One year monitoring of volatile organic compounds (VOCs) from an oil and gas station in northwest China

Huang Zheng [1,2], Shaofei Kong [1], Xinli Xing [2,3], Yao Mao [3], Tianpeng Hu [2], Yang Ding [2], Gang Li [4], Dantong Liu [5], Shuanglin Li [1], and Shihua Qi [1,3]

[1] Department of Atmospheric Sciences, School of Environmental Studies, China University of Geosciences, Wuhan, 430074, China
[2] Department of Environmental Science and Technology, School of Environmental Studies, China University of Geosciences, Wuhan, 430074, China
[3] State Key Laboratory of Biogeology and Environmental Geology, China University of Geosciences, Wuhan, 430074, China
[4] Karamay Environmental Monitoring Center Station, Karamay, 834000, China
[5] School of Earth and Environmental Sciences, the University of Manchester, M13 9PL, UK

Correspondence to: Shaofei Kong (kongshaofei@cug.edu.cn) or Xinli Xing (xingxinli5300225@163.com)

Abstract. Oil and natural gas are important for energy supply around the world. The exploring, drilling, transportation and processing in oil and gas regions can release a lot of volatile organic compounds (VOCs). To understand the VOC levels, compositions and sources in such region, an oil and gas station in northwest China was chosen as the research site and fifty-seven VOCs designed as the photochemical precursors were continuously measured for an entire year (September 2014–August 2015) using an on-line monitoring system. The average concentrations of total VOCs were 297 ± 372 ppbv and the main contributor was alkanes, accounting for 87.5% of the total VOCs. According to the propylene-equivalent concentration and maximum incremental reactivity methods, alkanes were identified as the most important VOC groups for the ozone formation potential. Positive matrix factorization (PMF) analysis showed that the annual average contributions from natural gas, fuel evaporation, combustion sources, oil refining process and asphalt (anthropogenic and natural sources) to the total VOCs were 62.6 ± 3.04%, 21.5 ± .99%, 10.9 ± 1.57%, 3.8 ± 0.50% and 1.3 ± 0.69%, respectively. The five identified VOC sources exhibited various diurnal variation patterns due to their different emission patterns and the impact of meteorological parameters. Potential source contribution function (PSCF) and contribution weighted trajectory (CWT) models based on backward trajectories indicated that the five identified sources had similar geographic origins. Raster analysis based on CWT analysis indicated that the local emissions contributed 48.4–74.6% to the total VOCs. Based on the high-resolution observation data, this study clearly described and analyzed the temporal variation of VOC emission characteristics at a typical oil and gas filed, which exhibited different atmospheric behaviors compared with that in urban and industrial areas.

***Keywords*:** Volatile organic compounds; oil and gas field; temporal variation; photochemical behavior; source apportionment; local-regional contribution

1 Introduction

Volatile organic compounds (VOCs) are ubiquitous in ambient air and originate from both natural processes (i.e., vegetation emission, volcanic eruption and forest fire) and anthropogenic activities such as the fossil fuel combustion, industrial processes and solvent usage (Cai et al., 2010; Leuchner and Rappenglück, 2010; Baudic et al., 2016). As the key precursors of O_3 formation (Fujita, 2001; Geng et al., 2008; Ran et al., 2009; Lyu et al., 2016), different VOC categories exhibited different ozone formation potential (Carter, 1994; Atkinson and Arey, 2003; Zou et al., 2015). Some VOC species (i.e., benzene) exhibit detrimental effects on human health (Colman Lerner et al., 2012; He et al., 2015) and they have negative impacts on air quality (Vega et al., 2011). Till now, research concerning atmospheric VOCs including their emission, atmospheric transformation, health impact and so on is still a hot topic around the world.

