5}$ estimation, based on climatic factors in East China, is more suitable for summer and autumn. In spring, the proportion of the variance explained by the two leading modes is 38.7%, suggesting that almost 60% of PM$_{2.5}$ variability may not be estimated. The frequent transformation of the synoptic system, such as dust events and cold air breaks (Li et al., 2016; Liu et al., 2004), results in the irregular annual distribution of PM$_{2.5}$ in spring. Although the total contribution of the two leading modes is not least, the application of this estimation
method to winter data reveals difficulties. On the one hand, internal forces are changed artificially every year, such as increasing coal combustions and emission controls by government. On the other hand, external forces such as complex meteorological-PM$_{2.5}$ interactions, are very difficult to evaluate quantitatively. Hence, these disturbances strongly affect PM$_{2.5}$ variability, making the impact of climatic factors insignificant by comparison with other seasons. In contrast, fixed emissions and regular atmospheric circulation in summer and autumn favor stable PM$_{2.5}$ concentrations. In some cases, low-variance EOF modes may be of particular importance, suggesting that filtered noise signals influence forecast skills (Ding, 2011).

3.2.2 PM$_{2.5}$ estimation for 2016

To further verify the suitability of climatic factors for PM$_{2.5}$ estimation, we applied our method to determine the PM$_{2.5}$ anomaly pattern for 2016 using the last 16 years of data (2000-2015) as a training period. The observed satellite-based PM$_{2.5}$ and estimated PM$_{2.5}$ in the summer and autumn of 2016 are shown in Figure 5; unfavorable results for spring and winter are not shown. The PM$_{2.5}$ anomalies are obtained using this estimation method, therefore, these PM$_{2.5}$ anomalies are selected as the object of comparison between observed and estimated values. The signs of the values are used as a reference standard, and hence the areas with positive values are considered to be air quality warning regions, which need attention. On the contrary, negative values indicates the mitigation of PM$_{2.5}$ pollution, and these areas would not be the focus of government emissions regulation. Because estimate construction is based solely on climatic indices, the values are underestimated. Root mean square errors (RMSEs) and percentages of anomalies with the same sign in East China are calculated to evaluate forecast skill. The RMSEs are 17.22 μgm$^{-3}$ in summer and 15.43 μgm$^{-3}$ in autumn in 2016. The percentages of anomalies of the same sign were 63% in summer and 69.4% in autumn in 2016. These results show that the key features of the PM$_{2.5}$ anomaly pattern are captured with acceptable bias. Thus, the application of this estimation method for 2016 appears to be reasonable.

3.3 Contributions of climatic factors to long-term PM2.5 variations

To quantify the individual contributions of climatic factors during the past 16 years, the ratios of the variance in climatic factors to that in the observed PM$_{2.5}$ data were calculated. This calculation indicates the proportion of PM$_{2.5}$ variability that can be explained by climatic factors.

First, the variance in climatic factors was obtained by stepwise regression as follows:

\[
\frac{\text{variance(climatic factors)}}{\text{variance(PC)}} = \frac{\text{RSS}}{\text{TSS}} = \frac{\sum[b_j\bar{y}_j(t)]^2}{\sum[\bar{y}(t)]^2}
\]

(7)

where RSS is the regression sum of squares and TSS is the total sum of squares. The ratio of RSS to TSS is the contribution of climatic factors (predictor) to PCs (predictand), that is to say, the fraction of the variance in the PCs that can be explained by climatic factors in the stepwise regression model (Wilks, 2006).

