Understanding the regional background ozone using multiple methods: A case study in northern China

Fangting Wang[1,2], Kun Zhang[1,2], Jin Xue[1,2], Ling Huang[1,2], Yangjun Wang[1,2], Hui Chen[1,2], Joshua S Fu[3], Li Li [1,2*]

[1]School of Environmental and Chemical Engineering, Shanghai University, Shanghai 200444, China

[2]Key Laboratory of Organic Compound Pollution Control Engineering, Shanghai University, Shanghai 200444, China

[3]Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN 37996, USA

*Correspondence: lily@shu.edu.cn (Li Li).

Abstract

Uprising ground-level ozone (O_3) and its regional pollution in eastern China are attracting more attention. On top of local precursor emissions and photochemistry, background ozone and long-range transport also contribute significantly to O_3 concentrations. To quantify the regional O_3 background concentrations and its yearly and seasonal variations, multiple methods, including the principal component analysis (PCA) and the Texas Commission on Environmental Quality (TCEQ) method, were applied for a case study in Shandong (SD) province in North China, where serious O_3 pollution occurred frequently yet the background contributions have not been well quantified. Results derived from multiple methods show an overall consistent trend with 2018-2020 averaged regional background O_3 (MDA8) of 88.9 μg/m^3, accounting for 79.4% of total O_3 in the region. From 2018-2020, the changes of regional MDA8 O_3 estimated by Methods 1, 2, 3, and 4 are -3.8 μg/m^3,1.6 μg/m^3, -5.2 μg/m^3 and 0.9 μg/m^3, respectively. Clear seasonal variations in the regional background O_3 are observed, showing a pattern of summer > spring > autumn > winter. In addition, the regional ozone contribution at coastal cities was larger than that for inland cities whereas local O_3 formation gradually increased from coastal areas to inland areas. The sea-land wind contribution to O_3 in the eastern coastal cities in summer was around 2.1% at the three-year average level, while the local photochemistry to O_3 in the inland cities was about 29.7% during ozone season. Local photochemical

contribution to O_3 in inland cities during ozone pollution episodes can reach up to 55.8%.

Plain Language Summary:
Multiple methods including PCA with different sets of data inputs, PCA-MLR, TCEQ are applied to quantify the regional background ozone in a typical region (SD) in north China for year 2018-2020. Annual and seasonal changes of the regional background ozone are estimated. Contributions from sea-land wind circulation and local photochemical formation to ozone are also analyzed.

1. Introduction

Tropospheric ozone is a typical secondary pollutant, which adversely affects the public health, crop yields, and air quality [*Chen et al.*, 2007; *Schauberger et al.*, 2019; *Suciu et al.*, 2017; *Tai and Martin*, 2017]. Additionally, O_3 is the third most important greenhouse gas and thus has a significant impact on global climate change [*Morgenstern et al.*, 2014]. A small amount of tropospheric O_3 is transported from the stratosphere; it is mainly produced via photochemical reactions by precursors (NO_x, VOCs, and CO) in the atmosphere. In general, at any location, the measured surface O_3 is the sum of the regional background O_3 and locally produced O_3 [*Berlin et al.*, 2013; *Nielsen-Gammon et al.*, 2005]. Regional background O_3 refers to the amount of O_3 transported into the area by large-scale winds [*Langford et al.*, 2009; *Nielsengammon et al.*, 2005], which mainly includes the photochemical effects of natural emissions of VOCs, NO_x, and CO; long-range transport of O_3 from distant pollutant sources; and O_3 from stratosphere–troposphere gas exchange [*Langford et al.*, 2009; *Vingarzan*, 2004].

The regional background O_3 concentrations considered in this work is distinct from the definitions used elsewhere: “local background O_3”; “baseline O_3”; “policy-relevant background (PRB) O_3” and “Local background O_3”, which include O_3 redistributed by local circulation, originate without local anthropogenic O_3 precursors [*Langford et al.*, 2009]. “Baseline O_3” is defined as O_3 measured at a given site in the absence of strong local emissions of anthropogenic precursors [*Chan and Vet*, 2010]. “PRB O_3” is the O_3 level in the area in absence of local anthropogenic

O_3 precursors [*L Zhang et al.*, 2011]. PRB O_3 is a concept based on a model; therefore, it is distinct from other terms [*McDonald-Buller et al.*, 2011].

O_3 pollution has become increasingly prominent and has shown obvious regional pollution characteristics [*Dai et al.*, 2021; *Dang and Liao*, 2019]. To prevent and control O_3 pollution, it is of scientific significance is to quantify the background and local O_3 contributions to clarify the limit of O_3 reduction by controlling anthropogenic precursors [*Vingarzan*, 2004]. Regarding the concentration of regional background O_3, existing research has not fully addressed this problem. The most commonly used methods for calculating regional background O_3 concentrations are the background in-situ measurement, the PCA method, and the TCEQ regional background O_3 estimation method [*WU et al.*, 2017]. Langford et al. were the first to use PCA to analyze the regional background O_3 concentration for Texas in 2006 and identified the first principal component (explained variance of 84%) as the regional background O_3 concentration, which was demonstrated by the spatial distribution of load and meteorological conditions [*Langford et al.*, 2009]. Using the same method, Liang et al. analyzed the regional background O_3 in the Yangtze River Delta region in May 2016; they posited that local production had a significant contribution to the high concentrations of O_3 [*Liang et al.*, 2018]. Based on the aforementioned method, Suciu et al. innovatively inserted meteorological parameters into the PCA, restricted the regional background O_3, and obtained a more reasonable result [*Suciu et al.*, 2017]. The TCEQ regional background O_3 estimation method is simpler than PCA but has higher requirements for the number of monitoring stations, representativeness of the regional distribution, and integrity of the monitoring data [*WU et al.*, 2017]. The TCEQ method defines the minimum MDA8 O_3 for all monitoring sites in the study area as the regional background O_3, and the difference between the maximum and minimum values as locally generated O_3 [*Nielsengammon et al.*, 2005]. Xue et al. used the TCEQ method to study the relative contribution of the regional background O_3 and local O_3 generation to O_3 level in Hong Kong, and further investigated the long-term trend in regional background O_3 from 2002 to 2013. They found that the regional background contribution accounted for approximately 70% of the total O_3, and the increase in regional background O_3 concentration was the major factor for the increase in urban O_3 concentration [*Xue et al.*, 2014]. However, the estimation of regional background O_3 derived by different methods has obvious differences, causing

high uncertainties. It is of scientific significance to get an overview of the regional background O_3 and understand their variations based on multiple methods.

China has experienced significant O_3 pollution in recent years, particularly in the North China Plain (NCP), one of the most economically developed and polluted regions [*Ma et al.*, 2016; *Sun et al.*, 2021]. Shandong is one of the provinces with serious O_3 pollution in NCP, where the 90^{th} percentile of the annual average MDA8 O_3 climbed from 154 to 186 μg/m^3 in 2015-2019 [*M Zhang et al.*, 2021]. However, the regional background O_3 concentration in the SD region has rarely been studied. In this study, taking SD province as a case study, we used multiple methods, including the PCA method, PCL-MLR method, TCEQ method and background in-situ measurement method to quantify the regional background O_3 concentrations in the SD region. PCA is run twice with single variable (only MDA8 O_3) and multivariable (MDA8 O_3, NO_2, wind direction (WD), wind speed (WS), and temperature (T)) as inputs respectively. The results of PCA analyses, TCEQ method, and background in-situ measurement were compared. On the basis of the aforementioned analysis, we estimated annual changes in the regional background O_3 concentrations, their seasonal variations, and their spatial distributions in the SD region to evaluate the contributions of O_3 by region and provide effective scientific and technological support for the prevention and control of O_3 pollution in the SD region and can apply it to other regions.