Previous studies in China mainly focused on the measurements of VOCs in urban agglomerations such as the Pearl River Delta (PRD) region (Tang et al., 2007; Liu et al., 2008; Cheng et al., 2010; Ling et al., 2011), Yangtze River Delta (YRD) region (An et al., 2014; Li et al., 2016; Shao et al., 2016) and Beijing-Tianjin-Hebei (BTH) region (Li et al., 2015) and key megacities including Beijing (Song et al., 2007; Wang et al., 2010; Yuan et al., 2010), Shanghai (Cai et al., 2010; Wang, 2014), Guangzhou (Zou et al., 2015) and Wuhan (Lyu et al., 2016). These studies found that vehicle emission and solvent usage contributed most to the ambient VOCs in urban areas. A few studies were also conducted in industrial areas (An et al., 2014; Wei et al., 2015; Shao et al., 2016) and petrochemical industrial regions with a lot of VOC emissions (Lin et al., 2004; Wei et al., 2015; Jia et al., 2016; Mo et al., 2017). These studies found that the VOC sources and compositions are complex due to the different emission and atmospheric processes (Warneke et al., 2014). However, the research conducted in oil and gas area in China is still limited while the VOCs emission characteristics in this type of regions are common around the world (Buzcu-Guven and Fraser, 2008; Simpson et al., 2010; Rutter et al., 2015; Bari et al., 2016). For instance, Leuchner and Rappenglück. (2010) found that natural gas/crude oil sources contributed most to the VOC emissions in Houston. Gilman et al. (2013) found that oil and gas emissions strongly contribute to the propane and butanes in northeast Colorado. Therefore, study concerning VOC emission characteristics studies in oil and gas area in China is very important.

In previous studies, the ambient air was sampled for a few days (weeks) or at a certain season with low time-resolution. The diurnal, monthly and seasonal variations were mostly overlooked, which prevented the understanding of the VOCs' temporal behaviors influenced by the real-time emission, photochemical reaction and meteorological condition. Therefore, a long-term monitoring with a high-time resolution of VOCs is desired (Baudic et al., 2016; Liu et al., 2016). It should be emphasized that in the September of 2013, the VOCs control at petrochemical regions has been listed as one of the main objectives of the Action Plan of Atmospheric Pollution Control released by the central government of China (http://www.gov.cn/zwgk/2013-09/12/content_2486773.htm), which proposed new requirements to conduct researches in this type of field.

To identify the VOC sources, receptor models including chemical mass balance (CMB), positive matrix factorization (PMF) and principal component analysis/absolute principal component scores (PCA/APCS) have been widely used (Guo et al., 2004; Rodolfo Sosa et al., 2009; An et al., 2014; Liu et al., 2016). Meanwhile, dispersion models including conditional probability

function (CPF), backward trajectory, potential source contribution function (PSCF) and concentration weighted trajectory (CWT) are also employed to locate the potential source origins (Song et al., 2007; Chan et al., 2011; Liu et al., 2016). Recently, the combination of these two types of models is developed to figure out the locations of various air pollutant sources (Zhang et al., 2013a; Bressi et al., 2014; Chen et al., 2016).These practices mainly focus on the atmospheric fine particles ($PM_{2.5}$), few studies have concerned the local and regional source contributions of VOCs.

In this study, an oil and gas field located in northwest China was chosen as the study area to conduct a long-term monitoring of VOCs with high-time resolution. The main objectives are to (1) compare VOC concentrations, compositions and ozone formation potential at this oil and gas station with other areas, (2) discuss the relationships between VOC concentrations and meteorological parameters in different time scale, (3) identify the possible VOC sources by PMF, and (4) identify the local source contributions and regional origins of VOCs based on PMF and dispersion models. This study is the first VOCs research with high-time resolution at the oil and gas fields in China, which provides new information on the temporal variation, ozone formation potential and local/regional contributions of VOCs and are helpful to establish control measures of VOCs at this type of region around the world.

2 Materials and methods

2.1 Site description

The study area (44.1–46.3 ºN and 84.7–86.0 ºE) is located in northwest China and at the northwestern margin of Junggar Basin, which is an important oil and gas bearing basin (Fig. 1a). The proven deposits of oil and natural gas are 2.41×10^9 t and 1.97×10^{11} m^3, respectively. There are hundreds of oil and gas wells in this field with an annual gas deliverability of 1.20×10^{10} m^3 (Chen, 2015). Additionally, 126 petrochemical plants spread across this area. This area can be divided into two regions with oil and gas operation and oil refinery at the north direction (Region 1) and petrochemical industry at the south direction (Region 2). These two regions are about 150 km in distance (Fig. 1b). Region 1 is abundant in oil and gas resources and the main petrochemical factories are oil refineries and natural gas chemical plants. The main products include gasoline, diesel, asphalt and 1, 3-butadiene. The flow charts of chemical process are shown in Fig. S1. Region 2 is a key petrochemical base, with the production capacity of oil and ethylene being 6×10^6 t yr^{-1} and 2.2×10^5 t yr^{-1}, respectively. The sampling site is located on the rooftop of a building (15 m above the ground, 45.6 ºN, 85 ºE), about 11 km away from the southeast of the urban region. At the northeast of the sampling site, there are hundreds of oil and gas wells (Fig. 1c). The study area is in the hinterland of Eurasia. The typical temperate continental arid desert climate results in high temperature in summer (27.9 ℃) and low temperature in winter (–15.4 ℃). The sufficient solar radiation, little precipitation and low humidity (43–56%) result in high evaporation (> 3000 mm) in this region.

