



## A new approach of air pollution regionalization based on geographically weighted variations for multi-pollutants in China



Peipei Qiu <sup>a</sup>, Lin Zhang <sup>b,\*</sup>, Xuesong Wang <sup>a</sup>, Yafei Liu <sup>c</sup>, Shuai Wang <sup>d</sup>, Sunling Gong <sup>e</sup>, Yuanhang Zhang <sup>a,\*</sup>

<sup>a</sup> State Key Joint Laboratory of Environmental Simulation and Pollution Control, College of Environmental Sciences and Engineering, Peking University, Beijing 100871, China

<sup>b</sup> Laboratory for Climate and Ocean-Atmosphere Studies, Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China

<sup>c</sup> State Key Laboratory of Water Environment Simulation, School of Environment, Beijing Normal University, Beijing 100875, China

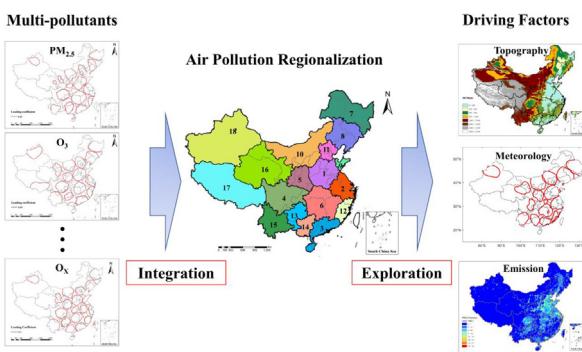
<sup>d</sup> China National Environmental Monitoring Centre, Beijing 100012, China

<sup>e</sup> State Key Laboratory of Severe Weather, Key Laboratory of Atmospheric Chemistry of CMA, Chinese Academy of Meteorological Sciences, Beijing 100081, China

### HIGHLIGHTS

### GRAPHICAL ABSTRACT

- We present the Geographically Weighted Rotation Empirical Orthogonal Function approach for air pollution regionalization.
- We propose an integrated air pollution regionalization result that can be applied to multiple pollutants and different years.
- PM<sub>2.5</sub>, O<sub>3</sub>, and O<sub>x</sub> correlations are analyzed for regional-specific and coordinated control measures.



### ARTICLE INFO

Editor: Jianmin Chen

#### Keywords:

Air pollution regionalization  
GWREOF  
Multi-pollutants  
Spatial-temporal variations  
PM<sub>2.5</sub> and ozone

### ABSTRACT

Air pollution regionalization is a key and necessary action to identify pollution regions for implementing control measures. Here we present a new approach called Geographically Weighted Rotation Empirical Orthogonal Function (GWREOF) for air pollution regionalization in China. Compared with previous methods, such as EOF, REOF, and K-mean, GWREOF better accounts for the variability of air pollution conditions driven by emission patterns and meteorology with centralized spatial locations. We apply GWREOF to multiple air pollutants (such as PM<sub>2.5</sub>, O<sub>3</sub>, and other monitored air pollutants) and air quality metrics using their measured spatial and temporal variations in 337 Chinese cities over 2015–2020. We find that the regionalization results for different air pollutants are highly similar, primarily determined by topography and meteorological conditions in China. Therefore, we propose an integrated regionalization result, which identifies 18 air pollution control regions in China and can be applied to multiple pollutants and different years. We further analyze PM<sub>2.5</sub>, O<sub>3</sub>, and O<sub>x</sub> (O<sub>3</sub> + NO<sub>2</sub>) pollution levels and their correlations in these regions. PM<sub>2.5</sub> and O<sub>3</sub> correlations are generally strongly positive in southern China while negative in northern China. However, PM<sub>2.5</sub> and O<sub>x</sub> correlations are broadly positive in China, reflecting the crucial role of atmospheric oxidizing capacity. Regional-specific and coordinated control measures are in need as China's air pollution strategy transits from PM<sub>2.5</sub>-focused to PM<sub>2.5</sub>-O<sub>3</sub> synergic control.

### 1. Introduction

\* Corresponding authors.

E-mail addresses: [zhanglg@pku.edu.cn](mailto:zhanglg@pku.edu.cn) (L. Zhang), [yhzhang@pku.edu.cn](mailto:yhzhang@pku.edu.cn) (Y. Zhang).

Air pollution, including both PM<sub>2.5</sub> (fine particle matter with an aerodynamic diameter of smaller than 2.5 μm) and ozone (O<sub>3</sub>) pollution, has

drawn considerable academic attention and raised substantial human health concerns (Fan et al., 2020; Guo et al., 2017; Kuerban et al., 2020; Lu et al., 2018; Silver et al., 2018; Ye et al., 2018). With the rapid development of urbanization and industrialization, air pollution has shown regionalized agglomeration characteristics (Ma et al., 2016). Geographically adjacent cities with similar topography and meteorological conditions tend to show similar temporal variations in air pollution (Chen et al., 2020). Due to atmospheric residential times of several days, PM<sub>2.5</sub> and O<sub>3</sub> events often occur at large regional scales. Severe PM<sub>2.5</sub> pollution events have been frequently observed in major megacity clusters of China, such as Beijing-Tianjin-Hebei (BTH), Yangtze River Delta (YRD), and Sichuan Basin (SCB) during wintertime (Mao et al., 2019; Wang et al., 2020; Wang et al., 2019). Moreover, surface O<sub>3</sub> concentrations are rising rapidly across the country, especially in the YRD and Pearl River Delta (PRD), where O<sub>3</sub> has overtaken PM<sub>2.5</sub> as the prime pollutant (Li et al., 2020; Lu et al., 2018; Lu et al., 2020). Recent measurements also showed that the regional extent and duration of ozone air pollution have expanded in the summer or autumn (Li et al., 2021).

One of the most influential and essential strategies to address regional air pollution issues is identifying the pollution control regions to facilitate coordinated control measures. For policymakers to develop a targeted and regional pollution control strategy, air pollution control regions were first determined qualitatively. Dated back to 1968, the *California Air Resources Board* (CARB) divided California State into 15 air basins based on topography and meteorological conditions (<https://www.arb.ca.gov/>). The U.S. *Environmental Protection Agency* (EPA) later divided the U.S. into ten control areas based on administrative management (<https://www.epa.gov/>; Schmalensee and Stavins, 2019). Both actions have significantly facilitated air quality improvements in the U.S. and California (Dedoussi et al., 2020). In Europe, the *European Union Air Quality Legislation* has required member states to establish pollution zones by considering air pollution levels since 1984 (Holman et al., 2015; Marco and Bo, 2013). In recent years, the Chinese government has designated key control regions or city clusters based on social-economic development and pollution levels, such as BTH, YRD, and PRD, where the *Air Pollution Prevention and Control Action Plan* focused on (Feng et al., 2019). These qualitative regionalization methods generally only consider individual factors (e.g., pollution level, administrative management, topography, and meteorology).

Quantitative regionalization methods by clustering or zoning locations with similar variability of targeted variables have been applied to identify climatic zones, such as Empirical Orthogonal Function (EOF) (also called Principal Component Analysis; PCA), Rotated Empirical Orthogonal Function (REOF), and K-means clustering (Carro-Calvo et al., 2017; Carvalho et al., 2016; Hannachi et al., 2007; Lyapina et al., 2016; Pražníkár, 2017; Vienneau and Briggs, 2013). These clustering and regionalization methods have recently helped identify air pollution control regions within which pollution conditions have relatively similar characteristics (Cheng et al., 2018; Wang et al., 2015; Yao et al., 2020; Zhang et al., 2012). For example, Zhang et al. (2012), by using the REOF method, identified four major haze regions based on visibility measurements from 1957 to 2005. Wang et al. (2015) and Cheng et al. (2018), also using the REOF method, clustered the regions for PM<sub>2.5</sub> in the year 2014 and ozone air pollution during 2015–2017 in China. More recently, Yao et al. (2020) identified 17 affinity zones for PM<sub>2.5</sub> with 2013–2017 surface measurements by combining REOF and the K-medoids clustering methods and further accounting for the geographical locations of measurements. These studies can provide valuable results on the regionalization of PM<sub>2.5</sub> or O<sub>3</sub>, but for a limited time or a single air pollutant. In addition, the traditional REOF and clustering methods do not consider spatial distances among measurement locations, which often induce unregular spatial results (see Section 3.2 of this study).

In this study, we propose a new approach called Geographically Weighted Rotation Empirical Orthogonal Function (GWREOF) that accounts for geographically weighted spatial and temporal variations of pollution levels. We compare the air pollution regionalization results derived by GWREOF with those by REOF and find that the GWREOF approach

can better identify the centralized spatial patterns of air pollution. We apply the approach for air pollution regionalization using nationwide measurements of multiple air pollutants during 2015–2020 in China. This study aims to develop a regionalization outcome that is testified by climatic conditions and multiple air pollutants and thus help policymakers establish coordinated pollution prevention and control measures.

## 2. Data and methodology

### 2.1. Datasets

The China National Environmental Monitoring Center (CNEMC) started releasing hourly average values of the six pollutants (including PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, CO, and O<sub>3</sub>) nationwide in 2013. The air pollution concentrations were monitored in 74 cities in 2013, increasing to 161 cities in 2014 and 337 cities in 2015. The spatial locations of 337 cities are shown in Fig. S1. Here the six-year (from 2015 to 2020) hourly measurements of the six pollutants are accessed from the CNEMC (data available at <https://quotsoft.net/air/>). Each city has multiple sites; here, we average them to represent the city-level air pollution condition. Daily mean PM<sub>2.5</sub> (or PM<sub>10</sub>, SO<sub>2</sub>, NO<sub>2</sub>, and CO) concentrations and maximum daily 8-h average ozone concentrations (O<sub>3</sub>-8H) are calculated. A valid daily mean concentration is calculated for at least 20 hourly data during the day. The O<sub>3</sub>-8H is the maximum value of the 8-h moving average at each site, and a valid 8-h moving average shall have at least 6 of every 8 h. All data are processed following the Ambient Air Quality Standards (GB3095-2012) (<https://english.mee.gov.cn/Resources/standards>). Here we calculate the air quality index (AQI) and Air Quality Composite Index (AQCI) using the six pollutants. The calculation of AQI follows the Technical Regulation on Ambient Air Quality Index (HJ 633-2012) ([https://english.mee.gov.cn/Resources/standards/Air\\_Environment/](https://english.mee.gov.cn/Resources/standards/Air_Environment/)). The AQCI is the sum of the six pollutant concentrations' quotients divided by corresponding secondary standard values. Furthermore, the particle matter index (PMI, including PM<sub>2.5</sub> and PM<sub>10</sub>) and total primary gas pollutants (PGP, including SO<sub>2</sub>, NO<sub>2</sub>, and CO) are constructed by first normalizing their components and then normalizing the sum. We use PMI and PGP as integrated metrics denoting, respectively, particle and gaseous pollution levels.

We further analyze meteorological conditions, topography, anthropogenic emissions, and population to support the regionalization results. Meteorological reanalysis data for 2015–2020 is obtained from the National Centers for Environmental Prediction (NCEP) Final Analysis (FNL) Operational Global Analysis data (<https://rda.ucar.edu/datasets/ds083.2/>). The daily average values of meteorological parameters are used, including near-surface air temperature (T), relative humidity (RH), wind speed (WS), and planetary boundary layer height (PBLH). The NCEP meteorological datasets have a 1° × 1° spatial resolution, and here we sample the daily values at the 337 city centers in China. We have evaluated the reanalysis data with the meteorological measurements at the 337 cities obtained from the China Meteorological Data Network (<http://data.cma.cn/>) over 2015–2020. The NCEP reanalysis data are highly consistent with the meteorological measurements, with correlation coefficients above 0.85 for all variables (figure not shown).

We also use the nationwide daily visibility measurements over 2015–2020. Measurements of daily visibility in China are downloaded from the National Earth System Science Data Center, National Science & Technology Infrastructure of China (<http://www.geodata.cn>). More details of the dataset can be found in Mao et al. (2019).

The anthropogenic emissions of sulfur dioxide (SO<sub>2</sub>), nitrogen dioxide (NO<sub>2</sub>), and primary PM<sub>2.5</sub> are downloaded from the Tsinghua University Multi-resolution Emission Inventory for China (MEIC) inventory ([www.meicmodel.org](http://www.meicmodel.org)) for the year 2017. The MEIC dataset is a widely used bottom-up anthropogenic emission inventory for air quality studies in China (Li et al., 2017). The national topographic data at ~1 km spatial resolution is accessed from the Resource and Environmental Data Cloud Platform (<http://www.gscloud.cn>) obtained by the Shuttle Radar Topography Mission (SRTM) of the Space Shuttle of the United States Endeavour. The

national population density data at ~1 km spatial resolution is downloaded from Socioeconomic Data and Applications Center (SEDAC, <https://sedac.ciesin.columbia.edu/>). The economic dataset (including gross domestic product (GDP) and GDP per capita of provinces and cities) for the year 2020 is collected from China Statistical Yearbooks Database (<https://data.cnki.net/Yearbook>).