2. Methodology

2.1 Data collection and preprocessing

Hourly data on O_3 and NO_2 were collected from 96 Air Quality Monitoring Stations (AQMSs) in Shandong Province from 2018 to 2020. These data were measured by the China National Environmental Monitoring Centre (http://datacenter.mee.gov.cn). The data processing method used in this study is similar to the process reported by Chu et al. [*Chu et al.*, 2020]. First, we deleted the missing values and zero values of site data. Second, we calculated the efficiency of the data [*Shamsipour et al.*, 2014], sites with an efficiency lower than 90% were excluded. Thus, 66 AQMSs were selected after

data screen, which covered all cities. Third, the missing values and zero values were filled with linear interpolation to calculate O_3 MDA8 [*Ottosen and Kumar*, 2019]. For data missing for more than 3 consecutive days, linear interpolation was not used, instead the average of the continues observational data at the remaining sites were used as replacement for such data. The spatial distribution of these sites is shown in Fig. 1.

Meteorological data were extracted from the National Centers for Environmental Prediction (NCEP) final operational global analysis data files with temporal and spatial resolutions of six hours and 2.5°×2.5°, respectively (https://www.psl.noaa.gov/data/gridded/data.ncep.reanalysis.html). A large subset of these data is available from the Physical Sciences Laboratory in its original four-times-daily format and as daily averages. Seven grids covered the SD region, and the corresponding grid meteorological data were matched to the AQMS. The meteorological data contained three elements, temperature, u-wind, and v-wind; temperature is the data for ground 2m height and both winds are at sigma level 995.

2.2 PCA Method

The variation in O_3 concentration varies significantly over time and is influenced by the emission of local O_3 precursors and by meteorological conditions [*Wang et al.*, 2019]. When the meteorological conditions are relatively stable, O_3 concentrations are more likely to be affected by the local photochemistry [*Shan et al.*, 2009] whereas the influence of regional transmission increases as the atmospheric diffusion conditions improve. Therefore, we used the PCA method to analyze the regional background O_3 concentrations by analyzing the multi-site MDA8 O_3 and the single-site MDA8 O_3 with NO_2, WD, WS, and T data at various sites in the SD region after stripping out the impact of anthropogenic emissions on the changes in O_3 concentration. Then, we interpreted the results of PCA in combination with meteorological data.

PCA is effective for dimensionality reduction and simplifying the system structure by converting multiple indicators into several uncorrelated comprehensive indicators (principal components) under the premise of less information loss through the correlation coefficient (or variance covariance) matrix [*MURTAGH et al.*, 1987]. In

general, the first few principal components can explain most of the variance in the original variables, and the results of these principal components are used to explain the original observations [*Abdul-Wahab et al.*, 2005]. PCA can be combined with multiple linear regression (MLR) methods, where the resolved principal components are considered as ozone sources, factor scores are considered as independent variables, and pollutant concentrations after normalization are considered as dependent variables, to predict and further determine the contribution rates of regional background O_3 [*Jolliffe*, 2005; *Statheropoulos et al.*, 1998].

PCA was used to calculate the regional background O_3 concentration. First, we assumed that all stations in the study area were affected by regional transport air masses; therefore, the principal components representing the regional background could be extracted [*WU et al.*, 2017]. Next, using the prior methods as a reference [*Langford et al.*, 2009; *Suciu et al.*, 2017], we used the results of loadings and factor scores to explain which principal components represent the regional background and then inversely calculated the regional background O_3 according to equation (1). This method has been widely applied in O_3 regional background research [*Huang et al.*, 2021; *Liang et al.*, 2018; *Yao et al.*, 2021].

$$O_3 = \bar{O_3} + \sigma\ (O_3)\ \sum_I^{N=66} f_i \alpha_i\ (t) \quad (1)$$

$$O_3^{PC1} = \bar{O_3} + \sigma\ (O_3)\ f_1 \alpha_1\ (t) \quad (2)$$

where $\bar{O}_3$ is the mean of the MDA8 O_3 at 66 sites, $\sigma(O_3)$ is the standard deviation of the data set, f_i is the PC_i variance contribution of the results of the PCA, and α_i is the daily PC_i amplitudes. When PC1 represents the regional background, use equation (2) to calculate 8-hour regional background O_3.

2.3 TCEQ Method

The TCEQ method was proposed by the Texas Commission on Environmental Quality. A rural site in the upwind direction was chosen among all the monitoring sites in the study area and the O_3 concentration at the site was utilized as the regional background [*Langford et al.*, 2009; *WU et al.*, 2017]. Nielsen-Gammon et al [*Nielsengammon et al.*, 2005] presented a TCEQ method based on a larger air quality monitoring network , which is simpler, more reliable, and more widely adopted . This approach calculates

the highest 8-hour O_3 concentration at each site and uses the lowest 8-hour O_3 concentration measured across all site as the regional background value. The improved TCEQ method considers data from a well-established monitoring network with good coverage in all directions in the study region, ensuring that at least one site is not affected by local emissions regardless of wind direction changes. Additionally, the difference between the highest and lowest 8-hour O_3 concentrations at all the sites is defined as the O_3 generated by local photochemical reactions. The daily 8-hour regional background O_3 and locally generated O_3 can be calculated by equation (3).

$$O_{3(R)} = O_{3\_MIN}$$

$$O_{3(L)} = \left(O_{3\_MAX}\right) - \left(O_{3\_MIN}\right) \quad (3)$$

where $O_{3(R)}$ represents the regional background O_3, and $O_{3(L)}$ represents the locally generated O_3.

2.4 Experimental design

We conducted three distinct PCA experiments to analyze single and multiple variables at various stations in the SD region to calculate the regional background O_3 concentrations. Method 1 was the most conventional approach; we used only MDA8 O_3 to run PCAs for the selected 66 AQMSs in the SD region in the 3 years from 2018 to 2020. Method 2 considered information such as meteorological parameters and precursors (NO_2, WD, WS, T) to constrain it, but fewer sites (only five) than in Method 1 were selected, and these sites were distributed in different regions of the SD region to better represent the regional characteristics. We ran five independent PCAs on five sites to extract the regional background O_3 concentrations [*Suciu et al.*, 2017]. Unlike Methods 1 (single variable, multiple sites) and 2 (multiple variables, multiple sites), Method 3 is a relatively innovative method that combines PCA with MLR, usually used for pollutant source analysis [*Bian et al.*, 2013; *Feng et al.*, 2020]. Method 3 includes three steps: First, to assume that regional contribution, local contribution, and other contributions such as ocean, local small air mass contribution, etc. are several sources of ozone. Second, using PCA/MLR to analyze MDA8 O_3 from 66 AQMSs, the PC factor score derived from PCA as independent variable and the standardization result of the mean value of all sites is used as a dependent variable to predict the contribution rate of different sources. Finally, regional background O_3 was

estimated based on the regional contribution rates and factor scores.

Simultaneously, regional background O_3 was calculated using the TCEQ method, which is named as Method 4 in this study. However, considering the influence of the MDA8 O_3 minimum data and location of specific sites on the results, we first found the frequency distribution of the sites with the smallest MDA8 O_3 values and found that one site had the smallest MDA8 value among all the sites on 377 days over 3 years, and it could not adequately capture regional transport air masses due to its location in the urban; thus, this site was removed. Second, the remaining data were cleaned using a phase-line approach, deleting outliers higher than Q3 + 1.5(Q3-Q1) or less than Q1-1.5(Q3-Q1) from the MDA8 O_3 sub-dataset on every day of every month (Q1 and Q3 represent the first and third quartiles, respectively) [*Mousavinezhad et al.*, 2021; *Yin et al.*, 2019]. Moreover, to evaluate the reliability of the results of the four distinct methods, the regional background O_3 was calculated through the different methods and compared with the observations at Yangkou station, Qingdao （the location is shown by the purple star in Fig 1）, which was defined as a background site (https://www.mee.gov.cn/gkml/hbb/bwj/201204/t20120401_250935.htm). Specific information for each method is presented in Table 1.

Table1. Summary of parameters for methods of calculating regional background O_3.

Approach	Areas	Input Parameters
Method 1 (PCA)	66 AQMS in the SD region	MDA8 O_3
Method 2 (PCA)	5 AQMS sites in the SD region	MDA8 O_3, NO_2, WD, WS, T
Method 3 (PCL/MLR)	66 AQMS in the SD region	MDA8 O_3
Method 4 (TCEQ)	65AQMS in the SD region	MDA8 O_3

3. Results and discussions

3.1 Regional and local contributions to MDA8 O_3 (Method 1-PCA)

After cleaning the data of all AQMSs in the SD region from 2018 to 2020, 66 sites fulfilled the data-quality requirements. We ran three independent PCAs on the MDA8 O_3 at these sites per year, and only the components with eigenvalues greater than 1

were judged as the main components. The results of this analysis are summarized in Table 2. The PCA resulted in four components for MDA8 O_3 over 3 years: the first principal component could explain the highest percentage of the variance of O_3, nearly 80%, and the cumulative variance of the four principal components reached more than 90%.