2.2 Descriptions of instruments and QA/QC

From September 2014 to August 2015, 57 ambient VOCs designed as the O_3 precursors by the Photochemical Assessment Monitoring Stations (PAMS) were continuously sampled and measured using an online monitor system (TH-300B, Wuhan-Tianhong Instrument Co., Ltd, China) with two-hour time resolution. The sampling and analysis procedures were described elsewhere (Lyu et al., 2016). Briefly, two-channels were installed to analyze VOCs separately. The water and carbon dioxide in the sampled air was firstly removed at a cold trap maintaining at –80 ℃ and then concentrated at –150 ℃ at another cold trap. After the purification and concentration, the VOCs were desorbed by rapid heating to 100 ℃. The C_2–C_5 VOCs were separated with a PLOT column (diameter: 0.32 mm, thickness of membrane: 1.5 μm, length: 60 m) and were quantified by the gas chromatograph-flame ionization detector (GC-FID, Agilent 7890). C_5–C_{12} were separated by a DB-624 column (diameter: 0.25 mm, thickness of membrane: 3 μm and length: 60 m) and were quantified using mass spectrometer detector (MSD, Agilent 5975).

The target compounds involved 57 VOC species: alkanes (30), alkenes (9), alkynes (acetylene) and aromatics (17). The standard gases from PAMS were used for the equipment calibration and verification through the 5-point method every two weeks (Lyu et al., 2016). The correlation coefficients of the calibration curves usually varied from 0.991 to 0.998. The detection limits were in the range of 0.04 to 0.12 ppbv (Table 1). The missing value was due to power failure or instrument maintenance and was not included in the data analysis.

2.3 Data sources and analysis

2.3.1 Meteorological parameters and air pollutants

Other dataset such as the three-hour resolution meteorological parameters (atmospheric pressure (P), temperature (T), relative humidity (RH), wind speed (WS) and direction (WD)) were collected from the Meteomanz (www.meteomanz.com) and are shown in Fig. 2. The boundary layer height (BLH) was computed every three hours each day through the NOAA's READY Archived Meteorological website (http://www.ready.noaa.gov/READYamet.php).

The hourly CO, NO_2, O_3, SO_2, ambient inhalable particles (PM_{10}) and fine particles ($PM_{2.5}$) were measured using an ambient air quality continuous automated monitors (TH-2000 series, Wuhan-Tianhong Instrument Co., Ltd, China) and the data were acquired from the Qingyue Open Environmental Data Center (https://data.epmap.org). It should be noted that the NO_2 (NO_2 = NO_x–NO) concentrations were in fact overestimated. This is because some oxidized reactive nitrogen that is converted by the molybdenum during the NO_x measurement while the NO measurement is accurate using the chemiluminescence technic. Therefore, the NO_2 concentrations discussed below are considered greater than the actual values (Dunlea et al., 2007; Zou et al., 2015). According to the ambient air quality standards-Ⅱ (GB/3095-2012), the main air pollutants were PM_{10} and $PM_{2.5}$ in winter and NO_2 in autumn (Fig. S2).

2.3.2 VOCs source apportionment and ozone formation potential

Positive matrix factorization (PMF) model has been widely employed for VOCs source apportionment (Buzcu-Guven and Fraser, 2008; Leuchner and Rappenglück, 2010; Liu et al., 2016; Lyu et al., 2016). In this study, the EPA PMF 5.0 (US EPA, 2014) was employed and additional information was given in Appendix A.