## 2.2. Methodology

Regionalization generally aims to find highly correlated spatial regions or zones where the pollution variations show consistency or similarity. Several clustering approaches, such as Empirical Orthogonal Function (EOF) (also called Principal Component Analysis, PCA), Rotated Empirical Orthogonal Function (REOF), and K-means clustering, have been applied in previous studies to regionalize air pollution conditions (Carro-Calvo et al., 2017; Carvalho et al., 2016; Gao et al., 2011; Lyapina et al., 2016; Pražníkár, 2017; Vienneau and Briggs, 2013). These approaches have different advantages and disadvantages. For example, the EOF (or PCA) approach decomposes a space-time field into spatial patterns and associated time series. The spatial patterns can provide the basis for regionalization, but it is difficult to interpret the physical meanings of spatial patterns because of their global features and orthogonality. The K-means method can cluster measurement locations (sites or grid cells) with similar temporal changes into one group. As an unsupervised learning method, the clustering K-means results are sensitive to the initial setting parameters (Govender and Sivakumar, 2020). Although REOF could compensate for the drawbacks of EOF or PCA, this method cannot deal with the spatial heterogeneity or non-stationarity (Hannachi et al., 2007). Most importantly, all these methods only use the correlation of the space-time field of a specific variable and tend to neglect its spatial information. Therefore, based on the REOF method, a new approach was developed, and the details are described below.

### 2.2.1. REOF and geographically weighted REOF

Here we describe both REOF and Geographically Weighted REOF (GWREOF) approaches. REOF (or Rotated principal component analysis) is commonly used for extracting the high-correlated spatial pattern (also called spatial mode) and capturing the principal temporal variations (also called time coefficients) of time series. The spatial patterns can then be used as the basis for regionalization. This method has previously been applied in the meteorological science and climatology (Hannachi et al., 2007; Zhou and Liu, 2018) and recently used for the air pollution regionalization (Cheng et al., 2018; Wang et al., 2015; Yao et al., 2020). The REOF method was employed to extract representative spatial patterns from the space-time field to approximate the clustering of sites or cities with similar time series variations. REOF can be expressed by Eq. (1), which decomposes a normalized space-time field ( $X_{mn}$ ) into spatial patterns and the principal time series.

$$X_{mn} = V_{mp}^* T_{pn}^* + \varepsilon_{mn} \quad (1)$$

$$R = X_{mn} X_{mn}^T = V \Lambda V^T \quad (2)$$

$$V_{mp}^* = \text{Rotated}\left(V_{mp} \Lambda_p^{1/2}\right) \quad (3)$$

In Eq. (1),  $p$  is the number of spatial modes,  $m$  is the number of measurement locations,  $n$  is the number of measurements at each location, and  $\varepsilon_{mn}$  represents the error matrix.  $V_{mp}^*$  represents the loading coefficients of spatial modes, where high values mean highly correlated areas.  $T_{pn}^*$  represents the time series of the spatial modes.  $V_{mp}^*$  is calculated by Eqs. (2) and (3), where the correlation matrix  $R$  (correlation matrix between locations) can be decomposed into eigenvectors  $V$  and eigenvalues  $\Lambda$ . The spatial modes  $V_{mp}^*$  are then estimated after the VARIMAX rotation (i.e., varimax orthogonal rotation) as denoted by Eq. (3). The spatial modes  $V_{mp}^*$  with high

loading coefficients can distinguish the regions with similar temporal variation and are used for the subsequent regionalization.

The REOF method has a limitation: it cannot effectively integrate geographic information of sites, resulting in irregular and non-adjacent spatial modes (Hannachi et al., 2007; Vienneau and Briggs, 2013; Yao et al., 2020). Air pollution levels can be affected by large-scale atmospheric circulations or have similar temporal variations (e.g., seasonal variations) on a broad spatial scale. That may lead to large-scale spatial mode features unsuitable for clear regionalization. A geographic weighting may help solve the limitation, as previously suggested by Harris et al. (2011), in which they developed geographically weighted PCA (GWPCA) to explore local features of population characteristics in Greater Dublin. This geographic weighting approach has not been applied to air pollution regionalization analyses. We propose in this study a new method named the Geographically Weighted REOF (GWREOF) method to get more localized spatial patterns, which allows a more precise division. The GWREOF method is described by Eqs. (4)–(6). Instead of decomposing  $R$ , a geographically weighted matrix  $K$  is constructed as below.

$$K = R \cdot W = X_{mn} X_{mn}^T \cdot W \quad (4)$$

Here, matrix  $K$  is calculated by the element-wise product of correlation matrix  $R$  and geographically weighted matrix  $W$ .  $X_{mn}$  is the normalized data matrix with  $m$  rows representing the number of observational locations and  $n$  columns representing the time series of air pollution concentrations, and  $W$  is a geographical weight matrix given as:

$$w_{ij} = \exp\left(-d_{ij}^2/h^2\right) \quad (5)$$

$$d_{ij} = R_e * \arccos(\sin v_i \cdot \sin v_j + \cos v_i \cdot \cos v_j \cdot \cos(u_i - u_j)) \quad (6)$$

Where  $w_{ij}$  represents the spatial weight between two locations  $i$  and  $j$ , which is the distance decay function, and it is calculated by the real spherical distance  $d_{ij}$  and the selected bandwidth  $h$ . The real spherical distance  $d_{ij}$  is calculated by Eq. (6), with  $R_e$  representing the earth's radius, and  $u$  and  $v$  denoting their latitudes and longitudes, respectively. The spatial modes are then calculated using Eqs. (1) and (3).

Fig. S2 shows a schematic diagram of the geographical weight between two locations and the GWREOF calculation process. Here we apply the GWREOF approach to different air pollutant concentrations and air quality indices. For air pollutants, including PM<sub>2.5</sub>, PM<sub>10</sub>, SO<sub>2</sub>, CO, NO<sub>2</sub>, O<sub>3</sub>–8H, O<sub>x</sub>, and indices (AQI, ACQI), the normalized space-time field  $X_{mn}$  is constructed by their daily values. For total particle matters (TPM) and total primary pollutants (TPP), the normalized space-time field  $X_{mn}$  is not constructed by the sum of pollutant concentrations but by normalizing their components and then normalizing the sum. The calculation of  $X_{mn}$  is proposed by the three-dimensional factor analysis method, as the concentrations of these components are highly and positively correlated (Shi et al., 2017; Tian et al., 2016).

### 2.2.2. Determination of the geographic weighting parameters

As described above, applications of the GWREOF will need to determine two parameters: the number of spatial modes to represent the original field and the distance bandwidth ( $h$  in Eq. (5)) for the geographic weighting. To define the optimal number of modes, we calculate each mode's variance contribution and the cumulative variance contribution. In general, the number of required modes was determined by the accumulated variance reaching >80% (Jolliffe and Cadima, 2016; Wang et al., 2015). In this study, we use the total variance contribution to determine the optimal number of spatial modes. We test the variance explained by each spatial mode by analyzing the PM<sub>2.5</sub> measurements over 2015–2020 with the GWREOF method. As shown in Supplementary Fig. S2c, the variances explained decrease rapidly as the rank of spatial modes increases. Summing up about 20 modes, we find that the cumulative variance of spatial modes can

reach over 84%. Thus, in this study, the first 20 spatial modes diagnosed by GWREOF are used for subsequent regionalization analyses.

After selecting the optimal number of spatial modes, we further determine the optimal bandwidth  $h$ . We test the bandwidth values ranging from 300 to 1200 km, as shown in Fig. S2d. By using a bandwidth of 800 km, the spatial modes can explain the original field to the greatest extent of variance. In the method, the bandwidth  $h$  could determine the spatial coverage and regional boundaries of air pollution regionalization. The determination of optimal bandwidth can depend on the number of spatial modes. When the number of spatial modes is determined as 20, a bandwidth of 800 km represents the overall spatial coverage of 20 modes covering almost all measurement sites of China.

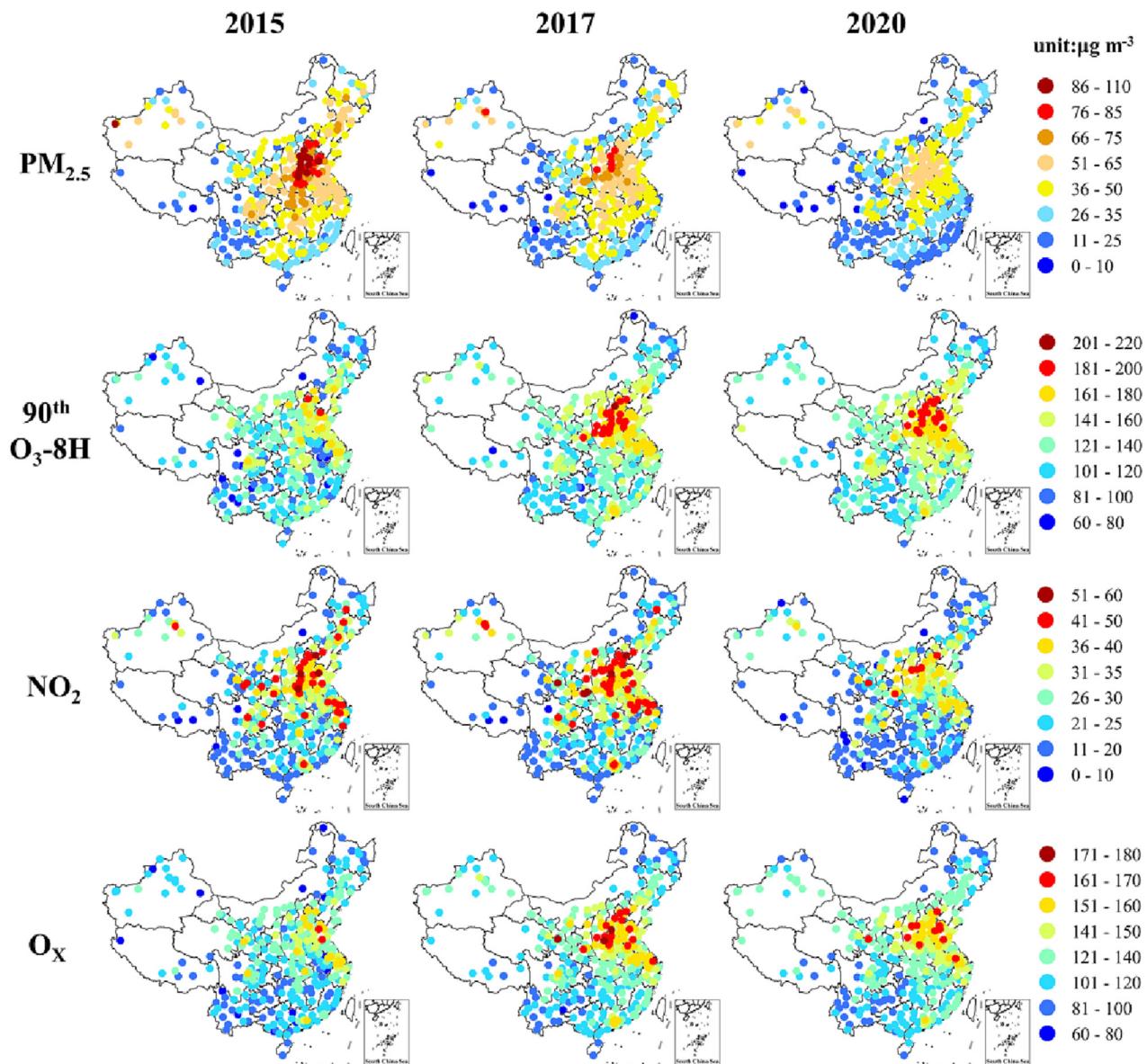
### 3. Results and discussions

#### 3.1. Spatial-temporal variation of air pollution across China

We first analyze the spatial and temporal variations of air pollutants concentrations measured in 337 Chinese cities from 2015 to 2020. Fig. 1 shows the spatial distributions of annual mean concentrations of  $\text{PM}_{2.5}$ ,  $\text{NO}_2$ , and

$\text{O}_x$ , as well as the annual 90th percentile of  $\text{O}_3\text{-}8\text{H}$  concentrations. Fig. S3 shows the interannual variations of national mean concentrations. Since 2013, the Chinese government has implemented stringent air pollution control policies, including the 'Air Pollution Prevention and Control Action Plan' and 'Three-year action plan for cleaner air', which have significantly reduced the emissions of primary air pollutants and thus improved air quality. Concentrations of air pollutants such as  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ,  $\text{SO}_2$ ,  $\text{NO}_2$ , and  $\text{CO}$ , except for  $\text{O}_3$ , show strong decreasing trends (Fig. S3). The national annual mean  $\text{PM}_{2.5}$ ,  $\text{PM}_{10}$ ,  $\text{SO}_2$ ,  $\text{NO}_2$ , and  $\text{CO}$  concentrations are, respectively,  $33 \mu\text{g m}^{-3}$ ,  $56 \mu\text{g m}^{-3}$ ,  $10 \mu\text{g m}^{-3}$ ,  $24 \mu\text{g m}^{-3}$ , and  $1.3 \text{ mg m}^{-3}$  in 2020, and decreased by 28.26 %, 27.27 %, 56.52 %, 11.11 %, and 31.58 % compared with 2015. However, the average 90th percentile  $\text{O}_3\text{-}8\text{H}$  concentration increased by 12.20 %, likely caused by insufficient and uncoordinated emission controls for VOCs (Li et al., 2017; Li et al., 2019).