A notable clustering phenomenon was observed when we mapped the principal component loadings for each site in Fig. 1 to reveal its spatial distribution characteristics. Different colors represent different principal components, and the coefficients ranging from -1 to +1 represent the mean contribution of each component to each site during 2018-2020. The length of the column represents the size of the load, with the upward direction corresponding to positive values and the downward to negative values. The amplitude (scores) and loadings jointly determine the daily increase or decrease in the O_3 concentration at a certain site. The loading coefficient and amplitude with the same positive and negative mean that the O_3 concentration at the site increases, and vice versa, the O_3 concentration at the site decreases. By comparing the spatial and temporal information provided by the scores and loadings with meteorological data such as wind and temperature, we could infer the potential physical and chemical processes.

The spatial distribution of loadings is shown in Fig. 1, where the loadings of the first principal component (PC1) were all positive. Loadings associated with each principal component using Method 1 are presented in Table S1. The loadings range from +0.63 to +0.97 and PC1 averagely accounts for 77.8% of the variance at each of the 66 sites. The widespread cluster of PC1 suggests that the O_3 and PC1 values at the sites were mostly controlled by the regional background O_3. This interpretation is supported by Fig. 2, which compares the PC1 amplitudes against the NCEP winds. For PC1, the spatial load coefficients of all sites were positive; according to the vector scatter plot of PC1 amplitude and wind speed, the principal components on O_3 pollution days were positive as well, indicating that PC1 contributed positively to the O_3 concentration at all sites. PC1 represents the regional background, and the southerly wind prevails on the day when the O_3 exceeds the ambient air quality standards of China (160 $\mu g/m^3$).

The positive loadings of PC2 are distributed in the coastal area, which shows that PC2 contributes significantly to the stations in the coastal area and is largely influenced by the sea and land breezes. Based on the relationship between PC2 scores and meteorological variables, the influence of WS and WD on PC2 was analyzed. On the monthly scale, PC2 scores were low in the high O_3 season, which was related to air mass transportation in the eastern coastal area. Therefore, we interpreted that PC2 represented mainly the land-sea breeze circulation. The spatial distributions of PC3 and PC4 with positive loadings also showed an obvious feature: PC3 was mainly distributed in the northern region of the SD region, and PC4 was low in the central region and high on the east and west sides. Thus, PC3 and PC4 might be affected mostly by the local photochemistry. Based on the temporal variation in PC3 and PC4 scores and their relationship with meteorological variables (Fig. S2-3), there is no clear characteristic that may be influenced by either specific meteorology or regional transportation, it is therefore named as contributions from local generation.

Table2. Results of PCA Analysis (Method 1).

	2018			2019			2020		
PC	Eigenvalues	Variance Contribution	Cumulative Variance	Eigenvalues	Variance Contribution	Cumulative Variance	Eigenvalues	Variance Contribution	Cumulative Variance
PC1	50.99	72.26	72.26	54.09	81.95	81.95	52.26	79.18	79.18
PC2	5.68	8.60	85.87	4.22	6.40	88.35	4.67	7.07	86.25
PC3	1.98	3.00	88.87	1.66	2.52	90.87	2.39	3.62	89.87
PC4	1.10	1.67	90.54	1.005	1.52	92.39	1.16	1.76	91.63

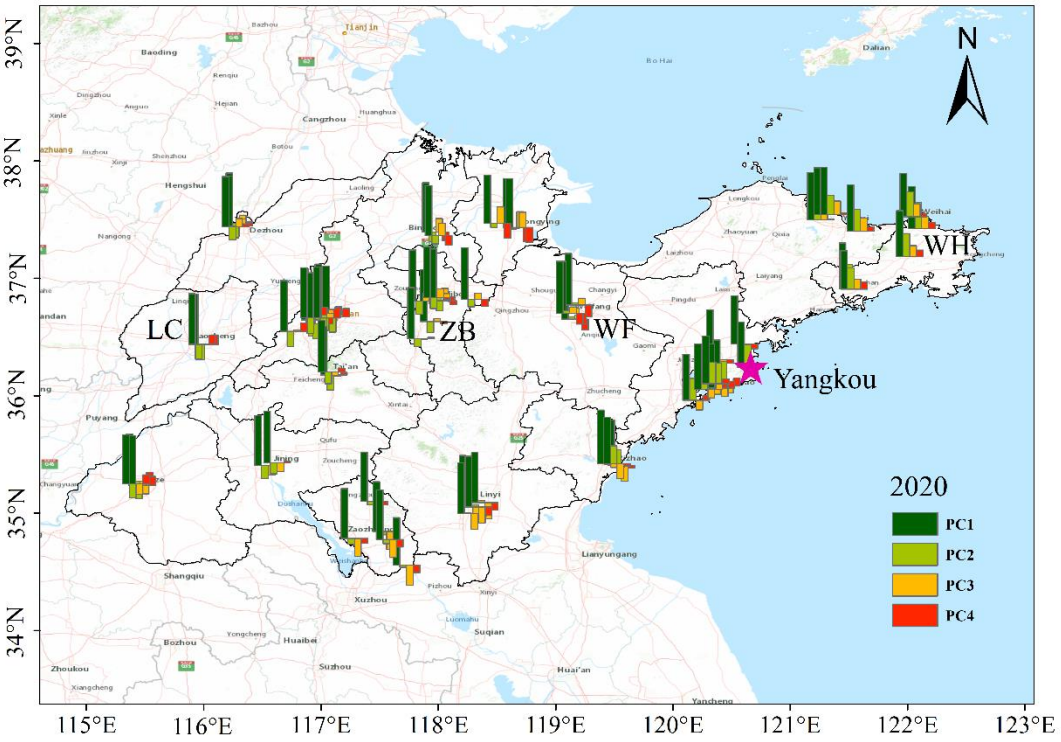

Figure 1. Spatial distribution of component coefficients (loadings) for PC1, PC2, PC3, and PC4 for 2020. Column length represents the size of the loading. Specific information of the loadings values for 2018-2019 and 2020 can be found in Table S1. And the purple star marks the location of the background site (Yangkou).