The VOC concentrations are not proportional to the ozone formation potential due to their wide ranges of photochemical reactivity with OH radicals (Table 1). Two methods including propylene-equivalent concentrations (Propy-Equiv) and the maximum incremental reactivity (MIR) were adopted to analyze the ozone formation potential of VOCs. More details can be found in the research of Atkinson and Arey. (2003) and Zou et al. (2015).

2. 4 Geographic origins of the VOCs

2. 4.1 Conditional probability function (CPF)

The CPF is widely used to locate the direction of sources based on wind direction data (Song et al., 2007). In this study, the directions of various VOC sources were explored based on the *G* matrix in PMF analysis and wind directions. The CPF is defined as:

$$\mathrm{CPF} = \frac{m_{\Delta\theta}}{n_{\Delta\theta}} \quad (1)$$

where $m_{\Delta\theta}$ is the number of data from wind sector $\Delta\theta$ (each is 22.5 degree) that exceed the threshold value (75th percentile of each source contribution); $n_{\Delta\theta}$ is the total number of occurrence from the same wind direction. Calm conditions (wind speed < 1 m/s) were excluded from the calculation for its difficulty in defining the wind direction.

2.4.2 Backward trajectory analysis

The 48 h backward trajectories with 2 h intervals (starting from 0:00 to 20:00 local time, LT) were run each day by the TrajStat –plugin of Meteoinfo software (http://www.meteothinker.com/downloads/index.html) using the Hybrid Single Particle Lagrangian Integrated Trajectory (HYSPLIT) model (Wang et al., 2009; Squizzato and Masiol, 2015). The start height was set as 500 m above ground level (Zhao et al., 2015; Liu et al., 2016). The FNL global analysis data produced by the National Center for Environmental Prediction's Global Data Assimilation System (GDAS) wind field re-analysis was introduced into the calculation. A total of 2743 backward trajectories were generated and then were grouped into four clusters according to their geographic sources and histories. As shown in Fig. S3, the trajectories were mainly originated from northwest of the sampling site during the whole observation period.

2.4.3 Local and regional transport contribution

The PSCF and CWT model are previously used to identify the possible source regions based on the backward trajectory analysis (Cheng et al., 2013; Bressi et al., 2014; Liu et al., 2016). The PSCF gives the proportion of air pollution trajectory in

a given grid and the CWT reflects the concentration levels of trajectories. The geographic domain (31 º–71 ºN, 36 º–107 ºE) was found to be within the annual range of 48 h backward trajectories. The total number of grids was 11360 with a resolution of 0.5 º×0.5 º. More information about the PSCF and CWT analysis can be found in Appendix B.

Local and regional source contributions of the observed VOCs were calculated by raster analysis. In previous studies, the domain was divided into 12 sectors (each was 30 º) to study the regional contributions (Bari et al., 2003; Wang et al., 2015; Wang et al., 2016). However, in this study, the domain was briefly divided into two sections (local and regional), with the sampling site as the original point. The range of local sources was defined as a polar with a radius of 12 h backward trajectories and the range of regional sources was outside of the circle (detailed descriptions can be found in Appendix C). The concentration of each grid was calculated by CWT analysis. By counting and averaging in each section, contributions of local emission and regional transportation were produced. To reduce the effects of background values, the lowest CWT values (C_b) in each section was deduced from the concentrations. The contribution (%) of local source and regional transportation was defined as follows:

$$\%C_{\mathrm{i}} = \frac{(C_i - C_{bi}) \times N_i}{\sum_{i=1}^{2}(C_i - C_{bi}) \times N_i} \times 100\% \quad (2)$$

where C_i is the mean CWT value in the *i*th section (local or regional), C_{bi} is the background value of the *i*th section, N_i is the number of grid with non-zero CWT concentrations in the *i*th section.

Several factors affect the calculated results of regional and local source contributions, including the radius of the circle and CWT value. In this study, the 12 h backward trajectories were chosen to differentiate the local area from regional area. In fact, the longer the backward trajectories were, the lower regional contributions produced. In addition, the PMF model was employed to VOCs source apportionment and the contribution of each identified source was introduced into CWT calculation. However, the negative value of source contribution was inevitably generated despite the application of F-peak in PMF analysis. Therefore, the negative CWT value was excluded in raster analysis and this would affect the results of regional and local source contributions. Overall, although flaws existed in this new method, it gave a new insight to understand the quantitative contributions of local and regional contribution to the VOCs in the study area.