For the spatial distributions, concentration levels of air pollutants, in general, are higher in eastern China than in western China. High polluted regions are found in central and eastern China, particularly in the BTH and surrounding areas, Fenwei Plain (FWP), YRD, and SCB (Fig. 1). These  $\text{PM}_{2.5}$  heavily polluted regions also tend to show severe  $\text{O}_3$  pollution as indicated by the high 90th percentile  $\text{O}_3\text{-}8\text{H}$  concentrations. Although the



**Fig. 1.** Spatial distributions of annual mean  $\text{PM}_{2.5}$  (top panels), the 90th percentile  $\text{O}_3\text{-}8\text{H}$  (second row), annual mean  $\text{NO}_2$  (third row), and mean  $\text{O}_x$  (bottom panels) concentrations measured at the 337 cities from 2015 to 2020.

concentrations of different air pollutants show different interannual changes, their spatial distribution patterns are relatively stable over 2015–2020. As discussed below, the spatial distributions of air pollution levels may be primarily determined by topography, meteorological conditions, population density, and emission levels.

Distinct seasonal variations are also seen for various air pollutants (Supplementary Fig. S4). Severe PM<sub>2.5</sub> pollution events have been frequently observed in winter or spring. In contrast, high O<sub>3</sub> pollution events always happen in summer and autumn. As shown in Fig. S4, the highest PM<sub>2.5</sub> concentrations appeared in winter ( $61 \pm 28 \mu\text{g m}^{-3}$ , December to February), followed by spring ( $36 \pm 12 \mu\text{g m}^{-3}$ , March to May) and autumn ( $35 \pm 13 \mu\text{g m}^{-3}$ , September to November), with the lowest in summer ( $23 \pm 8 \mu\text{g m}^{-3}$ , June to August). Much heavier PM<sub>2.5</sub> air pollution in North China in winter than in summer is associated with poorer atmospheric ventilation and larger pollutants emissions from the heating supply there (An et al., 2019; Gao et al., 2018; Liu et al., 2022). For ozone, the highest 90th percentile O<sub>3</sub>-8H concentrations appeared in summer ( $156 \pm 34 \mu\text{g m}^{-3}$ ), followed by spring ( $146 \pm 23 \mu\text{g m}^{-3}$ ) and autumn ( $131 \pm 28 \mu\text{g m}^{-3}$ ), with the lowest in winter ( $92 \pm 12 \mu\text{g m}^{-3}$ ). The seasonal averaged O<sub>3</sub> levels peak in different seasons for different areas: about 64 % of Chinese cities have the highest O<sub>3</sub> concentrations in summer, while 16 % of cities in Southwest China occur in spring and about 20 % in South and central China in autumn. This may be driven by a combined effect of ozone precursors' emissions and meteorological conditions (Li et al., 2020; Yin et al., 2019).

The spatial and seasonal variations of measured surface NO<sub>2</sub> concentrations share similar patterns as PM<sub>2.5</sub>, with the highest concentrations mainly occurring in winter (national mean value of  $34 \pm 12 \mu\text{g m}^{-3}$ ) and the lowest value in summer ( $19 \pm 7 \mu\text{g m}^{-3}$ ). As NO<sub>2</sub> can be quickly photolyzed to form ozone, O<sub>x</sub> (O<sub>3</sub> + NO<sub>2</sub>) is often used as an indicator of atmospheric oxidizing capacity. The highest concentrations of O<sub>x</sub> mainly occurred in summer, with a mean value of  $138 \pm 32 \mu\text{g m}^{-3}$ , followed by spring ( $137 \pm 21 \mu\text{g m}^{-3}$ ) and autumn ( $122 \pm 22 \mu\text{g m}^{-3}$ ), and winter ( $105 \pm 12 \mu\text{g m}^{-3}$ ), similar to O<sub>3</sub> but with a smaller seasonal variation. The high values of O<sub>x</sub> in spring and summer are likely associated with temperature and solar radiation conditions (Li et al., 2021; Wang et al., 2022). These similarities and differences in spatial and temporal variations among

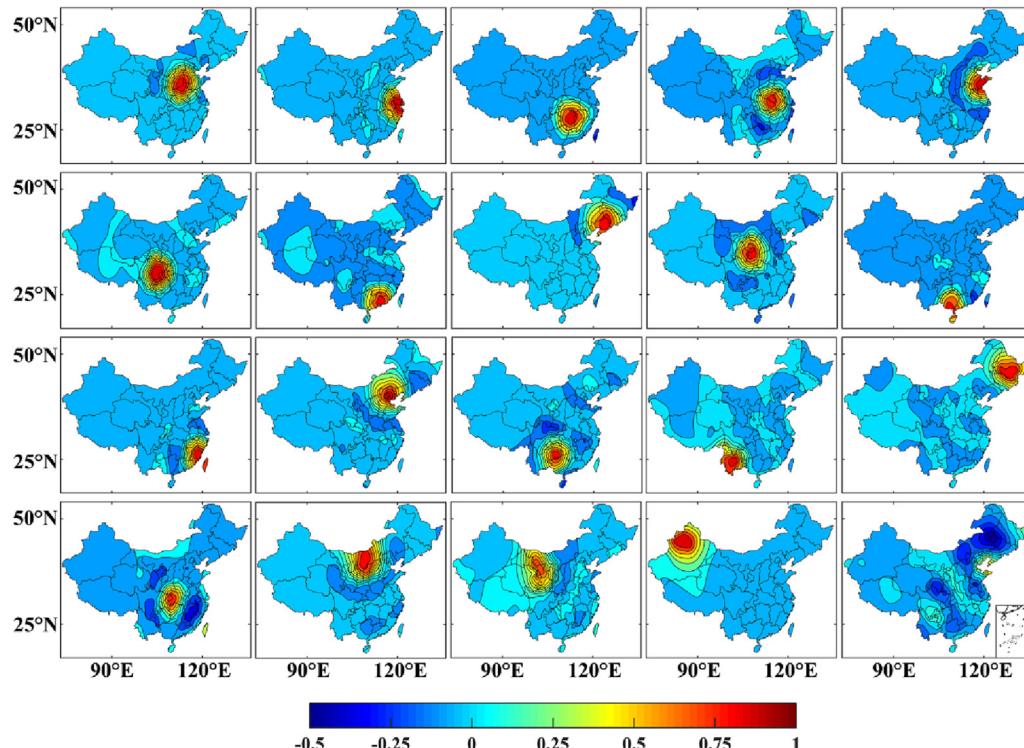
air pollutants shall be accounted for when developing a climatological regionalization.

### 3.2. Regionalization effects based on GWREOF and its comparison with REOF

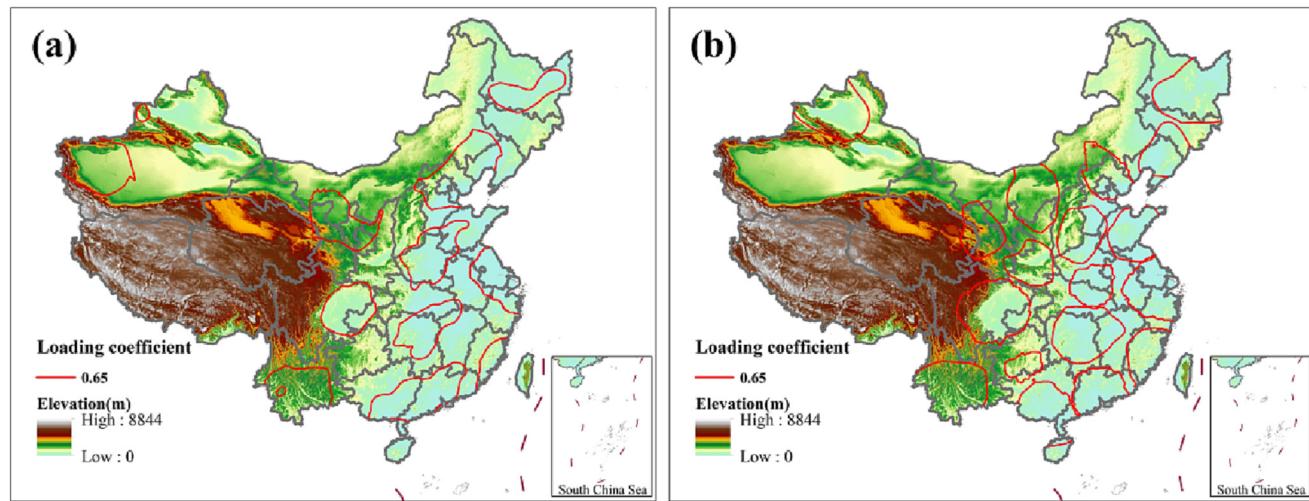
The spatial and temporal variations in air pollution concentrations can be analyzed using the regionalization approaches so that the cities with similar pollution concentration variations are clustered into one region. Here we compare the regionalization results from the REOF and GWREOF approaches to demonstrate the advance of GWREOF. Fig. 2 shows the distribution of loading coefficients for the first 20 spatial modes when analyzing the whole 2015–2020 PM<sub>2.5</sub> measurements at 337 Chinese cities by the GWREOF method. For comparison, Fig. S5 shows the results analyzed by REOF. For the first 20 spatial modes, GWREOF identifies 19 modes with distinct spatial patterns with high loading coefficient centers. These 19 modes can explain 82.5 % of the total variance in the PM<sub>2.5</sub> measurements. In contrast, among the 20 spatial modes analyzed by REOF (Fig. S5), only 15 modes have clear spatial patterns and are much less centralized than the GWREOF modes. Less percentage (72.5 %) of the total variance is explained by the 15 REOF modes.

Such differences in spatial modes can also be seen when regionalizing the nationwide O<sub>3</sub>-8H concentrations. Both GWREOF and REOF methods are applied to the 2015–2020 O<sub>3</sub>-8H concentrations at the 337 Chinese cities, and the loading coefficients of the first 20 spatial modes are, respectively, shown in Fig. S6 and Fig. S7. The GWREOF-derived spatial modes for O<sub>3</sub>-8H show centralized spatial patterns. The spatial modes derived by REOF, however, show much broader features. For example, the first mode analyzed by REOF, which explains 35 % of the total measured variance, shows similar loading coefficients covering the whole of northern China. Such a broad spatial mode may be largely driven by the sensitivity of ozone to meteorological variables, such as temperature.

Therefore, by comparing the spatial modes analyzed separately by the two methods (i.e., REOF method and GWREOF method), the spatial patterns based on GWREOF have more clear boundaries and centralized loading coefficients. Typically, the high loading coefficients are determined using the threshold value of greater than or equal to 0.6 (Wang et al., 2015; Yao et al.,



**Fig. 2.** Loading coefficients of the first 20 spatial modes as decomposed by the GWREOF method for 2015–2020 surface PM<sub>2.5</sub> measurements at the 337 Chinese cities.



**Fig. 3.** Comparison of the regionalization results derived by (a) the REOF method and (b) the GWREOF method, based on surface PM<sub>2.5</sub> measurements at the 337 Chinese cities over 2015–2020. The threshold value of 0.65 is used to identify the regional boundaries of spatial modes shown in Fig. 2.

2020). Here we tested threshold values around 0.6 and found that a value of 0.65 best identified the spatial modes with small overlapping.

Fig. 3 shows the regionalization results of PM<sub>2.5</sub> by using a threshold loading coefficient of 0.65 as the regional boundaries. Compared with REOF results, the GWREOF method based on integrating the geographic distance information between sites could make the clustering results more spatially explicit, and the boundaries of spatial patterns are more evident and regular. Therefore, the GWREOF method will be applied to regionalize all pollutants in the subsequent analysis.