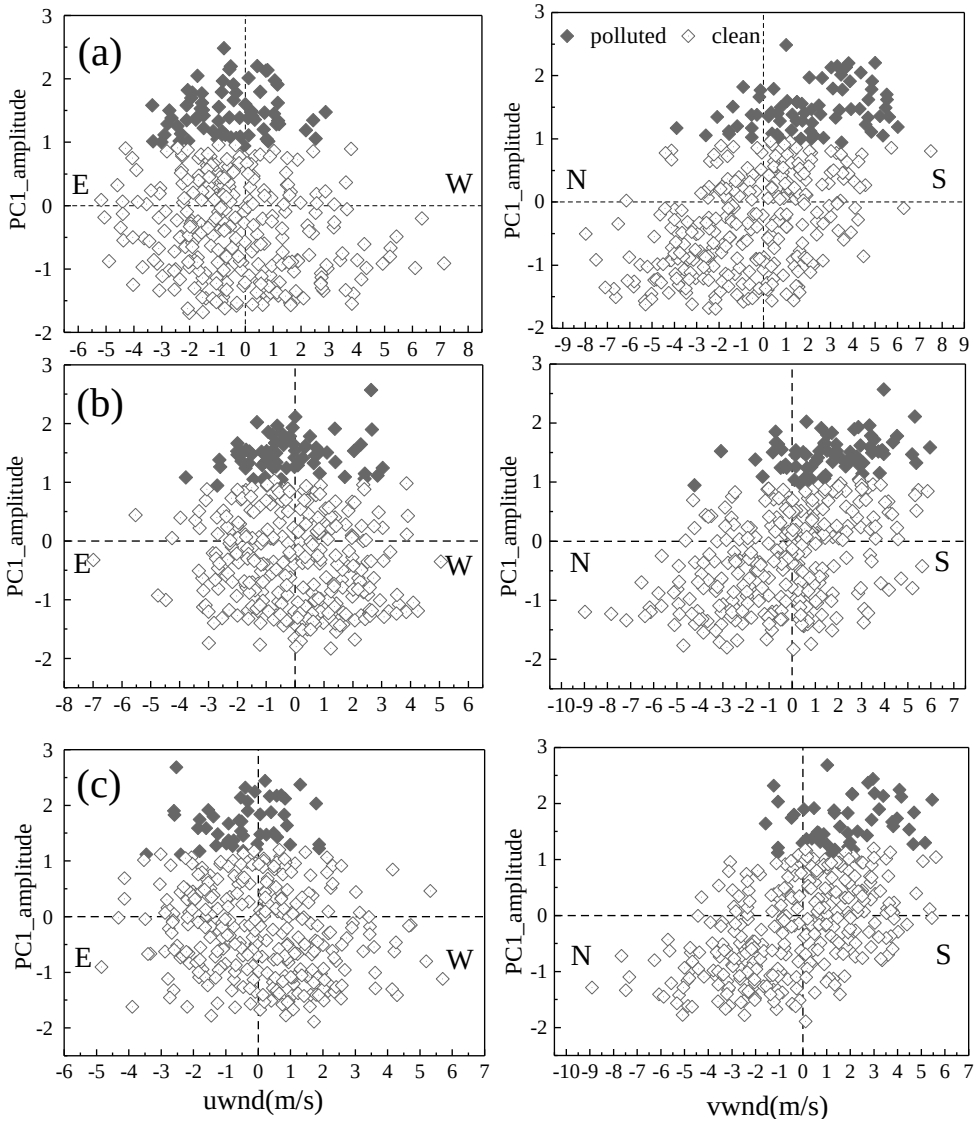

Figure 2. Scatterplots of correlations between PC1 amplitudes and mean NCEP reanalysis winds. Solid diamonds represent O_3 clean conditions (O_3<160μg/m^3); open diamonds represent O_3 polluted conditions. (a)-(c) represent 2018-2020, respectively.

Based on the spatial distribution of the sites in Fig. 1 (d), four sites (Weihai: SDFX; Weifang: HTJCZ; Zibo: DFHGC; Liaocheng: QZF) with relatively complete data were selected and marked on the map. The O_3 season (April-September) was used to illustrate the changes from inland to coastal areas. The difference between the measurements and the regional background O_3 represents the local contribution, which includes not only locally produced O_3 but also the O_3 in air transported to the site by local circulation. As shown in Fig. 3, the local contribution increases as the distance from the coast increases from Weihai to Liaocheng; this was expected based

on the gradient created by land-sea breeze circulation (PC2). In summer, the PC2 amplitude was mostly negative (Fig. S1), and the local contribution to Weihai becomes negative when O_3 from the ocean is transported to this region. In April, May, and September, PC2 was generally positive, and the local contribution increased in Weihai.

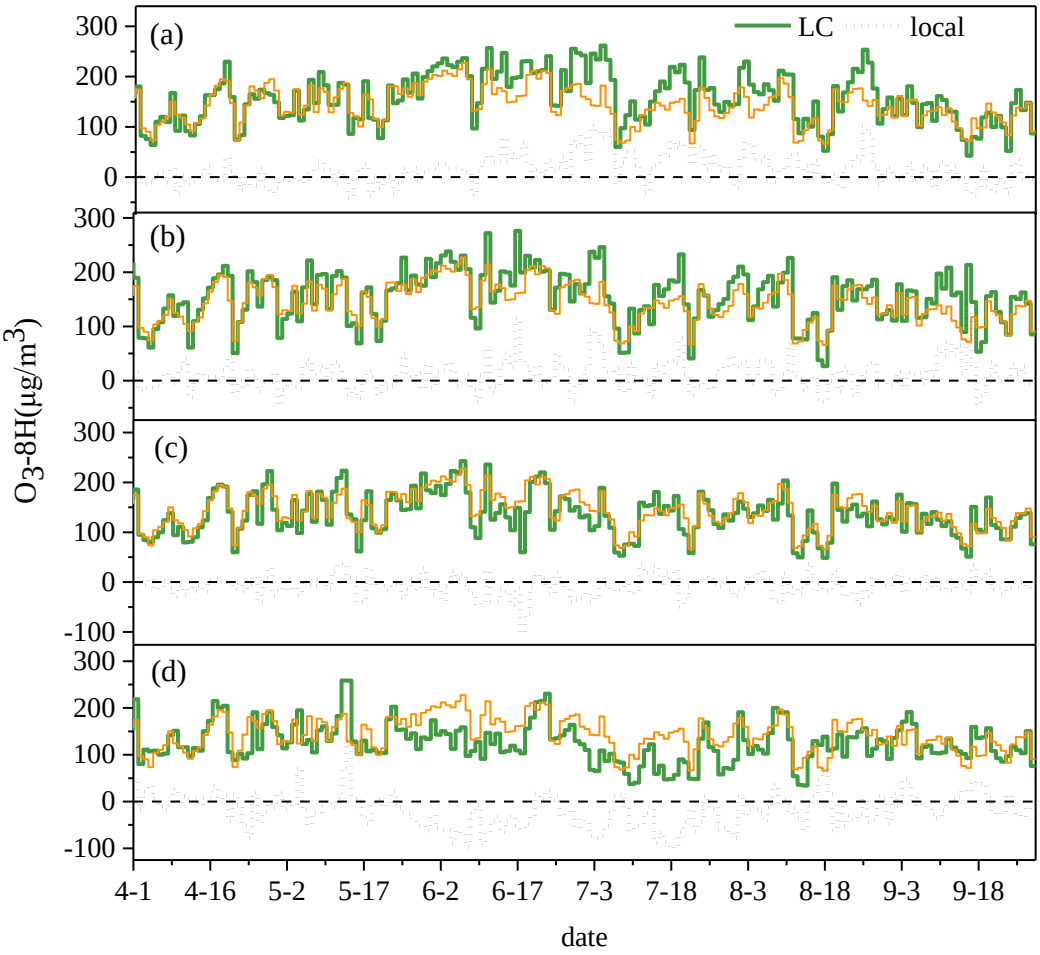

Figure 3. Daily 8-h maximum O_3 measured at Liaocheng (a), Zibo (b), Weifang (c), Weihai (d), compared with the regional background O_3 from inland to coast derived from PCA. The thin orange line represents the O_3 regional background, and the thick green line represents the O_3 concentration in the city. The dotted line represents the local contribution defined as the difference between the measurements and the regional background O_3.

In addition to the differences in local contributions from inland to coastal cities, the contribution of sea-land winds to coastal cities and locally generated O_3 in inland cities were further explored. Fig. 3 shows that the coastal areas are usually affected by

sea-land wind from June to August. Therefore, we considered the impact of sea-land wind to O_3 in coastal cities from June to August during 2018-2020, and the ozone concentration affected by sea-land wind is calculated using equation (1). Its contribution is calculated using the average value of ozone concentration in coastal cities, and the results show that the contribution of sea-land wind to coastal cities in the past three years was 3.3%, 1.8%, 1.3%, respectively. Ozone pollution days during the ozone season (April-September) were chosen to illustrate the contribution of local generation to O_3 in inland cities. Results indicate that the local generation of O_3 during ozone season in 2018, 2019 and 2020 was 35.5%, 29.0%, and 24.7%, respectively, while during O_3 pollution days, the contribution of local formation increased to 50.3%, 43.9%, 55.8%, in the year 2018, 2019, and 2020, respectively.

As shown in Fig. 4, the seasonal variations in the regional background O_3 showed the characteristic pattern of summer > spring > autumn > winter from 2018 to 2020. The regional background O_3 ranged from 60 to 152μg/m^3. In terms of interannual variation, the regional background O_3 decreased slightly in spring and summer during the 3 years and increased slightly in autumn and winter. The regional background O_3 during the Spring, Summer, Autumn, and Winter changed by -11, -8, +3, and +2 μg/m^3, respectively from 2018 to 2020.