3 Results and discussions

3.1 VOCs levels and compositions

The statistics of observed VOCs are summarized in Table 1 and every two hours' variation of four VOC categories are shown in Fig. 3. Among the four different VOC groups, the average concentrations of alkanes were highest (129 ± 173 ppbv), followed by alkenes (9.52 ± 14.5 ppbv), aromatic hydrocarbons (4.28 ± 8.24 ppbv) and acetylene (3.03 ± 5.55 ppbv). The top four alkanes were ethane (39.7 ± 57.3 ppbv), propane (22.6 ± 33.5 ppbv), *n*-butane (15.8 ± 21.4 ppbv) and *i*-butane (12.5 ± 17.5 ppbv). These four species totally accounted for 64.8% of the alkanes. Among the alkenes, 1-pentene, propylene and ethylene were the most abundant species with their average concentrations of 4.47 ± 6.72 ppbv, 1.88 ± 10.2 ppbv and 1.42 ± 1.69 ppbv,

respectively. They totally represented 71.8% of the alkenes. 96.7% of the aromatic hydrocarbons were composed by benzene, toluene and *m, p*-xylene, with corresponding average concentrations of 1.13 ± 1.62 ppbv, 1.06 ± 1.91 ppbv and 0.72 ± 1.94 ppbv, respectively. High concentrations of alkanes, ethane and propane in ambient were also reported in other oil and natural gas operation and industrial areas in the US (Pétron et al., 2012; Helmig et al., 2014; Warneke et al., 2014). For instance, the average concentrations of ethane and propane were 74 ± 79 ppbv and 33 ± 33 ppbv, respectively in Horse pool and Uintah Basin in the winter of 2012. Despite the highly enhanced VOC levels were due to the temperature inversion, the VOC levels in Uintah Basin were still higher than those in the regional background areas as the existence of oil and gas exploitation activities (Helmig et al., 2014). A distinct chemical signature of collected air samples from the Boulder Atmospheric Observatory in northeast Colorado was also found with enhanced concentrations of most alkanes (propane, *n*-butane, *i*-pentane and *n*-pentane) (Pétron et al., 2012).

The VOC concentrations, compositions and the top five species in this study and other areas around the world were compared and are shown in Fig. 4. The total VOC concentrations in this study (297 ± 372 ppbv) were 1–50 times higher than those in urban areas like Beijing (34.5 ppbv), Shanghai (32.4 ppbv), Guangzhou (43.6 ppbv), Seoul (122 ppbv), Mexico (117 ppbv) and 28 cities in the US (9.91 ppbv) as well as industrial areas, including Houston (31.2 ppbv), northeast Colorado (96.1 ppbv), Alberta oil sands (2.87 ppbv), Ulsan (91.7 ppbv), YRD (22.9 ppbv) and Nanjing (34.5 ppbv) (Fig. 4a). As shown in Fig. 3e and Fig. 4b, the alkanes were the most abundant groups (87.5% on average) during the whole sampling period, which was quite higher than other urban/industrial areas (45.3–67.2%, Fig. 4b). Similar relative high proportions of alkanes were found in Houston (77.1%) (Leuchner and Rappenglück, 2010), Alberta oil sands area (74.8%) (Simpson et al., 2010) and Northeast Colorado (97.4%) (Gilman et al., 2013), which were all related to oil and gas operations. In urban areas, the aromatics accounted for about 10.1–47.9% of the total VOCs, with toluene as one of the most abundant species. Toluene is mainly from solvent usage (Guo et al., 2004; Yuan et al., 2010) or vehicle exhaust emission (Wang et al., 2010) in cities. Another dominant compound in the urban air is propane (1.45–14.7 ppbv), which is the main component of liquid petroleum gas/natural gas (LPG/NG) (McCarthy et al., 2013). In industrial areas, alkanes and alkenes contribute most to the total VOCs (43.4–97.4% and 1.8–43.6%, respectively) with ethane, propane and ethylene as the top species usually (Fig. 4c). They may originate from the incomplete combustion or LPG/NG usage (Durana et al., 2006; Tang et al., 2007; Guo et al., 2011). To sum up, the concentrations of VOCs in this study were higher than many other regions and cities. The compositions and the top five species of VOCs exhibited typical characteristics of oil and gas exploring regions, such as Houston (Leuchner and Rappenglück, 2010), North Colorado (Gilman et al., 2013) and Alberta oil sands area (Simpson et al., 2010).