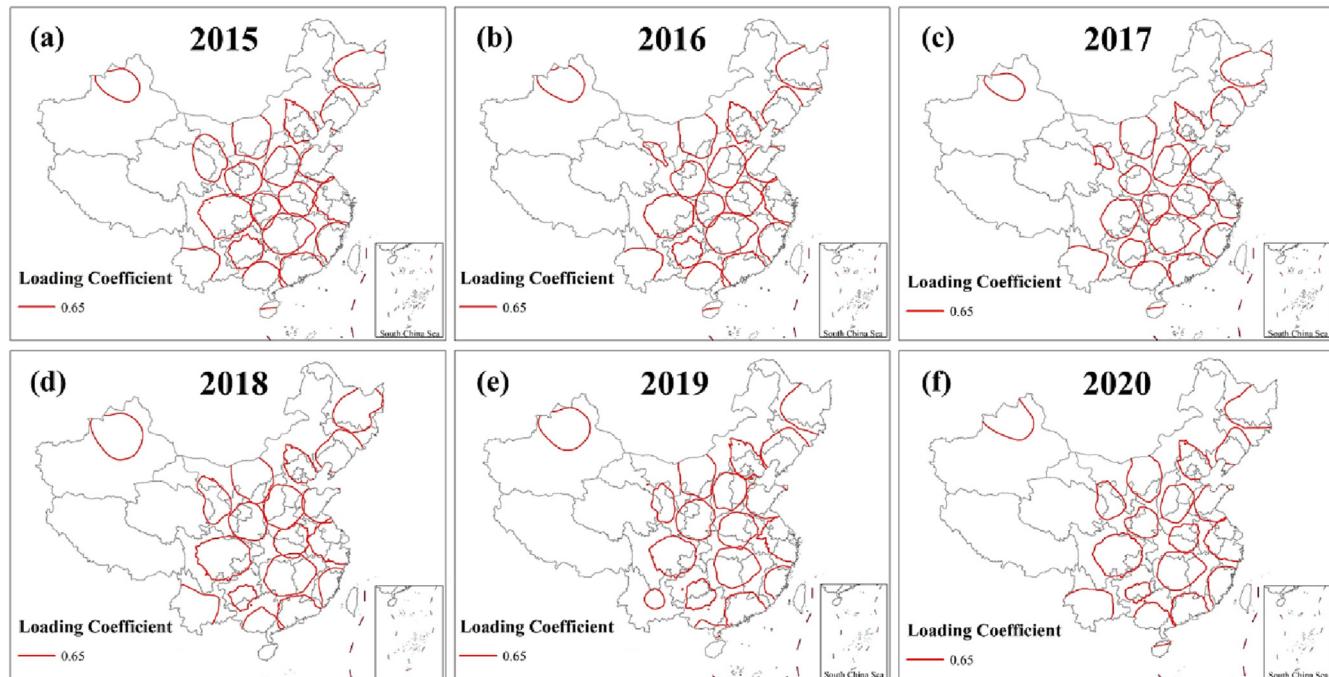
### 3.3. Geographically weighted regionalization results for air pollutants in China

#### 3.3.1. Annual and seasonal regionalization results for PM<sub>2.5</sub>

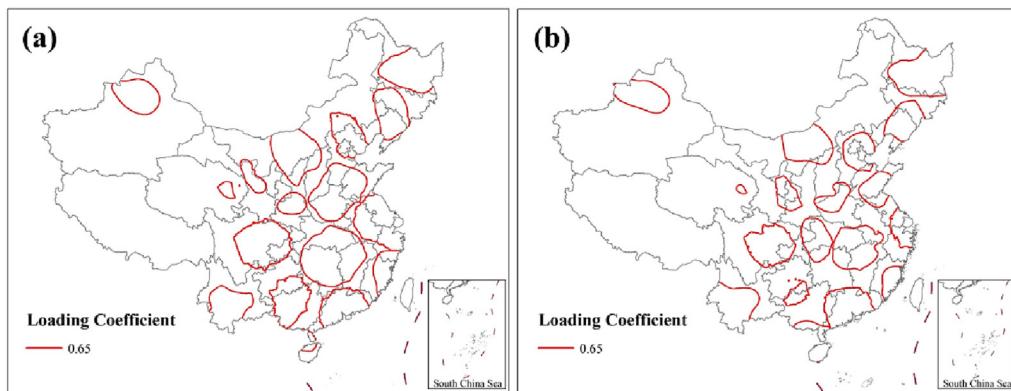
As shown in Fig. 3 above, the loading coefficient values of spatial modes can provide boundaries for air pollution regionalization. To explore the

interannual stability of spatial modes, we extract the spatial modes of PM<sub>2.5</sub> spatiotemporal field as derived by the GWREOF method in individual years of 2015–2020. Results are shown in Fig. 4. The 19 spatial modes now correspond to 19 regions identified using a threshold loading coefficient of 0.65, as described in the section above. Comparing the multi-year results (Fig. 3) with the individual-year results (Fig. 4), we can see that the distributions of spatial modes in different years are highly consistent. Nearly all 19 regions show similar loading coefficient high centers, with some differences in their spatial coverage. These regions include the key populated areas, such as North China Plain (NCP), Yangtze River Delta (YRD), Pearl River Delta (PRD), and Sichuan Basin (SCB), as will be discussed later.

Although the PM<sub>2.5</sub> concentrations have substantially decreased over 2015–2020, the spatial modes estimated by the GWREOF method do not show much change. Such interannual stability of spatial modes can reflect relatively stable PM<sub>2.5</sub> temporal variations in these regions. In addition,



**Fig. 4.** Spatial modes with high loading coefficients for PM<sub>2.5</sub> concentrations across China in individual years for 2015–2020. The threshold value of 0.65 for high loading coefficient is used to identify the regional boundaries of spatial modes.



**Fig. 5.** Comparison of the PM<sub>2.5</sub> regionalization results during (a) the pollution season (November and December, January and February) and (b) the clean season (May to August) over 2015–2020.

even though PM<sub>2.5</sub> levels have significantly reduced, their spatial distributions remain relatively unchanged, as shown in Fig. 1. The interannual stability of spatial modes provides valuable information to develop a climatological air pollution regionalization.

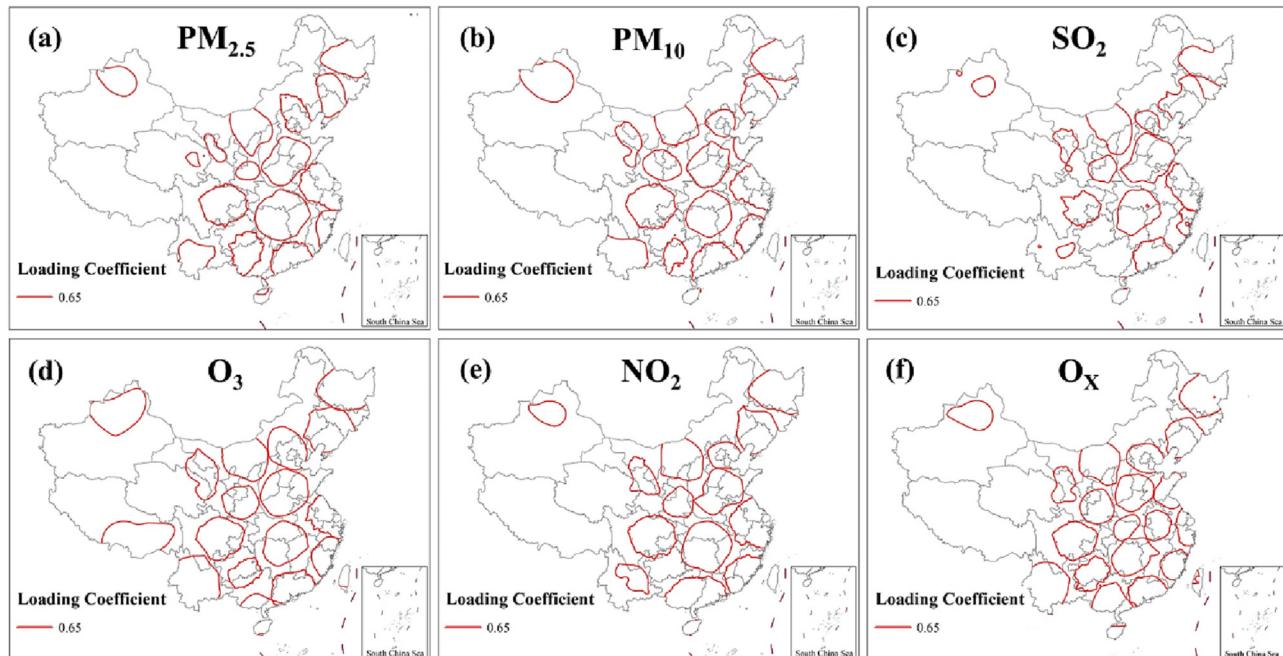
The PM<sub>2.5</sub> air pollution levels have significant seasonal variations, and thus it is necessary to investigate regionalization results for different seasons. As for PM<sub>2.5</sub>, we analyze the spatial-temporal variability of PM<sub>2.5</sub> levels in the pollution season (November and December, January and February) and in the clean season (May to August). We then calculate their spatial modes using the GWREOF method and compare them in Fig. 5. The main spatial modes for the two seasons are highly similar but also have some differences in the spatial coverages. In general, the spatial modes for the pollution season cover relatively larger areas than the results for the clean season, as would be expected because heavy PM<sub>2.5</sub> pollution frequently occurs at a large regional scale during the pollution season, as PM<sub>2.5</sub> has a residence time of several days to weeks in the air (Chen et al., 2020).

### 3.3.2. Regionalization results for other air pollutants

We extend our analysis to other air pollutants and summarize the regionalization results here. The regionalization results for other primary

pollutants (PM<sub>10</sub>, NO<sub>2</sub>, and SO<sub>2</sub>) and total oxidant O<sub>x</sub> have been investigated using the GWREOF method. Fig. 6 shows the distributions of spatial modes for different air pollutants, with 14–18 high-loading coefficient centers identified. We can see that the spatial modes for the different air pollutants are highly similar. For example, the key populated areas, such as NCP, YRD, SCB, etc., are all identified for the different air pollutants, although their spatial coverages vary to some degree. Due to the scarcity of measurement sites in western China, only one spatial mode centered in southern Xinjiang is identified for the different air pollutants. For the primary pollutants, such as PM<sub>2.5</sub>, PM<sub>10</sub>, NO<sub>2</sub>, and SO<sub>2</sub>, we find the spatial and temporal variations of their concentrations are highly correlated at the regional scale (Fig. 1 and Fig. S4), thus driving the similarities in their spatial modes (Chu et al., 2020).

The spatial modes for O<sub>3</sub> and O<sub>x</sub> also show similar patterns of the high loading coefficient centers, even though the concentrations of PM<sub>2.5</sub> and ozone are sometimes negatively correlated (Fig. S4). We will discuss in Section 4 that the PM<sub>2.5</sub> vs. ozone correlations vary in different regions of China, while the PM<sub>2.5</sub> vs. O<sub>x</sub> correlations tend to be positive. We further analyze the regionalization results based on air quality indices, including Air Quality Index (AQI), Air Quality Composite Index (AQCI), total



**Fig. 6.** Spatial modes analyzed by the GWREOF method for (a) PM<sub>2.5</sub>, (b) PM<sub>10</sub>, (c) SO<sub>2</sub>, (d) O<sub>3</sub>–8H, (e) NO<sub>2</sub>, and (f) O<sub>x</sub>. Spatial modes are calculated with 2015–2020 all-year measurements except for O<sub>3</sub> with measurements during ozone pollution season.

Particulate Matter Index (PMI), and total Primary Gas Pollutant (PGP). The results support a consistent regionalization based on the different metrics (Fig. S8). The patterns of their spatial modes are similar to those of  $\text{PM}_{2.5}$ . As the predominant pollutant in most Chinese cities over 2015–2020,  $\text{PM}_{2.5}$  levels largely contribute to the variations of air quality indices (Hu et al., 2015).

### 3.3.3. Integration of multi-pollutants regionalization results

The analyses above indicate that the regionalization results based on the GWREOF approach show high similarity for different air pollutants concentrations and air quality metrics. Therefore, we aim to integrate these regionalization results so that the resulting air pollution regionalization is robust as a climatology and can be applied to different air pollutants.

Since  $\text{PM}_{2.5}$  and ozone are the two predominant pollutants, we use their GWREOF results during the pollution seasons to cluster the cities and identify spatial modes. Based on the spatial modes shown in Fig. 6a and d, each of the 337 Chinese cities is clustered into the nearest spatial mode. As for the cities outside spatial modes (the loading coefficient is smaller than 0.65), we classify them into their most correlated modes, i.e., the spatial mode with the largest loading coefficient for the city. Fig. 7 shows the spatial locations of the 337 cities, color-coded by the clustering numbers for  $\text{PM}_{2.5}$  and ozone. The clustering results of the 337 Chinese cities are generally consistent for  $\text{PM}_{2.5}$  and ozone. However, some cities are clustered into different regions, as would be expected due to the differences in  $\text{PM}_{2.5}$  vs. ozone-derived spatial modes. For these cities, we then use the spatial modes of other air pollutants (e.g.,  $\text{SO}_2$ ,  $\text{NO}_2$ , etc.) and cluster the cities to the mode that most air pollutants locate. One exception is the city of Xilin Gol League in central Inner Mongolia; although the spatial modes of both  $\text{PM}_{2.5}$  and  $\text{O}_3$  cluster the city with Beijing and northern Hebei (Fig. 7), this city is clustered with other areas in Inner Mongolia due to the sparseness of measurements around the area.