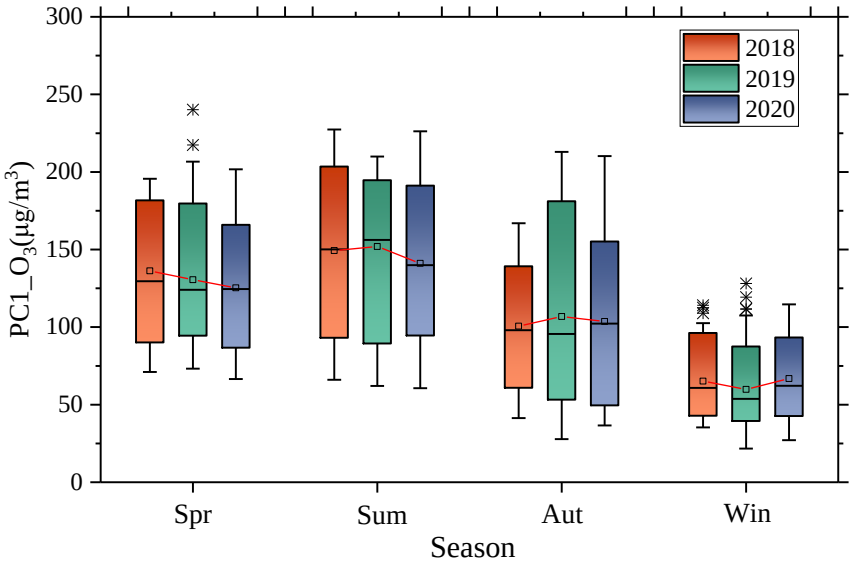

Figure 4. Regional background O_3 in different seasons of 2018-2020 (Method 1-PCA).

3.2 Regional and local contributions to MDA8 O_3 (Method 2-PCA)

Method 2 differs from Method 1 because it uses only MDA8 O_3, considers both O_3 precursors (NO_2) and meteorological variables (WS, WD, and T), and selects fewer sites (five sites). Additionally, site distribution is required, data must be complete. Data from five sites were used for the analysis: Zibo, Qingdao, Taian, Weihai, and Binzhou. PCA was performed on the five parameters: MDA8 O_3, daily mean NO_2, WD, WS, and T for the five sites from 2018 to 2020; the meteorological data sources were the daily mean data from the NCEP reanalysis data. The results of the PCA are shown in Table 3, where two components with eigenvalues greater than 1 were extracted for each site, and the eigenvalues of PCs from each site were similar; the mean value was approximately 1.6. The first component explained approximately 40% of the variance in the original variables, and the second component explained approximately 25% of the variance, indicating that both PCs were important in explaining the original variables.

Table 3. Results of PCA Analysis (Method 2).

City	PC	Eigenvalue	Variance Contribution	Cumulative Variance
Zibo	PC1	2.324	46.475	46.475
	PC2	1.106	22.122	68.597
Qingdao	PC1	2.071	41.420	41.420
	PC2	1.170	23.405	64.825
Taian	PC1	2.390	47.796	47.794
	PC2	1.140	22.810	70.604
Weihai	PC1	1.835	36.692	36.679
	PC2	1.275	25.508	62.201
Binzhou	PC1	2.155	43.092	43.092
	PC2	1.224	24.486	67.579

We infer the meaning of the components by considering the relationship between each principal component loading (absolute values greater than or equal to 0.5) and the variables. From the loadings of the two principal components at each site (Table 4), a clear pattern emerges: for each site, PC1 has high loadings on the factors O_3, NO_2, and T, reflecting the chemical generation process, and PC2 at all sites had larger values on the factors WD and WS, reflecting the physical transport process. We used

Weihai as an example: PC1 scores increased with O_3 and T, whereas NO_2 decreased (Fig. S4), which reflects NO_2 depletion and describes the chemistry, possibly local O_3 production. PC2 scores did not have a significant relationship with T and they increased with WS (Fig. S5), which reflects regional transport effects. Thus, further evidence suggests that PC1 and PC2 were primarily associated with chemical processes and physical processes, respectively.

Table 4. Loading or correlations of components with variables at each site by Method 2.

City	PC1					PC2				
	O_3	NO_2	T	WD	WS	O_3	NO_2	T	WD	WS
	μg/m^3	μg/m^3	℃	°	m/s	μg/m^3	μg/m^3	℃	°	m/s
Zibo	**0.916**	**-0.674**	**0.881**	0.498	0.081	0.091	**0.403**	0.014	0.490	**-0.834**
Qingdao	**0.775**	**-0.582**	**0.900**	0.561	-0.087	0.213	**0.596**	-0.005	0.199	**-0.854**
Taian	**0.892**	**-0.726**	**0.922**	0.443	0.139	0.142	**0.426**	0.036	0.581	**-0.775**
Weihai	**0.731**	**-0.210**	**0.807**	0.662	-0.409	0.031	**-0.854**	0.112	0.009	**0.729**
Binzhou	**0.922**	**-0.488**	**0.909**	0.484	0.075	0.087	**0.695**	-0.050	0.704	**-0.486**

Finally, based on the PCA results, referring to the method of Suciu et al. [*Suciu et al.*, *2017*], the PC scores for regional background O_3 were substituted as the mean of PC2 scores at each site, the PC scores for local contributions were replaced by the mean of PC1 scores at the site, and the cumulative contribution of the PCs was replaced by the results of the standardization of each component. Based on this calculation method, the regional background O_3 was back-calculated, and the results are shown in Fig. 5. Compared with other methods, there was no significant seasonal trend, and the regional background O_3 was approximately 110 μg/m^3 for each month, this result will be compared with previous studies.

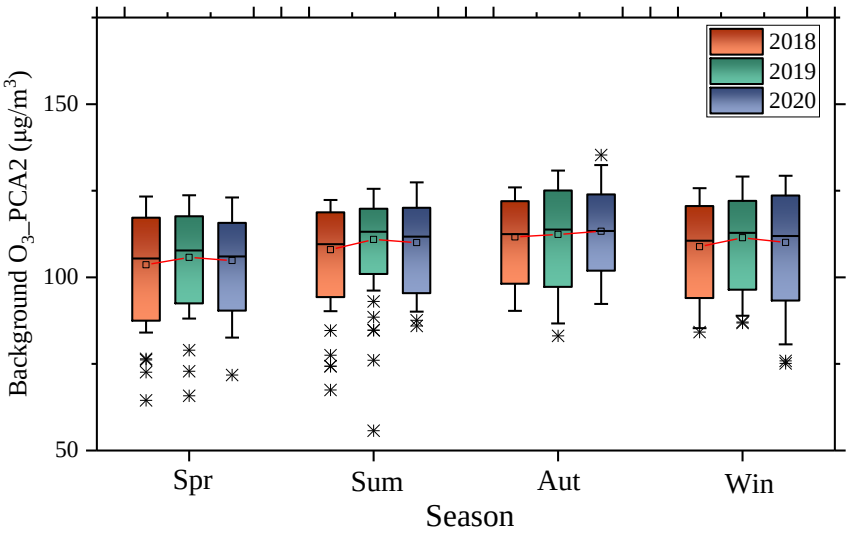

Figure 5. Regional background O_3 in the seasons of 2018-2020 (Method 2-PCA).

3.3 Regional and local contributions to MDA8 O_3 (Method 3-PCA/MLR)

PCA/MLR (Method 3), as a relatively novel method, uses the idea of source resolution and continues to use MLR to estimate the O_3 regional background based on the results of Method 1 (PC1 represents the regional background). Using the factor score of Method 1 as the independent variable, and the standardized results of the mean MDA8 O_3 of 66 AQMS sites in the SD region as the dependent variable, after MLR processing, the contribution proportions of the 2018-2020 O_3 regional background were obtained as follows: 60.2%, 57.3%, and 57.3%, which show a decreasing trend; the O_3 regional background was then calculated using equation (10) of Bian et al.[*Bian et al.*, 2013], and the results are shown in Fig. 6. The seasonal pattern of regional background O_3 remained consistent with that of Method 1, but the annual variation varied slightly by the season, especially in summer, for which there was a decreasing trend from year to year.