3.2 Contribution of VOCs to ozone formation potential (OFP)

The profiles of different VOCs categories with concentrations expressed in different scales are shown in Fig. 5. The top ten VOCs species to OFP obtained by Propy-Equiv and MIR method are listed in Table S1. Among the top ten compounds calculated by the two methods, six compounds were the same, but differing in their rank order. Considering the kinetic activity, 1-pentene ranked first by Propy-Equiv method. However, the *o*-xylene showed highest OFP based on the MIR method, which

may be related with the chemical mechanisms and the impacts of NO_x (Zou et al., 2015). Despite the two methods were different in mechanisms, the proportions of different VOC categories to the OFP were the same. From the non-weighted concentrations by volume and carbon atom, alkanes contributed 83 ± 9% and 82 ± 9%, respectively to the total VOCs concentrations, followed by alkenes (11 ± 6% and 9 ± 4%, respectively) and aromatics (5 ± 6% and 8 ± 7%, respectively). Although the proportions of alkenes and aromatics increased when compared to the values of non-weighted, the alkanes were still dominant, accounting for 45 ± 11% and 50 ± 14%, respectively. In summary, the alkanes had the highest concentrations (for both volume and carbon atom) and largest proportions to the OFP weighted by Propy-Equiv and MIR method. The results of this study were different from previous researches. For example, the alkanes with the highest concentrations (both for volume and carbon atom) contributed less to OFP, while alkenes and aromatics with less concentrations contributed most to the OFP in Guangzhou (73% and 83%, respectively) (Zou et al., 2015) and Tianjin (about 28–40% and 32–42%) (Liu et al., 2016) as well as a petrochemical industrialized city (48–49% and 37–49%, respectively) (Jia et al., 2016).

3.3 Temporal variations

Fig. 6 shows the temporal variations of ethane, ethylene, acetylene and benzene in different timescales. Though differences existed, the selected compounds broadly represent the respective alkanes, alkenes, alkynes and aromatics (Lyu et al., 2016). Significant differences were found between the meteorological parameters in different seasons ($p < 0.01$) and the highest concentrations of these species in winter were observed. The seasonal variation of VOCs is controlled by meteorological conditions, photochemical activities and source emissions. The highest values in winter was due to inhibited photochemical activities under suppressed dispersion conditions (averaged BLH as 121 ± 71.7 m, wind speed as 1.20 ± 0.76 m s^{-1}) and low temperature (–11.8 ± 5.00 ℃). For instance, all these species were negatively correlated with BLH, exhibiting higher VOC levels under lower BLH (Fig. 6). The wind speed and temperature were also found negatively correlated with VOC concentrations and ethylene showed the highest negative correlation coefficient with these parameters (Table S2). The less photochemical reactions can result in the high concentrations in winter, which was proved by negative correlation between VOCs and O_3 (Table S2). Additional sources (i.e., combustion) may be also present, in view of the obvious increase of acetylene (Fig. 6c) from summer (0.87 ± 1.00 ppbv) to winter (10.5 ± 8.51ppbv). Conversely, the high temperature, wind speed and BLH favor the dilution and dispersion of ambient VOCs and the photochemical depletion in summer.

The diurnal variations of VOCs and trace gases (NO_2 and O_3) related to photochemical reaction are shown in Fig. 7. The VOCs had a reverse trend with O_3 ($r = –0.82$, $p < 0.01$). The lower BLH and less photochemical activities resulted in peak values for VOCs and low O_3 concentrations before sunrise (6:00 local time). After sunrise, with the initiation of photochemical oxidation and the increasing of BLH, the concentrations of VOCs decreased while the O_3 increased rapidly. The minimum of VOCs and occurred at about 12:00–14:00 LT was resulted from both dispersion or dilution conditions and photochemical reactions (with highest O_3 concentrations at l4:00 LT) in the afternoon. The diurnal variation of NO_2 was controlled by BLH, O_3 and photochemical reactions (i.e., OH radical) and showed a double peak. The similar diurnal patterns of different atmospheric