Following the procedure described above and further accounting for the city-level administrative divisions, we derive a regionalization result based on the GWREOF method applied to multi-pollutants in China. Fig. 8 shows the 18 clustered regions and Table 1 summarizes their main areas, key features, and annual pollution levels. Table S1 further summarizes their main topography features and the variance contributions. The 18 regions are NCP, YRD, PRD, SCB, GZB, CYR, HLJ, LNP, SDI, NMG, BTN, FJC, GZP, GX, YNP, QGP, TIB, and XJ regions. Comparing the regionalization results in previous studies (Cheng et al., 2018; Wang et al., 2015; Yao et al., 2020), our regionalization results have some similarities (i.e., the major regions including NCP, YRD, PRD, SCB, LNP, and XJ can always be identified), and some differences (i.e., The South China region can be divided into two regions PRD and GX). There are also some differences compared to the current socio-economic zones. For example, the Beijing-Tianjin-

Hebei region is usually classified as one megacity cluster (or urban agglomeration); however, in our result, Beijing, Tianjin, and northern Hebei are clustered into the BTN region, while southern Hebei is clustered into the NCP region. The main reasons are that the  $\text{PM}_{2.5}$  variations in BTN cities could be affected by frequent northwest wind during winter (Chen et al., 2018; Yang et al., 2020), and emissions in southern Hebei are more similar to NCP (Gao et al., 2018).

### 3.4. Driving factors and pollution control strategies for the regionalization

#### 3.4.1. The influences of topography and meteorology

The similarity of regionalization results using different air pollutants and years suggest common factors contributing to air pollution's spatial and temporal variability. The topography and meteorology conditions are candidates as they strongly affect air pollutants' spatial-temporal distributions. Fig. 9 shows the topography and regionalization results based on the nationwide visibility measurements over 2015–2020 in China. The metric of visibility reflects the haze pollution level and is also strongly determined by meteorological conditions (Mao et al., 2019). As shown in Fig. 9b, the GWREOF approach clusters the cities into 20 regions using the nationwide visibility measurements. We can see that 12 visibility regions are spatially consistent with air pollution regions, including LNP, YRD, BTN, GX, SCB, PRD, FJC, HLJ, YNP, GZP, QGP, and XJ (Table 1). Five regions, including Shandong, Zhongyuan Plain, Shanxi and southern Hebei, and Huhhot-Baotou-Erdos (Fig. 9b), overlap the air pollution regions of SDI, NCP, CYR, and NMG.

We find that spatial regionalization in China can be strongly influenced by the topography. Clustered regions tend to have relatively stable topography inside, and changes in the topography often lead to changes in clustered regions, as can be seen by comparing Fig. 9a and Fig. 9b. For example, the NCP region has a relatively flat terrain, surrounded by Yanshan and Taihang Mountains. The YRD, MYRD, HLJ, and LNP regions are also located in plain landforms. The basin and plateau terrains can also influence air pollution regionalization. The SCB and GZB regions are located in the typical basin terrain surrounded by mountains or plateaus. The regions of GZP, GX, YNP, QGP, TIP, and XJ are located in Guizhou Plateau, Guangxi Basin, Yunnan Plateau, Qinghai-Gansu Plateau, Tibetan Plateau, and Xinjiang, respectively.

As shown in Fig. 9, the topography may not only affect the air pollution regionalization, but also largely modulate the regions' economic developments, population densities, and emission intensities. Previous studies have found a significant impact of topography on the spatial distribution of air pollution in China (Wang et al., 2018; Xu et al., 2016). Taking SCB as an example, we analyze the emission intensity, economic development, and pollution levels inside and outside the region (Table S2). Considerable

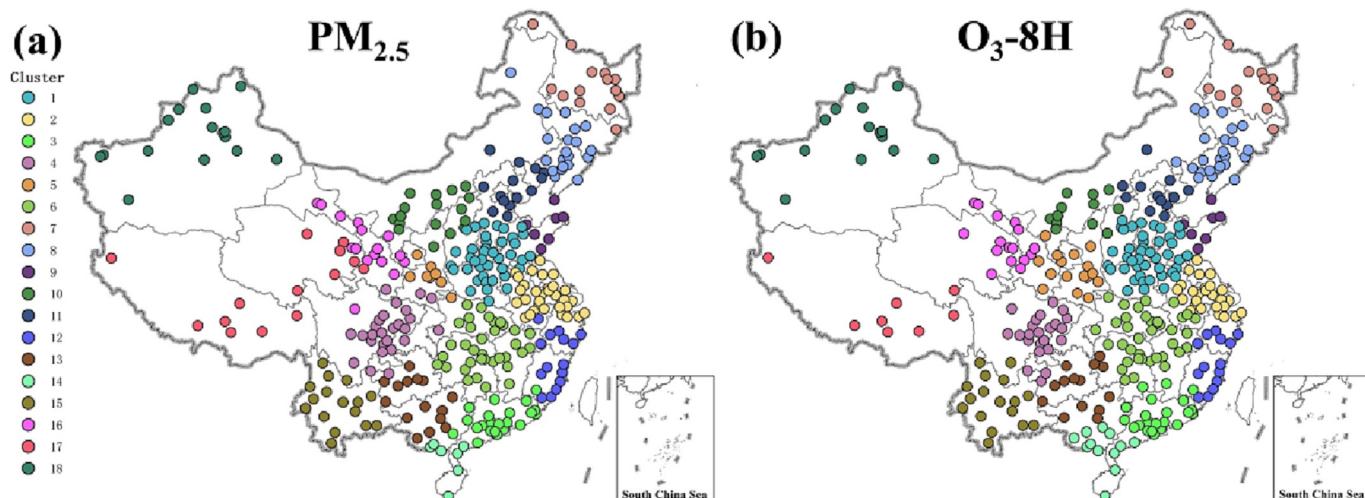


Fig. 7. Cities attributed to different spatial modes based on GWREOF results for (a)  $\text{PM}_{2.5}$  pollution season and (b)  $\text{O}_3$ -8H pollution season of 2015–2020.



**Fig. 8.** Air pollution regionalization based on multi-pollutants for China. The 18 identified regions and their pollution levels are summarized in Table 1.

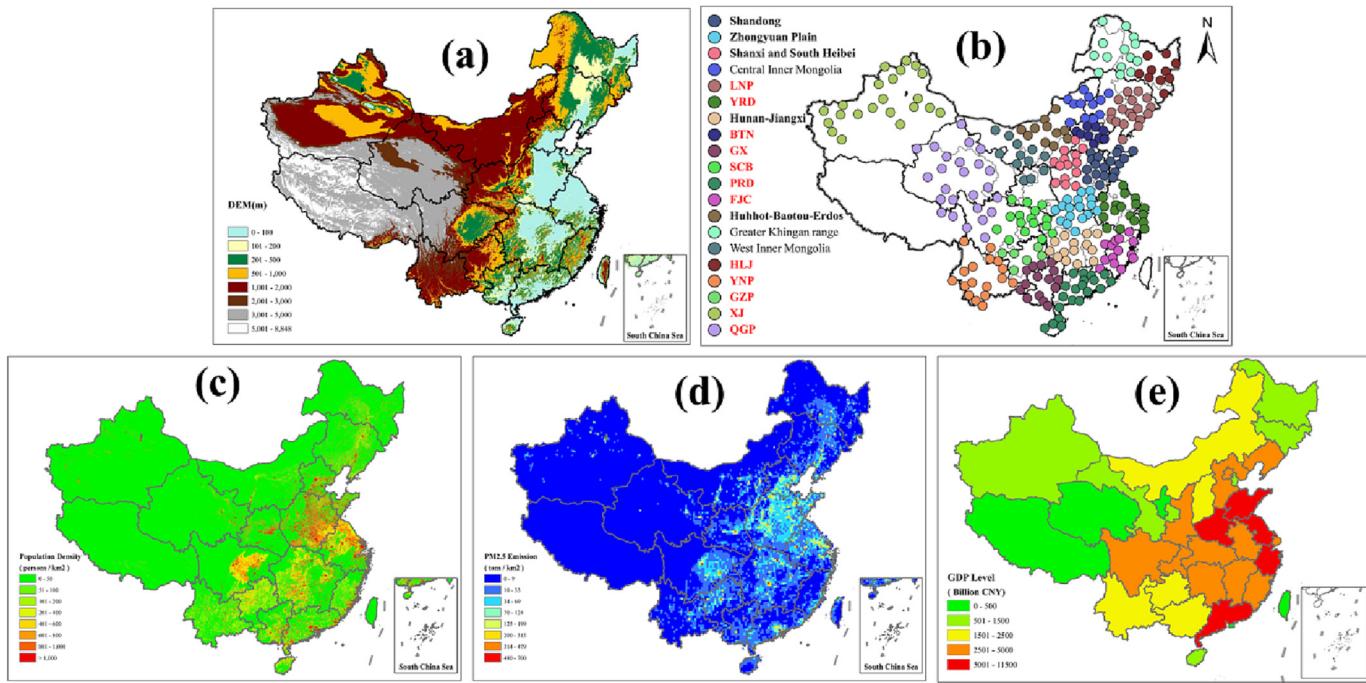
differences can be seen inside and outside SCB: population densities of 408 person/km<sup>2</sup> vs. 21 person/km<sup>2</sup>, GDP per capita of 59,467 CNY vs. 44,953 CNY, PM<sub>2.5</sub> emissions of 182,002 ton/year vs. 6504 ton/year, and regional mean PM<sub>2.5</sub> of 35  $\mu\text{g m}^{-3}$  vs. 15  $\mu\text{g m}^{-3}$ . A similar effect of terrain on air pollution can be found in GZB (Table S2). For the typical plain regions, such as YRD, we also find significant differences in the economic development and emission intensity when compared with its surrounding regions (e.g., FJC).

The population density, GDP per capita, and PM<sub>2.5</sub> emissions in YRD (800 person/km<sup>2</sup>, 123,445 CNY, and 139,676 tons/year) are much higher than in FJC (335 person/km<sup>2</sup>, 105,806 CNY, and 94,017 tons/year). Changes in the topography tend to be associated with population, economic developments, and pollution emissions (Liu et al., 2022; Xu et al., 2019; Yang et al., 2020), thus providing natural boundaries for air pollution regionalization.

**Table 1**

The 18 identified regions based on air pollution measurements over 2015–2020 and their regional mean air pollution levels.

Region number	Region abbreviation	Areas	PM <sub>2.5</sub> ( $\mu\text{g m}^{-3}$ )	90th O <sub>3</sub> -8H ( $\mu\text{g m}^{-3}$ )	SO <sub>2</sub> ( $\mu\text{g m}^{-3}$ )	NO <sub>2</sub> ( $\mu\text{g m}^{-3}$ )	O <sub>x</sub> ( $\mu\text{g m}^{-3}$ )
1	NCP	North China plain, including Hebei, Henan, Western Shandong	62	167	24	36	150
2	YRD	Yangtze River Delta, including Shanghai, Jiangsu, Zhejiang, and Anhui provinces	43	149	13	32	140
3	PRD	Pearl River Delta in Guangdong Province	29	136	11	25	124
4	SCB	Sichuan Basin, including Sichuan and Chongqing	35	126	12	26	117
5	GZB	Guanzhong Basin in southern Shaanxi province	42	135	15	32	131
6	CYR	Central Yangtze River, including Hubei, Hunan and Jiangxi provinces	43	133	14	25	120
7	HLJ	Heilongjiang Province	31	102	13	20	98
8	LNP	Lining and Jilin Province	40	135	21	26	122
9	SDI	Eastern Shandong	43	155	17	30	143
10	NMG	Midwest Inner Mongolia	36	143	30	31	135
11	BTN	Beijing, Tianjin, and northern Hebei	48	164	19	37	143
12	FJC	Fujian Coastal and Southern Zhejiang	26	123	10	24	119
13	GZP	Guizhou Plateau	25	107	15	17	96
14	GX	Guangxi Province	29	118	10	18	105
15	YNP	Yunnan Plateau	21	113	12	17	102
16	QGP	Qinghai-Gansu Plateau	27	128	18	23	128
17	TIB	Tibetan Plateau	13	112	9	13	103
18	XJ	Xinjiang region	42	117	12	27	117



**Fig. 9.** Spatial distributions of (a) the topography, (b) visibility regionalization result analyzed by GWREOF using the 2015–2020 daily visibility measurements, (c) population density, (d) primary PM<sub>2.5</sub> emissions for the year 2017, and (e) provincial GDP in China.

Furthermore, the similarity of air pollution regionalization results among different air pollutants can also be caused by their common correlations with meteorological conditions. Previous studies have demonstrated that meteorology significantly influences the spatial and temporal variations of air pollution levels in China (An et al., 2019; Zhang et al., 2019; Chen et al., 2020; Yang et al., 2020). Here using the measurements of air pollutants concentrations and meteorological variables, we also find strong correlations between them in the heavily polluted regions (Table S3). We further apply the GWREOF method to regionalize the meteorological fields, including planetary boundary layer height (PBLH), relative humidity (RH), temperature (T), and wind speed (WS) (Fig. S9). The regionalization results for these meteorological variables are also similar to the air pollution regionalization results. Most air pollution regions (such as NCP, YRD, PRD, SCB, etc.) can be identified when using meteorological measurements for regionalization.