The variance in the observed PM$_{2.5}$ data was obtained:

\[
C_{pc} = \frac{\text{variance(PC)}}{\text{variance(PM$_{2.5}$)}}
\]

(8)

where $C_{pc}$ is the ratio of the variance in the PCs to that in the observed PM$_{2.5}$ data, which equals the eigenvalue associated with the PCs divided by the sum of all eigenvalues in
the EOF analysis. The results are shown in Table 1. The contribution ($C_{climate}$) of climatic factors to PM$_{2.5}$ change during the past 16 years was calculated:

$$C_{climate} = \frac{\text{variance(climatic factors)}}{\text{variance}(\text{PM}_{2.5})} = \frac{RSS}{TSS} \times C_{PC}$$

(9)

These climatic factors explain 5-30% of seasonal PM$_{2.5}$ variability during the past 16 years (Table 4). The WPWPS is a dominant driver, with a 12.2% contribution in spring. In summer, the WNPTN and NHSHRP jointly account for 29.25% of the total variance in PM$_{2.5}$ concentrations. The APVA is the major contributor, with a 22.38% contribution in autumn. The joint contributions of CAA and NINO A SSTA to PM$_{2.5}$ concentrations are the largest among all climatic factors in winter, with a value of 16.61%. Previous studies have involved the quantitative evaluations of the effect of climatic factors on PM$_{2.5}$ and haze events. As demonstrated by Wang et al. (2015), approximately 45-67% of the interannual to interdecadal variability in haze occurrence can be attributed to Arctic sea ice. Yin and Wang (2015) revealed that subtropical Western Pacific SST anomaly can account for 40% of the variability of winter haze over NCP. It can be inferred that the contribution of climatic factors may be different to PM$_{2.5}$ concentrations and haze days, and climatic factors may contribute less to PM$_{2.5}$ concentrations. Luo et al. (2017) calculated about 1%-30% contribution to PM$_{2.5}$ across Mainland China during 1998-2012, which is quite similar to our results. It indicates that climatic factors play non-dominant roles in determining PM$_{2.5}$ concentrations over East China. Increased anthropogenic emissions are mostly considered as the primary factors driving PM$_{2.5}$ variability (Guo et al., 2011; Lin et al., 2014). Additionally, our results provide additional insight into the contributions of influential factors to PM$_{2.5}$ variability by statistical methods at long-time scales.

4 Conclusions

Based on seasonal surface-level PM$_{2.5}$ data derived from MODIS/AOD during 2000-2015, climatic factors were linked closely with PM$_{2.5}$ variability. These factors comprise the WPWPS and APVI in spring; the NHSHRP, WNPTN, and KCSST in summer; the APVA and EATI in autumn; and the CAA, NINO A SSTA, PDO, and PPVA in winter. These climatic factors dominate seasonal PM$_{2.5}$ distributions mainly through changes in atmospheric circulations and meteorological conditions, such as the pressure field, wind field, and precipitation. Then the monthly climatic indices were used to estimate the spatial distributions of PM$_{2.5}$ anomalies using combined EOF analysis and stepwise regression. By cross-validation and retrospective tests, the feasibility of climatic factors for PM$_{2.5}$ prediction was evaluated and discussed. The mean PCCs between PPD estimated based on climatic factors and observed PM$_{2.5}$ data are 0.12 in spring, 0.39 in summer, 0.39 in autumn, and 0.36 in winter. The area-averaged TCCs between PPD and observed PM$_{2.5}$ data are 0.41 in spring, 0.57 in summer, 0.48 in autumn, and 0.42 in winter. The impact of climatic factors on PM$_{2.5}$ is more significant in western East China than over coastal areas, and more significant in summer and autumn than spring and winter. An independent experiment was performed to evaluate PM$_{2.5}$ estimation in summer and autumn 2016, with satisfactory results. RMSE values between estimated and observed data are 17.22 $\mu$g m$^{-3}$ in the summer of 2016 and 15.43 $\mu$g m$^{-3}$ in the autumn of 2016, percentages of anomalies of the same sign are accurately predicted in more than 60% of the areas. Finally, individual contributions of climatic factors during the past 16 years over East China ranges from about 5% to 30%, indicating that climatic factors play non-negligible roles in modulating PM$_{2.5}$ concentrations over East China.
The novel method applied in this study shows favorable performance for most areas of East China. Based on the historical remote sensing data, it can be used to estimate long-term variations of PM$_{2.5}$ distributions. However, the prediction of PM$_{2.5}$ distributions in spring and winter still faces challenge by considering the climatic factors alone and needs to be investigated further. Recently, Yang et al. (2016) proposed an air quality forecasting method that considers both meteorological factors and emissions, producing satisfactory results, such an approach will be taken into consideration in our future work. In addition, simultaneous predictor-predictand relationships have been examined in our study, however, the improvement of other relationships in prediction, such as the lead-lag relationship (Wang et al., 2016; Zhu and Li, 2017), needs further explorations.