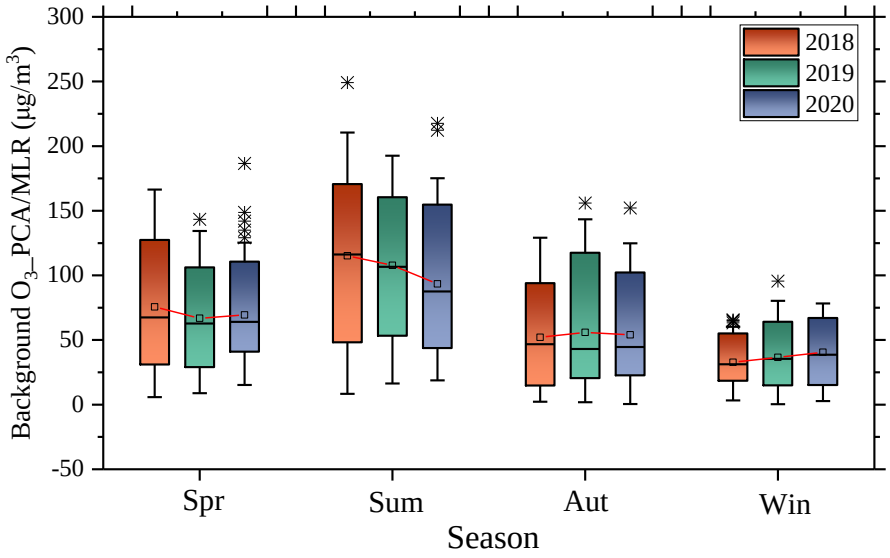

Figure 6. Regional background O_3 in the seasons of 2018-2020 (Method 3-PCA/MLR).

3.4 Regional and local contributions to MDA8 O_3 (Method 4-TCEQ)

The TCEQ method was used to estimate the regional background O_3 in the SD region. Because the lowest MDA8 O_3 at the AQMS selected by the TCEQ method represents the regional background value, to reduce the inaccuracy caused by the site, the distribution of the minimum MDA8 O_3 at all sites was calculated. Fig. 7 shows the regional background O_3 for the seasons from 2018-2020, and a characteristic pattern of summer > spring > autumn > winter can be observed, which is consistent with the pattern of results resolved by the PCA, ranging from 20 to 173 μg/m^3. In addition, summer and autumn show a trend of increasing and then decreasing, and the other two seasons show the opposite pattern, which is slightly different from that of the PCA, but overall, they all show a slightly increasing trend.

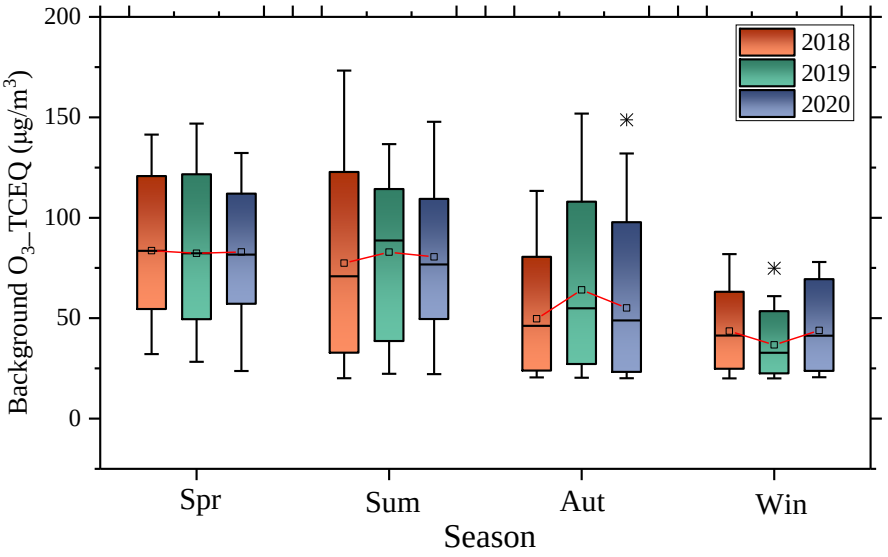

Figure 7. Regional background O_3 in the seasons of 2018-2020 (Method 4-TCEQ).

To further illustrate the contribution of regional background O_3 to coastal and inland cities in different years and seasons, the mean MDA8 O_3 was calculated for all AQMSs in coastal and inland cities in the SD region, and the ratio of the regional background O_3 to the mean MDA8 O_3 was defined to reflect the magnitude of the contribution of the regional background to O_3 concentration. As shown in Table 5, the contribution of the regional background to the coast is higher than that inland, which is consistent with the conclusion in PCA Method 1 that the local contribution from the coast to the inland is increasing. Regarding interannual variability, the regional contribution of O_3 to both showed an increasing trend, and for seasonal variability, it decreased sequentially in spring, summer, autumn, and winter.

Table 5. Contribution of regional background O_3 to coastal and inland cities in the seasons of 2018-2020.

Season	Coastal cities			Inland cities		
	2018	2019	2020	2018	2019	2020
Spring	65.1%	66.8%	72.0%	62.2%	63.6%	67.1%
Summer	57.6%	61.9%	63.7%	47.0%	52.2%	55.6%
Autumn	43.1%	54.7%	45.4%	42.8%	54.2%	44.0%
Winter	41.8%	43.2%	52.7%	41.3%	41.9%	51.1%

3.5 Comparisons among multiple methods

Due to the differences in the principles used to estimate the regional background O_3 concentrations, there were differences in the calculated results. In general, the results of Methods 1 and 2 were approximately 40 μg/m^3 larger than those of Method 3 and the TCEQ. For Method 3, this difference was mainly because after the MLR process, the resulting regional contribution decreased, only about 60%, therefore, the results were smaller when further estimating the regional background O_3. In the TCEQ method, the lowest MDA8 O_3 was selected to represent the regional background, and the selected site may be influenced by urban sites that do not capture the regional background well and are therefore lower in magnitude compared with Methods 1 and 2.

As shown by the prior analysis of the results, the seasonal trends of the regional background O_3 were generally consistent for the three methods, with a clear monthly variation characteristic, except for Method 2, which adds meteorological parameters as constraints and has a smooth trend. This phenomenon may be because Method 2 considers the meteorological factors of the station, indicating that the main component of the regional background value has almost no relationship with temperature. Therefore, there is no obvious monthly variation trend. The results of each method for interannual variability are presented in Table 6. Because Methods 1 and 3 use different analysis methods for the same dataset, the annual trends for both are consistent, showing a slight decrease of 3.8 μg/m^3 and 5.2 μg/m^3 for each of the three years. The results of Method 2, the TCEQ method, and the background sites show a consistent pattern of increasing and then decreasing, but overall, the values increase by 1.6 μg/m^3, 0.9 μg/m^3, and 14.7 μg/m^3, respectively, over the 3 years. The annual pattern of change for each method varied, but in general, there was an increase over the 3 years. Additionally, to reduce the error of a single method, the average value of the aforementioned results is expressed as the regional background O_3 in the SD region in the past 3 years, which were 89.2 μg/m^3, 89.8 μg/m^3 and 87.6 μg/m^3, respectively, and the three-year average value is 88.9 μg/m^3

Table 6. Comparison of all approaches in this study and the literature.

Method	2018	2019	2020	Average

	AVE	SD	AVE	SD	AVE	SD	AVE	SD
Method 1 (PCA)	113.1	45.9	112.6	49.6	109.3	43.5	111.7	46.4
Method 2 (PCA)	108.0	10.5	110.2	10.5	109.6	10.3	109.3	10.5
Method 3 (PCL/MLR)	69.7	45.8	67.6	42.0	64.6	36.6	67.3	41.7
Method 4 (TCEQ)	66.1	31.2	68.8	31.9	67.0	29.3	67.3	30.8
Average	89.2	42.3	89.8	42.5	87.6	39.1	88.9	10.5
Background site	91.5	35.1	104.1	39.6	106.2	34.3	100.6	37.0

[a] AVE represents the average value of a method for a given year; [b] SD represents the standard deviation of a method for a given year.