#### 3.4.2. Atmospheric environment characteristics of different regions

We now examine the air pollution characteristics in the 18 identified regions, which shall have distinct differences among them. Fig. 10 shows the statistics of PM<sub>2.5</sub> and O<sub>3</sub> concentrations and the corresponding probability density distributions for the 18 regions. Both severe PM<sub>2.5</sub> and O<sub>3</sub> pollution areas can be seen in NCP, YRD, GZB, BTN, SDP, and LNP regions, with frequent exceedances of PM<sub>2.5</sub> and O<sub>3</sub> national air quality standards occurring. O<sub>3</sub> pollution is also prominent in southern China, such as the PRD region.

The probability density distributions of PM<sub>2.5</sub> for these regions show unimodal distributions with different means and variances of pollution levels. The probability density distribution of the clean TIB region presents a higher peak and narrower width, reflecting a small variance ( $6 \mu\text{g m}^{-3}$ ) around a low average value ( $13 \mu\text{g m}^{-3}$ ). In contrast, the NCP region presents a lower peak and broader width, which means the pollution concentrations fluctuate in a large variance ( $40 \mu\text{g m}^{-3}$ ) around a high average value ( $62 \mu\text{g m}^{-3}$ ). When comparing the probability density distributions of PM<sub>2.5</sub> between 2015 and 2020 (Fig. S10a and Fig. S10c), all means for these regions declined in 2020 compared to their means in 2015. The variances in heavily polluted regions in 2020 have decreased to  $33 \mu\text{g m}^{-3}$  in NCP,  $21 \mu\text{g m}^{-3}$  in YRD, and  $29 \mu\text{g m}^{-3}$  in GZB, from, respectively,  $42 \mu\text{g m}^{-3}$ ,  $27 \mu\text{g m}^{-3}$ , and  $32 \mu\text{g m}^{-3}$  in those regions in 2015. It reflects

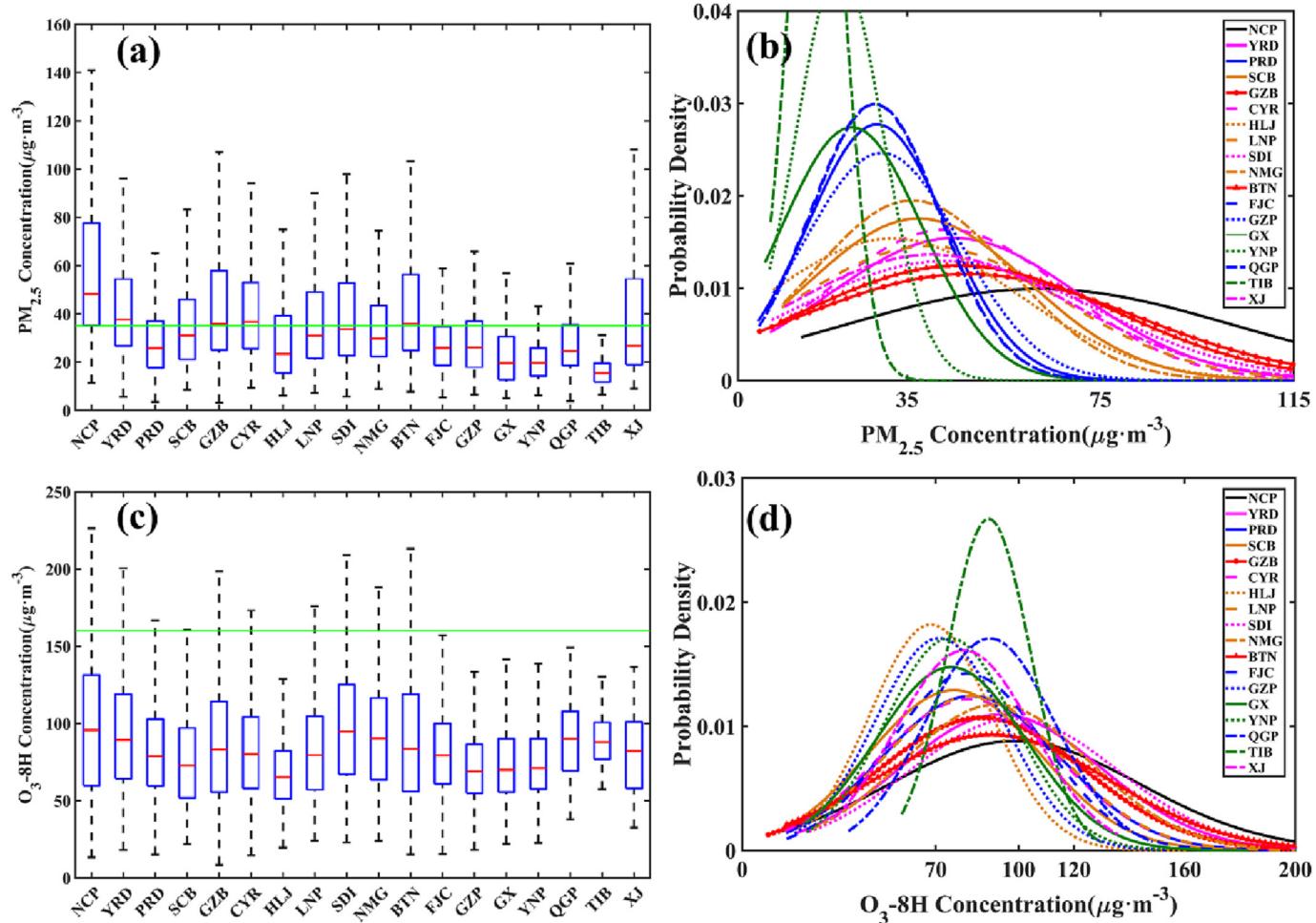
that severe PM<sub>2.5</sub> pollution concentrations in heavily polluted areas have been effectively reduced (Xue et al., 2019; Zhang et al., 2019).

Due to the effect of background ozone concentration, the probability density distributions of O<sub>3</sub> for the different regions show broad overlaps in the range of  $70 \mu\text{g m}^{-3}$  to  $100 \mu\text{g m}^{-3}$ . Their 90th percentile O<sub>3</sub>-8H concentrations, however, show considerable differences. For example, the 90th percentile O<sub>3</sub>-8H in heavily polluted regions can reach  $187 \mu\text{g m}^{-3}$  in NCP,  $176 \mu\text{g m}^{-3}$  in YRD, and  $163 \mu\text{g m}^{-3}$  in PRD. The 90th percentile O<sub>3</sub>-8H levels in clean regions, such as HLJ in northeast China, GZB and TIB in western China are about  $102$ – $112 \mu\text{g m}^{-3}$ . The probability density distributions of O<sub>3</sub> also show a notable upward shift in 2020 relative to that in 2015 (Fig. S10b and Fig. S10d).

Despite the different probability distributions for PM<sub>2.5</sub> and O<sub>3</sub>, we find that the high PM<sub>2.5</sub> and high O<sub>3</sub> pollution events are strongly co-located in these regions. Fig. 11 shows the scatterplots of annual 90th percentile O<sub>3</sub>-8H vs. annual mean PM<sub>2.5</sub> concentrations averaged among regions. The strong spatial correlations indicate that regions with high PM<sub>2.5</sub> concentrations, such as NCP, BTN, and YRD, are also where high O<sub>3</sub> pollution levels tend to happen, although the two pollution events may occur in different seasons. Such a spatial consistency is mainly driven by the spatial distribution of pollutant emissions that share similar spatial patterns as population and GDP (Fig. 9). The temporal correlations between daily O<sub>3</sub>-8H and PM<sub>2.5</sub> concentrations vary among different regions and cities, reflecting their different temporal variations (i.e., seasonal variations). As described in Section 3.3, NCP, YRD, and GZB regions are prone to severe PM<sub>2.5</sub> pollution in winter and O<sub>3</sub> pollution in summer. The 2015–2020 annual mean PM<sub>2.5</sub> concentration reaches  $62 \mu\text{g m}^{-3}$  in NCP,  $43 \mu\text{g m}^{-3}$  in YRD, and  $43 \mu\text{g m}^{-3}$  in GZB. There the 90th percentile O<sub>3</sub> concentrations are  $167 \mu\text{g m}^{-3}$ ,  $149 \mu\text{g m}^{-3}$ , and  $136 \mu\text{g m}^{-3}$ , respectively.

#### 3.4.3. Implication for pollution control

Our analyses indicate that the regionalization results are largely determined by topography and meteorology, while their pollution conditions are strongly affected by emission patterns. Further analyzing the temporal PM<sub>2.5</sub> and O<sub>3</sub> correlations shows a distinct difference between northern and southern China. Fig. 12 shows the correlation coefficients between PM<sub>2.5</sub> and O<sub>3</sub> in 337 cities across China. Strong positive PM<sub>2.5</sub> vs. O<sub>3</sub>

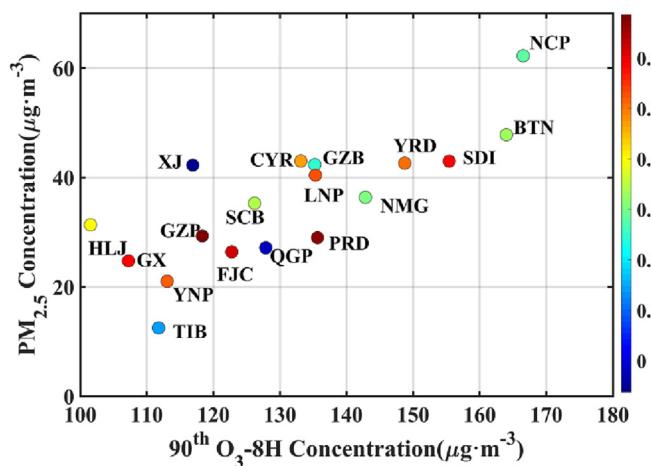


**Fig. 10.** Box-and-whisker plots (minimum, 25th percentile, median, 75th percentile, and maximum) of daily PM<sub>2.5</sub> (a) and O<sub>3</sub>-8H (c) concentrations over 2015–2020, and the corresponding probability density curves of PM<sub>2.5</sub> (b) and O<sub>3</sub>-8H (d) for the 18 regions. The horizontal green lines in the left panels denote national air quality standards for PM<sub>2.5</sub> (35  $\mu\text{g}\cdot\text{m}^{-3}$ ) and O<sub>3</sub>-8H (160  $\mu\text{g}\cdot\text{m}^{-3}$ ).

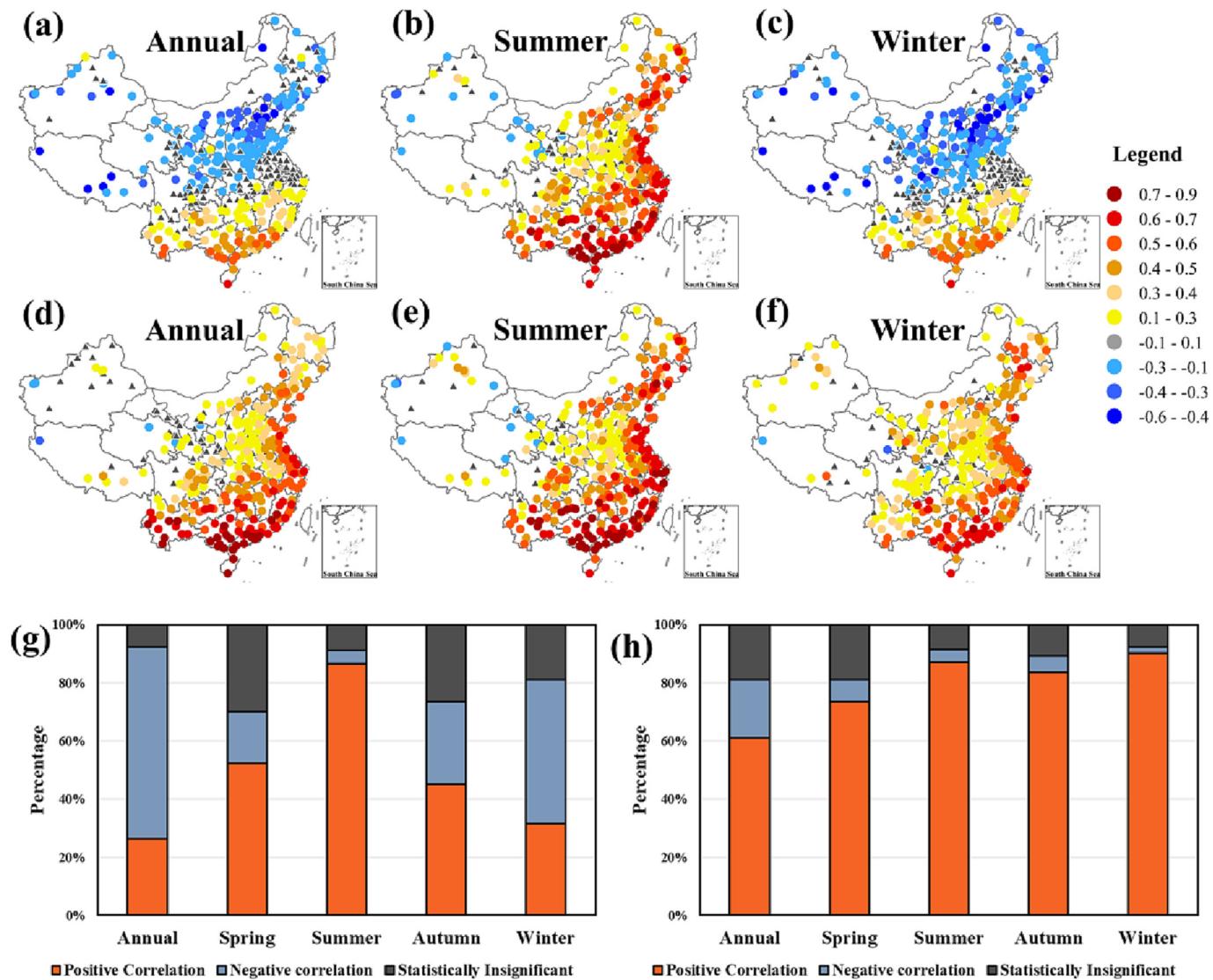
correlations ( $r > 0.70$ ) are found in the south of the Yangtze River, and negative or insignificant correlations occur in the north of the Yangtze River. The positive PM<sub>2.5</sub> vs. O<sub>3</sub> correlations in southern China can be largely associated with more active sunlight-driven photochemistry that promotes