lifetime compounds including ethane, ethylene, acetylene and benzene (the most abundant contributors to its categories) were also found (Fig. S4). To better understand the effects of BLH and photochemical reactions on VOCs, the diurnal variations of VOCs, BLH and O_3 in winter and summer were analyzed (Fig. 7b, c). VOC concentrations in winter (213 ±97.7 ppbv) were significantly higher than those in summer (130 ±100 ppbv). However, the VOCs in summer and winter decreased by 8.3 times and 2.3 times, respectively, from maximum to minimum. This was due to the BLH increased by 8.2 times in summer while the BLH in winter only increased by 2.3 times. The effects of photochemical reactions on VOCs in two seasons were comparable, which was explained by similar O_3 increment in winter (0.78 times up) and summer (0.71 times up). Therefore, we can conclude that the role of BLH variation was more important than the photochemical reaction for the diurnal variation of VOCs.

3.4 Ambient ratios: sources and photochemical removal

Ambient ratios for VOC species holding similar reaction rates with OH radicals can reflect the source features, as these compounds are equally affected by the photochemical processing and the new emission inputs (Russo et al., 2010; Baltrėnas et al., 2011; Miller et al., 2012). For example, *n*-butane and *i*-butane have similar reaction rates with the OH radicals, with the differences less than < 10% and the ratios of these pair species indicated different sources. The butanes are associated with NG, LPG, vehicle emission and biomass burning and the ratios of *i*-butane / *n*-butane varied according to sources (i.e., 0.2–0.3 for vehicle, 0.46 for LPG and 0.6–1.0 for NG) (Buzcu and Fraser, 2006; Russo et al., 2010). In this study, the slope of *i*-butane / *n*-butane (0.80–0.82, Fig. 8a) was within the range of reported emissions from natural gas. Additionally, *i*-pentane and *n*-pentane have similar physical and chemical characteristics (i.e., boiling point and reaction rate coefficients with hydroxyl radical), which result in less susceptible of the *i*-pentane / *n*-pentane ratio in source identification (Gilman et al., 2013). The pentanes are always form the NG emission, vehicle emission, liquid gasoline and fuel evaporation with the *i*-pentane/*n*-pentane ratios ranged between 0.82–0.89 (Gilman et al., 2010, 2013), ~2.2–3.8 (Conner et al., 1995; McGaughey et al., 2004), 1.5–3.0, and 1.8–4.6 (Watson et al., 2001), respectively. As shown in Fig. 8b, the slopes of *i*-pentane / *n*-pentane were 1.03–1.24 in this study, suggested that the pentanes were more likely from the mixed sources of NG and fuel evaporation. This assumption was proved by the high loadings of pentanes in NG and fuel evaporation source compositions in section 3.5.5

Information on the photochemical removal process can be obtained by comparing the ambient ratios of aromatics due to their differences in atmospheric lifetimes. For example, the atmospheric lifetimes of benzene (9.4 days), toluene (1.9 days) and ethylbenzene (1.6 days) are relative longer than *m*-xylene (11.8 h) and *p*-xylene (19.4 h) (Monod et al., 2001). The commonly used ratios are benzene / toluene, *m, p*-xylene / ethylbenzene, benzene / ethylbenzene and toluene / ethylbenzene. The diurnal variation of these compounds and ratios are shown in Fig. 9. A continuous decreasing of these compounds and ratios were observed from 08:00 to 14:00 LT, indicating the increased photochemical removal processes due to the increase in reactive radicals (i.e., hydroxyl radical). The diurnal patterns of benzene / ethylbenzene and toluene / ethylbenzene in this study (Fig. 9b, d) were opposite to those observed in Dallas, which was mainly influenced by vehicle emission (Qin et al., 2007). After 14:00 LT, the increasing of the ratios and aromatic concentrations were due to the weakening of photochemical activities. The

unusual high concentrations of ethylbenzene and *m, p*-xylene were observed at about 02:00 LT (Fig. 9c), which might be related to new emissions. This assumption was verified by a small peak occurred at 02:00 LT in the diurnal profile of oil refinery source (see section 3.5.1). After 12 h dispersion, dilution and photochemical reaction, the concentrations of these two compounds reached its minimum values at about 14:00 LT.