3.6 Comparisons with previous studies

Fig.8 summarizes the regional background O_3 concentrations reported in previous studies estimated by different methods for various regions. The results of several methods are within reasonable limits based on comparisons with other studies. We compared the results of this study with other studies, such as those of Liang et al. [*Liang et al.*, 2018] in the YRD region, Huang et al. [*Huang et al.*, 2021] in Shanghai, Xue et al. [*Xue et al.*, 2014] in Hong Kong, Berlin et al. [*Berlin et al.*, 2013] and Suciu et al. [*Suciu et al.*, 2017] in the Houston–Galveston–Brazoria (HGB) region, and Souri et al. [*Souri et al.*, 2016] in the Houston. Method 2 is referenced in Suciu's Method 2， from our results, PC2 represents the regional background, which differs from the results of Suciu [*Suciu et al.*, 2017], and the seasonal variation in regional background O_3 is not significant but remains similar in the magnitude of the background values. Compared with the results of Berlin et al[*Berlin et al.*, 2013], the region background O_3 calculated by the same PCA method is slightly larger than that of TCEQ, roughly 20 μg/m^3. Our results are significantly lower than those of Souri et al.[*Souri et al.*, 2016], who focused on the regional background O_3 at different wind directions and showed that the regional background O_3 were greatest at east-northeast winds. The results of Liang et al. and Huang et al. do not express specific regional background O_3 concentrations, but they report ranges of 68.8 to 154.7 μg/m^3 and 66.38 to 219.83 μg/m^3, respectively, and the maximum values are higher than the

overall average value because both study periods are in the O_3 season.

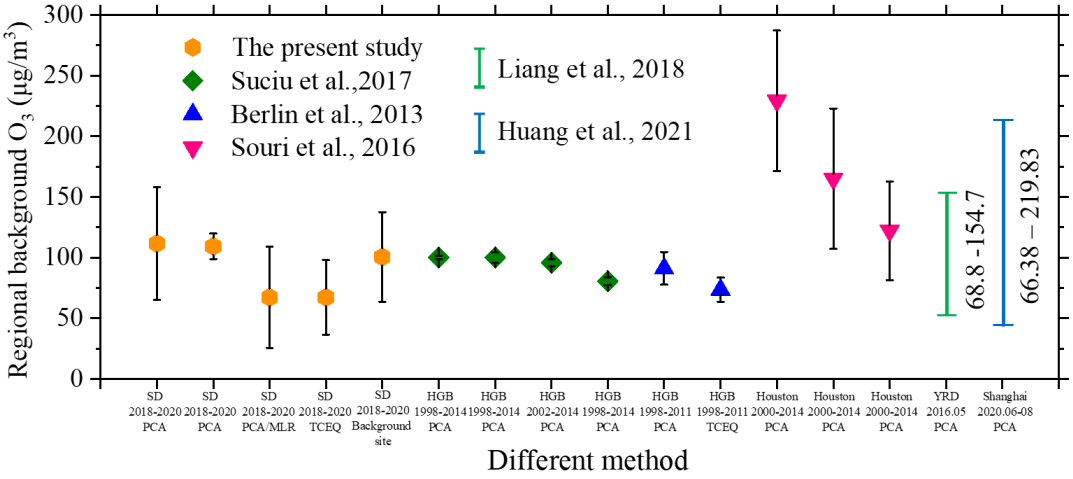

Figure 8. Comparison between this study and other research [*Berlin et al.*, 2013; *Huang et al.*, 2021; *Liang et al.*, 2018; *Souri et al.*, 2016; *Suciu et al.*, 2017].

4. Conclusions

Three PCA methods with differernet parameters and a TCEQ method was used to estimate the regional background O_3 concentrations in the SD region, where ozone pollution is severe in recent years. The regional background O_3 calculated using different PCA and TCEQ methods did not differ significantly and showed an overall consistent trend. Method 1 is the most commonly used method for resolving regional background O_3 using PCA and produces the highest O_3 concentration. Method 2 incorporates NO_2 and meteorological parameters as constraints and yields a flat monthly trend. Method 3 combines PCA with MLR and resolves relatively smaller O_3 background concentrations, which makes it close to the result of the TCEQ. Based on the results of the four methods and background sites, the three-year regional background O_3 showed an overall increasing trend, and the three-year average values for Method1, 2, 3, 4 and background in-situ measurement were $111.7 \pm 46.4 \mu g/m^3$, $109.3 \pm 10.5 \mu g/m^3$, $67.3 \pm 41.7 \mu g/m^3$, $67.3 \pm 30.8 \mu g/m^3$, $100.6 \pm 37.0 \mu g/m^3$, respectively. There was a clear seasonal pattern of regional O_3 background, with high values in spring and summer and low values in autumn and winter. Furthermore, the regional background O_3 differs spatially with the eastern coastal area seeing more influences from the marine environment. The concentration of locally generated O_3 gradually increased from coastal to inland cities while the opposite is observed for regional ozone contribution. Uncertainties exist in terms of estimating the regional

522 background O_3 concentrations. Additional factors can also be considered for
523 multivariate analysis, such as adding constraints on the precursor (e.g., VOC) and
524 additional relevant meteorological variables (e.g., solar radiation, relative humidity).
525 Further research is necessary to reduce these uncertainties.
526

527 **Author contributions**

528 L. Li designed and led the research. F.T. Wang performed the data analysis and
529 prepared the manuscript with contributions from all co-authors. All co-authors
530 contributed to the discussions. L. Li and Joshua S Fu reviewed the paper.

531 **Competing interests**

532 The authors declare that they have no conflict of interest.

533 **Acknowledgement**

534 This study was financially supported by the National Key Research on Air Pollution
535 Formation and Control (DQGG202119), the National Natural Science Foundation of
536 China (42075144).

537 **References**

538 Abdul-Wahab, S. A., C. S. Bakheit, and S. M. Al-Alawi (2005), Principal component
539 and multiple regression analysis in modelling of ground-level ozone and factors
540 affecting its concentrations, *Environmental Modelling & Software*, *20*(10), 1263-1271,
541 doi:10.1016/j.envsoft.2004.09.001.
542 Berlin, S. R., A. O. Langford, M. Estes, M. Dong, and D. D. Parrish (2013),
543 Magnitude, decadal changes, and impact of regional background ozone transported
544 into the greater Houston, Texas, area, *Environ Sci Technol*, *47*(24), 13985-13992,
545 doi:10.1021/es4037644.
546 Bian, L., T. Li, and J. Hou (2013), Source Apportionment of Polycyclic Aromatic
547 Hydrocarbons Using Two Mathematical Models for Runoff of the Shanghai Elevated
548 Inner Highway, China, *Environmental Science*, *34*(10), 3840-3846,
549 doi:10.13227/j.hjkx.2013.10.027.
550 Chan, E., and R. J. Vet (2010), Baseline levels and trends of ground level ozone in
551 Canada and the United States, *Atmospheric Chemistry and Physics*, *10*(18),
552 8629-8647, doi:10.5194/acp-10-8629-2010.
553 Chen, T. M., J. Gokhale, S. Shofer, and W. G. Kuschner (2007), Outdoor air pollution:
554 Ozone health effects, *American Journal of the Medical Sciences*, *333*(4), 244-248,