the formation of both secondary aerosols and ozone (Chen et al., 2019; Qiu et al., 2022). Previous studies have reported a dominant secondary component in PM<sub>2.5</sub> for measurements south of the Yangtze River (Tao et al., 2017; Wang et al., 2019; Zhu et al., 2018). For example, secondary components (such as sulfate, nitrate, and secondary organic aerosol) account for about 70 % to 80 % of PM<sub>2.5</sub> chemical components in PRD (Yan et al., 2020). In northern China, primary emissions of PM<sub>2.5</sub> and its components are still the dominant contributors (Gao et al., 2018; Gautam et al., 2018), and O<sub>3</sub> is further titrated by NO (Tan et al., 2019). Therefore, controlling primary PM<sub>2.5</sub> emissions is still the key measure in northern China, such as in NCP, GZB, and SDI, particularly for addressing the heavy wintertime PM<sub>2.5</sub> air pollution.

Considering NO<sub>2</sub> as a potential atmospheric oxidant, we analyze the correlations of PM<sub>2.5</sub> with the total oxidant O<sub>x</sub>. As shown in Fig. 12, the PM<sub>2.5</sub> vs. O<sub>x</sub> correlations become almost positive across China throughout the year. Even in winter, the correlation coefficients between PM<sub>2.5</sub> and O<sub>x</sub> can reach 0.45 in NCP, 0.48 in LNP, and 0.55 in SDI. The proportion of cities with high correlation coefficients ( $>0.5$ ) could reach above 70 % in the whole year and above 85 % in both summer and winter. The strong positive correlations between PM<sub>2.5</sub> and O<sub>x</sub> indicate the important role of atmospheric oxidants and oxidizing capacity, which is the key driver of atmospheric pollution complex with both high PM<sub>2.5</sub> and O<sub>3</sub> air pollution. Therefore, it is necessary to implement synergistic controls for PM<sub>2.5</sub> and O<sub>3</sub> air pollution from the perspective of atmospheric oxidizing capacity, not only in southern China but also extending to northern China.



**Fig. 11.** Relationship of 90th percentile O<sub>3</sub>-8H and annual mean PM<sub>2.5</sub> concentrations for the 18 regions. Colors denote the temporal correlation coefficients for each region's daily O<sub>3</sub>-8H and PM<sub>2.5</sub> concentrations.



**Fig. 12.** Spatial distribution of correlation coefficients between daily PM<sub>2.5</sub> and O<sub>3</sub>-8H (top panels, (a)-(c)), correlation coefficients between daily PM<sub>2.5</sub> and O<sub>x</sub> (middle panels, (d)-(f)) at 337 Chinese cities over 2015–2020, and percentage of PM<sub>2.5</sub> vs. O<sub>3</sub> correlations (bottom left, (g)) and PM<sub>2.5</sub> vs. O<sub>x</sub> correlations (bottom right, (h)) that are positive, negative, or statistically insignificant ( $p$ -value  $<0.05$ ) at the Chinese cities annually and in different seasons.

The air pollution control measures since 2013 have significantly reduced the primary air pollutants emissions, such as SO<sub>2</sub>, NO<sub>2</sub>, and primary PM<sub>2.5</sub>, while their decrease rates are rather different. We find that in all 18 regions, there are rising trends of NO<sub>2</sub>/SO<sub>2</sub> concentration ratios, and the O<sub>3</sub> exceedance over PM<sub>2.5</sub> exceedance ratios are generally increasing during 2015–2020 (Supplementary Fig. S11). This feature reflects a transition of pollution control strategy we are facing, from PM<sub>2.5</sub> focused to PM<sub>2.5</sub> and O<sub>3</sub> coordinated controls. We find high O<sub>3</sub> exceedance over PM<sub>2.5</sub> exceedance ratios in regions of PRD, NCP, CYR, FJC, and GX. On the other hand, in regions of GZB, NMG, HLJ, LNP, QGP, and XJ regions, PM<sub>2.5</sub> exceedance still dominates. Our regionalization results will help design region-specific pollution control strategies that can be further investigated in future work.

#### 4. Conclusions

Air pollution in China has shown distinct regionalized agglomeration characteristics. Both PM<sub>2.5</sub> and O<sub>3</sub> air pollution can occur over regional scales. Identifying air pollution control regions is a key and necessary strategy to facilitate the implementation of air pollution control measures. Previous air pollution control regions are generally based on the administrative division, topography, and climatic conditions. Some recent studies applied

quantitative methods to cluster the air pollution zones based on their variations in concentrations but only applied to a limited period. In this study, we develop a new approach called Geographically Weighted Rotation Empirical Orthogonal Function (GWREOF) and apply it to analyze air pollution regionalization in China based on multiple pollutants. Compared with previous regionalization methods, GWREOF better accounts for the variability of air pollution conditions driven by emission patterns and meteorology with centralized spatial locations.

We apply the GWREOF approach to multiple air pollutants and air quality metrics using nationwide surface measurements in China at 337 cities over 2015–2020. By considering geographical weighting, the GWREOF regionalization results show much more distinct centralized spatial modes than the REOF approach, which only considers pollution variations. We find that the regionalization results show a high consistency for different air pollutants, air quality metrics, and years. Although recent China's air pollution control actions have significantly decreased the concentration levels of pollutants, including PM<sub>2.5</sub> and primary air pollutants (except for O<sub>3</sub> with an increasing trend), their spatial patterns are relatively stable, explaining the consistency in their regionalization results. We propose an integrated regionalization result, which reflects a climatological condition and can be applied to multiple air pollutants. The regionalization identifies

18 air pollution control regions, including North China Plain, Yangtze River Delta, Pearl River Delta, Sichuan Basin, Guanzhong Plain, Central Yangtze River, Heilongjiang Province, Liaoning region, Shandong Peninsula, Yunnan Plateau, Guizhou Plateau, Fujian Coastal region, etc.

We find that the regionalization results are primarily determined by topography and meteorology conditions, and the severity of air pollution levels in these regions strongly reflect the emission patterns. High emissions in eastern China lead to both high PM<sub>2.5</sub> pollution and high annual 90th percentile O<sub>3</sub> levels, although the two air pollutants may occur in different seasons. Analyzing the PM<sub>2.5</sub>-O<sub>3</sub> concentration correlations in these regions generally show positive correlations in southern China while negative correlations in northern China, suggesting primary emission controls are still the critical measure for PM<sub>2.5</sub> pollution in the north. However, the PM<sub>2.5</sub>-O<sub>x</sub> concentrations show positive correlations broadly over China. As the air pollution condition in China transits from PM<sub>2.5</sub>-focused to PM<sub>2.5</sub>-O<sub>3</sub> synergic control, region-specific and collaborative pollution control strategies are in need from the perspective of atmospheric oxidizing capacity.

#### CRediT authorship contribution statement

**Peipei Qiu:** Conceptualization, Methodology, Data Collection and Data Curation, Software Modeling, Formal analysis, Writing—Original and revised Draft. **Lin Zhang:** Conceptualization, Writing—review, Editing & Supervision. **Xuesong Wang:** Funding acquisition, Data collection and Discussion. **Yafei Liu:** Data collection and analysis, Discussion. **Shuai Wang:** Data collection and analysis, Discussion. **Sunling Gong:** Data collection and Discussion. **Yuanyang Zhang:** Funding acquisition, Conceptualization, Writing—review, Editing & Supervision. All the authors commented on the paper.

#### Data availability

Data will be made available on request.

#### Declaration of competing interest

The authors declare they have no conflict of interest.

#### Acknowledgments

This work was supported by the Special Funds of the State Environmental Protection Public Welfare Industry (No. 201509001) and Supported by the National Key Research and Development Program of China (No. 2018YFC0213203).

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.scitotenv.2023.162431>.