Acknowledgments

We sincerely acknowledge the Editor and anonymous reviewers and their kind and valuable comments greatly improved the manuscript. We acknowledge the MODIS Science Data Support Team and the Earth Observing System Data Gateway for processing and distributing the MODIS data used in this paper. We gratefully acknowledge the National Climate Center of China collecting the monthly climatic indices that from their website (http://cmdp.ncc.cma.gov.cn/Monitoring/cn_index_130.php) and Physical Sciences Division, which is responsible for the NCEP/NCAR reanalysis archive (http://www.esrl.noaa.gov/psd/data/gridded/data.ncep.reanalysis.html). We also thank the China Meteorological Administration (CMA) operation issuing database for providing the precipitation and RH data. This study was supported by the National Natural Science Foundation of China (91637101, 41475040, 41675109, 41775129), the National key research and development plan (2017YFC1501701, 2017YFC1501405, 2016YFC0202003) and the Shanghai Science and Technology Committee Research Project (16ZR1431700).

References


Zhang, Y., Li Z. Remote sensing of atmospheric fine particulate matter (PM$_{2.5}$) mass concentration near the ground from satellite observation. Remote Sensing of Environment, 2015, 160:252-262. doi: 10.1016/j.rse.2015.02.005


Table 1. The proportion of each principal component (PC) to the total variance in all seasons ($C_{pc}$, %).

<table>
<thead>
<tr>
<th>Season</th>
<th>PC1</th>
<th>PC2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>24.9</td>
<td>13.8</td>
</tr>
<tr>
<td>Summer</td>
<td>50</td>
<td>15.9</td>
</tr>
<tr>
<td>Autumn</td>
<td>48.4</td>
<td>11.8</td>
</tr>
<tr>
<td>Winter</td>
<td>33.9</td>
<td>13.8</td>
</tr>
</tbody>
</table>
Table 2. Climatic factors, correlation coefficients between climatic factors and principal components (PCs), and regression equations for PC predictions in all seasons.

<table>
<thead>
<tr>
<th>Seasons</th>
<th>PC1</th>
<th>PC2</th>
<th>WPWPS</th>
<th>APVI</th>
<th>WNPTN</th>
<th>NHSHRP</th>
<th>KCSST</th>
<th>APVA</th>
<th>EATI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring PC1</td>
<td>1.00</td>
<td>0.70</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring PC1 = 0.7 × WPWPS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring PC2</td>
<td>1.00</td>
<td>0.75</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Spring PC2 = 0.75 × APVI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer PC1</td>
<td>1.00</td>
<td>0.54</td>
<td>0.55</td>
<td>1.00</td>
<td>0.01</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer PC1 = 0.53 × WNPTN + 0.54 × NHSHRP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer PC2</td>
<td>1.00</td>
<td>0.68</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summer PC2 = 0.68 × KCSST</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autumn PC1</td>
<td>1.00</td>
<td>0.68</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autumn PC1 = 0.68 × APVA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autumn PC2</td>
<td>1.00</td>
<td>-0.65</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Autumn PC2 = -0.65 × EATI</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter PC1</td>
<td>1.00</td>
<td>-0.57</td>
<td>0.53</td>
<td>1.00</td>
<td>0.24</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter PC1 = -0.47 × CAA + 0.42 × NINO A SSTA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter PC2</td>
<td>1.00</td>
<td>-0.54</td>
<td>-0.44</td>
<td>1.00</td>
<td>0.3</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Winter PC2 = -0.48 × PPVA - 0.37 × PDO</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 3  Areal-averaged temporal correlation coefficients (TCCs) between reconstructed PM$_{2.5}$ data (RPD)/ predicted PM$_{2.5}$ data (PPD) and observed PM$_{2.5}$ data.