555 doi:10.1097/MAJ.0b013e31803b8e8c.
556 Chu, B. W., et al. (2020), Air Pollutant Correlations in China: Secondary Air Pollutant
557 Responses to NOx and SO2 Control, *Environmental Science & Technology Letters*,
558 *7*(10), 695-700, doi:10.1021/acs.estlett.0c00403.
559 Dai, H., J. Zhu, H. Liao, J. Li, M. Liang, Y. Yang, and X. Yue (2021), Co-occurrence
560 of ozone and PM2.5 pollution in the Yangtze River Delta over 2013–2019:
561 Spatiotemporal distribution and meteorological conditions, *Atmospheric Research*,
562 *249*, doi:10.1016/j.atmosres.2020.105363.
563 Dang, R. J., and H. Liao (2019), Radiative Forcing and Health Impact of Aerosols and
564 Ozone in China as the Consequence of Clean Air Actions over 2012-2017,
565 *Geophysical Research Letters*, *46*(21), 12511-12519, doi:10.1029/2019gl084605.
566 Feng, J. S., N. N. Song, Y. X. Yu, and Y. X. Li (2020), Differential analysis of
567 FA-NNC, PCA-MLR, and PMF methods applied in source apportionment of PAHs in
568 street dust, *Environmental Monitoring and Assessment*, *192*(11),
569 doi:10.1007/s10661-020-08679-3.
570 Huang, Q., Y. Huang, S. Zhang, D. JIn, S. Gao, and G. Xiu (2021), O_3 Source
571 Characteristics of Industrial Area in Yangtze River Delta Based on Boundary
572 Observation, *Environmental Science*, *42*(10), 4621-4631,
573 doi:10.13227/j.hjkx.202101199.
574 Jolliffe, I. J. E. o. s. i. b. s. (2005), Principal component analysis.
575 Langford, A. O., C. J. Senff, R. M. Banta, R. M. Hardesty, R. J. Alvarez, S. P.
576 Sandberg, and L. S. Darby (2009), Regional and local background ozone in Houston
577 during Texas Air Quality Study 2006, *Journal of Geophysical Research*, *114*,
578 doi:10.1029/2008jd011687.
579 Liang, Y., Y. Liu, H. Wang, L. Li, Y. Duan, and K. Lu (2018), Regional
580 characteristics of ground-level ozone in Shanghai based on PCA analysis, *Acta*
581 *Scientiae Circumstantiae*, *38*(10), 3807-3815, doi:10.13671/j.hjkxxb.2018.0209.
582 Ma, Z. Q., J. Xu, W. J. Quan, Z. Y. Zhang, W. L. Lin, and X. B. Xu (2016), Significant
583 increase of surface ozone at a rural site, north of eastern China, *Atmospheric*
584 *Chemistry and Physics*, *16*(6), 3969-3977, doi:10.5194/acp-16-3969-2016.
585 McDonald-Buller, E. C., et al. (2011), Establishing Policy Relevant Background (PRB)
586 Ozone Concentrations in the United States, *Environmental Science & Technology*,
587 *45*(22), 9484-9497, doi:10.1021/es2022818.
588 Morgenstern, O., G. Zeng, S. M. Dean, M. Joshi, N. L. Abraham, and A. Osprey
589 (2014), Direct and ozone-mediated forcing of the Southern Annular Mode by
590 greenhouse gases, *Geophysical Research Letters*, *41*(24), 9050-9057,
591 doi:10.1002/2014gl062140.
592 Mousavinezhad, S., Y. Choi, A. Pouyaei, M. Ghahremanloo, and D. L. Nelson (2021),
593 A comprehensive investigation of surface ozone pollution in China, 2015-2019:
594 Separating the contributions from meteorology and precursor emissions, *Atmospheric*
595 *Research*, *257*, doi:10.1016/j.atmosres.2021.105599.
596 MURTAGH, F., A. J. A. HECK, and s. s. library (1987), Multivariate data analysis,

131.
Nielsen-Gammon, J. W., J. Tobin, and A. J. g. Mcneel (2005), A Conceptual Model for Eight-Hour Ozone Exceedances in Houston, Texas Part II: Eight-Hour Ozone Exceedances in the Houston-Galveston Metropolitan Area.
Nielsengammon, J., J. Tobin, A. Mcneel, and G. Li (2005), A Conceptual Model for Eight-Hour Ozone Exceedances in Houston, Texas Part I: Background Ozone Levels in Eastern Texas.
Ottosen, T. B., and P. Kumar (2019), Outlier detection and gap filling methodologies for low- cost air quality measurements, *Environmental Science-Processes & Impacts*, *21*(4), 701-713, doi:10.1039/c8em00593a.
Schauberger, B., S. Rolinski, S. Schaphoff, and C. Muller (2019), Global historical soybean and wheat yield loss estimates from ozone pollution considering water and temperature as modifying effects, *Agricultural and Forest Meteorology*, *265*, 1-15, doi:10.1016/j.agrformet.2018.11.004.
Shamsipour, M., et al. (2014), A Framework for Exploration and Cleaning of Environmental Data – Tehran Air Quality Data Experience, *17*(12), 0-0.
Shan, W., Y. Yin, H. Lu, and S. Liang (2009), A meteorological analysis of ozone episodes using HYSPLIT model and surface data, *Atmospheric Research*, *93*(4), 767-776, doi:10.1016/j.atmosres.2009.03.007.
Souri, A. H., Y. Choi, X. Li, A. Kotsakis, and X. Jiang (2016), A 15-year climatology of wind pattern impacts on surface ozone in Houston, Texas, *Atmospheric Research*, *174-175*, 124-134, doi:10.1016/j.atmosres.2016.02.007.
Statheropoulos, M., N. Vassiliadis, and A. J. A. e. Pappa (1998), Principal component and canonical correlation analysis for examining air pollution and meteorological data, *32*(6), 1087-1095.
Suciu, L. G., R. J. Griffin, and C. A. Masiello (2017), Regional background O3 and NOx in the Houston–Galveston–Brazoria (TX) region: a decadal-scale perspective, *Atmos. Chem. Phys.*, *17*(11), 6565-6581, doi:10.5194/acp-17-6565-2017.
Sun, J., et al. (2021), A comprehensive study on ozone pollution in a megacity in North China Plain during summertime: Observations, source attributions and ozone sensitivity, *Environment International*, *146*, doi:10.1016/j.envint.2020.106279.
Tai, A. P. K., and M. V. Martin (2017), Impacts of ozone air pollution and temperature extremes on crop yields: Spatial variability, adaptation and implications for future food security, *Atmospheric Environment*, *169*, 11-21, doi:10.1016/j.atmosenv.2017.09.002.
Vingarzan, R. (2004), A review of surface ozone background levels and trends, *Atmospheric Environment*, *38*(21), 3431-3442, doi:10.1016/j.atmosenv.2004.03.030.
Wang, J. Z., Y. Q. Yang, Y. M. Zhang, T. Niu, X. F. Jiang, Y. Q. Wang, and H. Z. Che (2019), Influence of meteorological conditions on explosive increase in O-3 concentration in troposphere, *Science of the Total Environment*, *652*, 1228-1241, doi:10.1016/j.scitotenv.2018.10.228.
WU, L., L. Xue, and W. Wang (2017), Review on the observation-based methods for

ozone air pollution research, *Journal of Earth Environment, 8*(06), 479-491, doi:10.7515/JEE201706001.
Xue, L., T. Wang, P. K. Louie, C. W. Luk, D. R. Blake, and Z. Xu (2014), Increasing external effects negate local efforts to control ozone air pollution: a case study of Hong Kong and implications for other Chinese cities, *Environ Sci Technol*, *48*(18), 10769-10775, doi:10.1021/es503278g.
Yao, Q., Z. Ma, T. Hao, W. Fan, X. Yang, Y. Tang, Z. Cai, and S. Han (2021), Temporal and spatial distribution characteristics and background concentration estimation of ozone in Beijing-Tianjin-Hebei region, *China Environmental Science*, *41*(11), 4999-5008, doi:10.19674/j.cnki.issn1000-6923.20210706.007.
Yin, C. Q., X. J. Deng, Y. Zou, F. Solmon, F. Li, and T. Deng (2019), Trend analysis of surface ozone at suburban Guangzhou, China, *Science of the Total Environment*, *695*, doi:10.1016/j.scitotenv.2019.133880.
Zhang, L., D. J. Jacob, N. V. Downey, D. A. Wood, D. Blewitt, C. C. Carouge, A. van Donkelaar, D. B. A. Jones, L. T. Murray, and Y. X. Wang (2011), Improved estimate of the policy-relevant background ozone in the United States using the GEOS-Chem global model with 1/2 degrees x 2/3 degrees horizontal resolution over North America, *Atmospheric Environment*, *45*(37), 6769-6776, doi:10.1016/j.atmosenv.2011.07.054.
Zhang, M., C. DING, Y. LI, G.-x. WANG, J.-j. LIN, H. MENG, and Y. XU (2021), Spatial and Temporal Distribution of Ozone and Influencing Factors in Shandong Province, *Environmental Science*, *42*(12), 5723-5735, doi:10.13227/j.hjkx.202102034.