#### References

- An, Z., Huang, R.J., Zhang, R., Tie, X., Li, G., Cao, J., Zhou, W., Shi, Z., Han, Y., Gu, Z., Ji, Y., 2019. Severe haze in northern China: a synergy of anthropogenic emissions and atmospheric processes. *Proc. Natl. Acad. Sci.* 116, 8657–8666.
- Carro-Calvo, L., Ordóñez, C., García-Herrera, R., Schnell, J.L., 2017. Spatial clustering and meteorological drivers of summer ozone in Europe. *Atmos. Environ.* 167, 496–510.
- Carvalho, M.J., Melo-Gonçalves, P., Teixeira, J.C., Rocha, A., 2016. Regionalization of Europe based on a K-means cluster analysis of the climate change of temperatures and precipitation. *Phys. Chem. Earth A/B/C* 94, 22–28.
- Chen, Z., Xie, X., Cai, J., Chen, D., Gao, B., He, B., Cheng, N., Xu, B., 2018. Understanding meteorological influences on PM<sub>2.5</sub> concentrations across China: a temporal and spatial perspective. *Atmos. Chem. Phys.* 18, 5343–5358.
- Chen, J., Shen, H., Li, T., Peng, X., Cheng, H., Ma, A.C., 2019. Temporal and spatial features of the correlation between PM<sub>2.5</sub> and O<sub>3</sub> concentrations in China. *Int. J. Environ. Res. Public Health* 16.
- Chen, Z., Chen, D., Zhao, C., Kwan, M.P., Cai, J., Zhuang, Y., Zhao, B., Wang, X., Chen, B., Yang, J., Li, R., He, B., Gao, B., Wang, K., Xu, B., 2020. Influence of meteorological conditions on PM<sub>2.5</sub> concentrations across China: a review of methodology and mechanism. *Environ. Int.* 139, 105558.
- Cheng, L., Wang, S., Gong, Z., Li, H., Yang, Q., Wang, Y., 2018. Regionalization based on spatial and seasonal variation in ground-level ozone concentrations across China. *J. Environ. Sci.* 67, 179–190.
- Chu, B., Ma, Q., Liu, J., Ma, J., Zhang, P., Chen, T., Feng, Q., Wang, C., Yang, N., Ma, H., Ma, J., Russell, A.G., He, H., 2020. Air pollutant correlations in China: secondary air pollutant responses to NO<sub>x</sub> and SO<sub>2</sub> control. *Environ. Sci. Technol. Lett.* 7, 695–700.
- Dedoussi, I.C., Eastham, S.D., Monier, E., Barrett, S.R.H., 2020. Premature mortality related to United States cross-state air pollution. *Nature* 578, 261–265.
- Fan, H., Zhao, C., Yang, Y., 2020. A comprehensive analysis of the spatio-temporal variation of urban air pollution in China during 2014–2018. *Atmos. Environ.* 220, 117066.
- Feng, Y., Ning, M., Lei, Y., Sun, Y., Liu, W., Wang, J., 2019. Defending blue sky in China: effectiveness of the "Air pollution prevention and control action Plan" on air quality improvements from 2013 to 2017. *J. Environ. Manag.* 252, 109603.
- Gao, H., Chen, J., Wang, B., Tan, S.-C., Lee, C.M., Yao, X., Yan, H., Shi, J., 2011. A study of air pollution of city clusters. *Atmos. Environ.* 45, 3069–3077.
- Gao, J., Wang, K., Wang, Y., Liu, S., Zhu, C., Hao, J., Liu, H., Hua, S., Tian, H., 2018. Temporal-spatial characteristics and source apportionment of PM<sub>2.5</sub> as well as its associated chemical species in the Beijing-Tianjin-Hebei region of China. *Environ. Pollut.* 233, 714–724.
- Gautam, S., Patra, A.K., Kumar, P., 2018. Status and chemical characteristics of ambient PM<sub>2.5</sub> pollutions in China: a review. *Environ. Dev. Sustain.* 21, 1649–1674.
- Govender, P., Sivakumar, V., 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). *Atmos. Pollut. Res.* 11, 40–56.
- Guo, H., Cheng, T., Gu, X., Wang, Y., Chen, H., Bao, F., Shi, S., Xu, B., Wang, W., Zuo, X., Zhang, X., Meng, C., 2017. Assessment of PM<sub>2.5</sub> concentrations and exposure throughout China using ground observations. *Sci. Total Environ.* 601–602, 1024–1030.
- Hannachi, A., Jolliffe, I.T., Stephenson, D.B., 2007. Empirical orthogonal functions and related techniques in atmospheric science: a review. *Int. J. Climatol.* 27, 1119–1152.
- Harris, P., Brunsdon, C., Charlton, M., 2011. Geographically weighted principal components analysis. *Int. J. Geogr. Inf. Sci.* 25, 1717–1736.
- Holman, C., Harrison, R., Querol, X., 2015. Review of the efficacy of low emission zones to improve urban air quality in European cities. *Atmos. Environ.* 111, 161–169.
- Hu, J., Ying, Q., Wang, Y., Zhang, H., 2015. Characterizing multi-pollutant air pollution in China: comparison of three air quality indices. *Environ. Int.* 84, 17–25.
- Jolliffe, I.T., Cadima, J., 2016. Principal component analysis: a review and recent developments. *Philos. Transact. A Math. Phys. Eng. Sci.* 374, 20150202.
- Kuerban, M., Waili, Y., Fan, F., Liu, Y., Qin, W., Dore, A.J., Peng, J., Xu, W., Zhang, F., 2020. Spatio-temporal patterns of air pollution in China from 2015 to 2018 and implications for health risks. *Environ. Pollut.* 258, 113659.
- Li, M., Liu, H., Geng, G., Hong, C., Liu, F., Song, Y., Tong, D., Zheng, B., Cui, H., Man, H., Zhang, Q., He, K., 2017. Anthropogenic emission inventories in China: a review. *Natl. Sci. Rev.* 4, 834–866.
- Li, M., Zhang, Q., Zheng, B., Tong, D., Lei, Y., Liu, F., Hong, C., Kang, S., Yan, L., Zhang, Y., Bo, Y., Su, H., Cheng, Y., He, K., 2019. Persistent growth of anthropogenic non-methane volatile organic compound (NMVOC) emissions in China during 1990–2017: drivers, speciation and ozone formation potential. *Atmos. Chem. Phys.* 19, 8897–8913.
- Li, K., Jacob, D.J., Shen, L., Lu, X., De Smedt, I., Liao, H., 2020. Increases in surface ozone pollution in China from 2013 to 2019: anthropogenic and meteorological influences. *Atmos. Chem. Phys.* 20, 11423–11433.
- Li, Y., Shi, G., Chen, Z., 2021. Spatial and temporal distribution characteristics of ground-level nitrogen dioxide and ozone across China during 2015–2020. *Environ. Res. Lett.* 16, 124031.
- Liu, J., Xue, L., Huang, X., Wang, Z., Lou, S., Ding, A., 2022. Intensified haze formation and meteorological feedback by complex terrain in the North China Plain region. *Atmos. Ocean. Sci. Lett.*, 100273.
- Lu, X., Hong, J., Zhang, L., Cooper, O.R., Schultz, M.G., Xu, X., Wang, T., Gao, M., Zhao, Y., Zhang, Y., 2018. Severe surface ozone pollution in China: a global perspective. *Environ. Sci. Technol. Lett.* 5, 487–494.
- Lu, X., Zhang, L., Wang, X., Gao, M., Li, K., Zhang, Y., Yue, X., Zhang, Y., 2020. Rapid increases in warm-season surface ozone and resulting health impact in China since 2013. *Environ. Sci. Technol. Lett.* 7, 240–247.
- Lyapina, O., Schultz, M.G., Hense, A., 2016. Cluster analysis of European surface ozone observations for evaluation of MACC reanalysis data. *Atmos. Chem. Phys.* 16, 6863–6881.
- Ma, Z., Hu, X., Sayer, A.M., Levy, R., Zhang, Q., Xue, Y., Tong, S., Bi, J., Huang, L., Liu, Y., 2016. Satellite-based spatiotemporal trends in PM<sub>2.5</sub> concentrations: China, 2004–2013. *Environ. Health Perspect.* 124, 184–192.
- Mao, L., Liu, R., Liao, W., Wang, X., Shao, M., Liu, S.C., Zhang, Y., 2019. An observation-based perspective of winter haze days in four major polluted regions of China. *Natl. Sci. Rev.* 6, 515–523.
- Marco, G., Bo, X., 2013. Air quality legislation and standards in the European Union: background, status and public participation. *Adv. Clim. Chang. Res.* 4, 50–59.
- Pražníkář, J., 2017. Particulate matter time-series and Köppen-Geiger climate classes in North America and Europe. *Atmos. Environ.* 150, 136–145.
- Qiu, Y., Wu, Z., Shang, D., Zhang, Z., Xu, N., Zong, T., Zhao, G., Tang, L., Guo, S., Wang, S., Dao, X., Wang, X., Tang, G., Hu, M., 2022. The temporal and spatial distribution of the correlation between PM<sub>2.5</sub> and O<sub>3</sub> contractions in the urban atmosphere of China. *Chin. Sci. Bull.* 67, 2008.
- Schmalensee, R., Stavins, R.N., 2019. Policy evolution under the clean air act. *J. Econ. Perspect.* 33, 27–50.
- Shi, G., Peng, X., Huangfu, Y., Wang, W., Xu, J., Tian, Y., Feng, Y., Ivey, C.E., Russell, A.G., 2017. Quantification of source impact to PM using three-dimensional weighted factor model analysis on multi-site data. *Atmos. Environ.* 160, 89–96.
- Silver, B., Reddington, C.L., Arnold, S.R., Spracklen, D.V., 2018. Substantial changes in air pollution across China during 2015–2017. *Environ. Res. Lett.* 13, 114012.
- Tan, Z., Lu, K., Jiang, M., Su, R., Wang, H., Lou, S., Fu, Q., Zhai, C., Tan, Q., Yue, D., Chen, D., Wang, Z., Xie, S., Zeng, L., Zhang, Y., 2019. Daytime atmospheric oxidation capacity in

- four chinese megacities during the photochemically polluted season: a case study based on box model simulation. *Atmos. Chem. Phys.* 19, 3493–3513.
- Tao, J., Zhang, L., Cao, J., Zhang, R., 2017. A review of current knowledge concerning PM2.5 chemical composition, aerosol optical properties and their relationships across China. *Atmos. Chem. Phys.* 17, 9485–9518.
- Tian, Y.-Z., Huang-Fu, Y.-Q., Shi, G.-L., Shi, X.-R., Han, B., Feng, Y.-C., 2016. Accuracy of advanced and traditional three-way factor analysis models for determining source contributions to particulate matter. *Aerosol Air Qual. Res.* 16, 2512–2522.
- Vienneau, D., Briggs, D.J., 2013. Delimiting affinity zones as a basis for air pollution mapping in Europe. *Environ. Int.* 51, 106–115.
- Wang, S., Li, G., Gong, Z., Du, L., Zhou, Q., Meng, X., Xie, S., Zhou, L., 2015. Spatial distribution, seasonal variation and regionalization of PM2.5 concentrations in China. *Sci. China Chem.* 58, 1435–1443.
- Wang, X., Dickinson, R.E., Su, L., Zhou, C., Wang, K., 2018. PM2.5 pollution in China and how it has been exacerbated by terrain and meteorological conditions. *Bull. Am. Meteorol. Soc.* 99, 105–119.
- Wang, Y., Li, W., Gao, W., Liu, Z., Tian, S., Shen, R., Ji, D., Wang, S., Wang, L., Tang, G., Song, T., Cheng, M., Wang, G., Gong, Z., Hao, J., Zhang, Y., 2019. Trends in particulate matter and its chemical compositions in China from 2013–2017. *Sci. China Earth Sci.* 62, 1857–1871.
- Wang, Y., Gao, W., Wang, S., Song, T., Gong, Z., Ji, D., Wang, L., Liu, Z., Tang, G., Huo, Y., Tian, S., Li, J., Li, M., Yang, Y., Chu, B., Petaja, T., Kermanen, V.M., He, H., Hao, J., Kulmala, M., Wang, Y., Zhang, Y., 2020. Contrasting trends of PM2.5 and surface-ozone concentrations in China from 2013 to 2017. *Natl. Sci. Rev.* 7, 1331–1339.
- Wang, Y., Jin, X., Liu, Z., Wang, G., Tang, G., Lu, K., Hu, B., Wang, S., Li, G., An, X., Wang, C., Hu, Q., He, L., Zhang, F., Zhang, Y., 2023. Progress in quantitative research on the relationship between atmospheric oxidation and air quality. *J. Environ. Sci.* 123, 350–366.
- Xu, X., Zhao, T., Liu, F., Gong, S.L., Kristovich, D., Lu, C., Guo, Y., Cheng, X., Wang, Y., Ding, G., 2016. Climate modulation of the tibetan plateau on haze in China. *Atmos. Chem. Phys.* 16, 1365–1375.
- Xu, W., Sun, J., Liu, Y., Xiao, Y., Tian, Y., Zhao, B., Zhang, X., 2019. Spatiotemporal variation and socioeconomic drivers of air pollution in China during 2005–2016. *J. Environ. Manag.* 245, 66–75.
- Xue, T., Liu, J., Zhang, Q., Geng, G., Zheng, Y., Tong, D., Liu, Z., Guan, D., Bo, Y., Zhu, T., He, K., Hao, J., 2019. Rapid improvement of PM2.5 pollution and associated health benefits in China during 2013–2017. *Sci. China Earth Sci.* 62, 1847–1856.
- Yan, F., Chen, W., Jia, S., Zhong, B., Yang, L., Mao, J., Chang, M., Shao, M., Yuan, B., Situ, S., Wang, X., Chen, D., Wang, X., 2020. Stabilization for the secondary species contribution to PM2.5 in the Pearl River Delta (PRD) over the past decade, China: a meta-analysis. *Atmos. Environ.* 242, 117817.
- Yang, Q., Yuan, Q., Yue, L., Li, T., 2020. Investigation of the spatially varying relationships of PM2.5 with meteorology, topography, and emissions over China in 2015 by using modified geographically weighted regression. *Environ. Pollut.* 262, 114257.
- Yao, X., Ge, B., Yang, W., Li, J., Xu, D., Wang, W., Zheng, H., Wang, Z., 2020. Affinity zone identification approach for joint control of PM2.5 pollution over China. *Environ. Pollut.* 265, 115086.
- Ye, W.F., Ma, Z.Y., Ha, X.Z., 2018. Spatial-temporal patterns of PM2.5 concentrations for 338 Chinese cities. *Sci. Total Environ.* 631–632, 524–533.
- Yin, C.Q., Solmon, F., Deng, X.J., Zou, Y., Deng, T., Wang, N., Li, F., Mai, B.R., Liu, L., 2019. Geographical distribution of ozone seasonality over China. *Sci. Total Environ.* 689, 625–633.
- Zhang, X.Y., Wang, Y.Q., Niu, T., Zhang, X.C., Gong, S.L., Zhang, Y.M., Sun, J.Y., 2012. Atmospheric aerosol compositions in China: spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* 12, 779–799.
- Zhang, Q., Zheng, Y., Tong, D., Shao, M., Wang, S., Zhang, Y., Xu, X., Wang, J., He, H., Liu, W., Ding, Y., Lei, Y., Li, J., Wang, Z., Zhang, X., Wang, Y., Cheng, J., Liu, Y., Shi, Q., Yan, L., Geng, G., Hong, C., Li, M., Liu, F., Zheng, B., Cao, J., Ding, A., Gao, J., Fu, Q., Huo, J., Liu, B., Liu, Z., Yang, F., He, K., Hao, J., 2019. Drivers of improved PM2.5 air quality in China from 2013 to 2017. *Proc. Natl. Acad. Sci.* 116, 24463–24469.
- Zhou, H., Liu, Y., 2018. Spatio-temporal pattern of meteorological droughts and its possible linkage with climate variability. *Int. J. Climatol.* 38, 2082–2096.
- Zhu, Y., Huang, L., Li, J., Ying, Q., Zhang, H., Liu, X., Liao, H., Li, N., Liu, Z., Mao, Y., Fang, H., Hu, J., 2018. Sources of particulate matter in China: insights from source apportionment studies published in 1987–2017. *Environ. Int.* 115, 343–357.