<table>
<thead>
<tr>
<th></th>
<th>Spring</th>
<th>Summer</th>
<th>Autumn</th>
<th>Winter</th>
</tr>
</thead>
<tbody>
<tr>
<td>RPD</td>
<td>0.58</td>
<td>0.77</td>
<td>0.75</td>
<td>0.56</td>
</tr>
<tr>
<td>PPD</td>
<td>0.41</td>
<td>0.57</td>
<td>0.48</td>
<td>0.42</td>
</tr>
</tbody>
</table>
Table 4 The contribution of climatic factors to long-term PM$_{2.5}$ variations ($C_{\text{climate}}$, %).

<table>
<thead>
<tr>
<th>Season</th>
<th>Climatic factors</th>
<th>Variance proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>WPWPS</td>
<td>12.20</td>
</tr>
<tr>
<td></td>
<td>APVI</td>
<td>7.76</td>
</tr>
<tr>
<td>Summer</td>
<td>WNPTN, NHSHRP, KCSST</td>
<td>29.25</td>
</tr>
<tr>
<td>Autumn</td>
<td>APVA</td>
<td>22.38</td>
</tr>
<tr>
<td></td>
<td>EATI</td>
<td>4.96</td>
</tr>
<tr>
<td>Winter</td>
<td>CAA, NINO A SSTa, PPVA, PDO</td>
<td>16.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>7.73</td>
</tr>
</tbody>
</table>
Figure 1. The relationship between the Western Pacific Warm Pool Strength (WPWPS) index associated with principal component 1 (PC1) and meteorological conditions in spring during 2000-2015. (a) Correlation coefficients between WPWPS and H500 in East China, (b) The mean wind anomalies at 850hPa with respect to the high index year (unit: m/s), (c) as (b) but the low index year, (d) correlation coefficients between WPWPS and precipitation. Dotted areas indicate the significance at over 90% confidence level. Note that there is blank exists in Shanghai because of missing values of monthly rainfall data.
Figure 2. Time series of (left) principal component 1 (PC1) / (right) principal component 2 (PC2) observations and predictions in the cross-validation test for all seasons. Red lines represent the predicted PCs by cross-validation, while the light gray bars are the observed PCs. R is the correlation between observed and predicted PCs.
Figure 3. Time series of pattern correlation coefficients (PCCs) for PM$_{2.5}$ prediction in all seasons during 2000-2015. The solid blue lines represent the PCC between PPD and observed PM$_{2.5}$ data. The dashed red lines represent the PCC between RPD and observed PM$_{2.5}$ data.
Figure 4. Temporal correlation coefficients (TCC) between RPD and observed PM$_{2.5}$ data in spring (a), summer (c), autumn (e) and winter (g). TCC between PPD and observed PM$_{2.5}$ data in spring (b), summer (d), autumn (f) and winter (h). Hatched regions represent exceeding the 90% significance level. Some blanks in grid cells of winter is due to the cloud mask of the unavailable data from satellite AOD retrieval.
Figure 5. The forecast (color shading) and observed (contours) PM$_{2.5}$ anomalies in (a) 2016 summer (blue contour \((-\)) starts at $-40 \, \mu\text{g m}^{-3}$, while red contour \(+) ends at $20 \, \mu\text{g m}^{-3}$ both with an interval of 20 $\mu\text{g m}^{-3}$), and in (b) 2016 autumn (blue contour \((-\)) starts at $-60 \, \mu\text{g m}^{-3}$, while red contour \(+) ends at $60 \, \mu\text{g m}^{-3}$ both with an interval of 30 $\mu\text{g m}^{-3}$).