Generally speaking, when the reaction with OH radicals was the only factor controlling the seasonal ratio of longer atmospheric lifetime to shorter lifetime compounds (i.e., benzene / toluene, *m, p*-xylene / ethylbenzene), an increasing in ratio value from winter to summer would be expected (Russo et al., 2010). However, the seasonal variation of BTEX ratios in this study was opposite to the general behavior. For example, the benzene / toluene ratio decreased from winter-spring (0.63–0.69) to summer-fall (0.52–0.57) (Fig. 8c) and ethylbenzene / *m, p*-xylene ratio also decreased from autumn-winter (0.47–0.69) to spring-summer (0.19–0.37) (Fig. 8d). Same results were also observed both in industry areas (Miller et al., 2012) and urban areas (Ho et al., 2004; Hoque et al., 2008; Russo et al., 2010). The results obtained in this study indicated that there were other factors affecting the seasonal variation such as source emissions. The BTEX mainly originate from vehicle exhaust (Wang et al., 2010), solvent usage (Guo et al., 2004; Yuan et al., 2010) and petrochemical industry (Na and Kim, 2001; Hsieh et al., 2006; Baltrėnas et al., 2011). The ratio of *m, p*-xylenes / ethylbenzene here (2.2 ± 1.2) was within the ranges reported at a petrochemical area in southern Taiwan (1.5–2.6) (Hsieh et al., 2006) and the vicinity of a crude oil refinery at the Baltic region (3.0–4.0) (Baltrėnas et al., 2011). Therefore, the BTEX in this area was mainly from the oil refinery emission. The unexpected low ratios of benzene / toluene and *m, p*-xylene / ethylbenzene in summer was due to the strong oil refinery emission strength and this finding was verified by the seasonal source contribution results in section 3.5.1.

3.5 Source apportionment: temporal variation and contribution to OFP

Five sources including oil refining process, NG, combustion source, asphalt and fuel evaporation were identified by the PMF analysis and their source profiles and daily contributions are shown in Fig. 10. The monthly, seasonal and annual contributions were calculated and shown in Fig. 11. The relationships between daily source contributions and meteorological parameters and trace gases were analyzed by scatter plots (Fig. 12). The source apportionment of this high-resolution dataset provided a unique opportunity to discuss the diurnal variation of different sources as shown in Fig. 13.

3.5.1 Oil refining

The emissions from the refining process are complex due to the diversities of VOC species, which depend on the production processes (Vega et al., 2011; Mo et al., 2015). The crude oil is composed of $\geq C_5$ alkanes, cycloalkanes, aromatics and asphaltics (Simpson et al., 2010) and they are supplied as the raw materials for various oil refining processes. High fractions of C_5–C_9 alkanes including hexane (32 ± 6.2%), cyclohexane (40 ± 7.9%), methylcyclohexane (47 ± 6.9%), *n*-octane (56 ± 4.2%), *n*-nonane (58 ± 2.9%) and aromatics (i.e., 22 ± 3.0% for benzene, 39 ± 5.4% for toluene and 45 ± 7.3% for xylenes) presented in this factor (Fig. 10a), which was similar to the chemical compositions measured from the oil refinery (Liu et al., 2008; Dumanoglu et al., 2014). The calculated daily source contributions from PMF model were well correlated with the high

loading species in its source profiles. For example, the methylcyclohexane showed significant correlation with this source contribution (Fig. S5a), suggesting that the tracers of oil refinery were well produced by the PMF model. The main products from oil refinery are gasoline, diesel, lube and kerosene in this area, consistent with the factor derived here.

The annual contribution of oil refining source was relatively stable throughout the year (3.8 ± 0.50%). The highest relative contribution was found in summer (5.3%) and the lowest in winter (2.4%) (Figure 11b). The Pearson analysis between the daily source contributions and wind speed disclosed a middle statistically negative correlation ($r = –0.12$, $p < 0.05$). However, no statistically correlations between the daily source contribution and other meteorological parameters were found (Table S3), even for the BLH. On the contrary, significant positive correlations between this source and trace gases (NO_2 and CO) were found, with *r* being 0.33 and 0.21, respectively (Fig. 12a). These trace gases are associated with oil refinery emission (Cetin et al., 2003). Therefore, the daily variation of oil refinery source in this study was more controlled by oil refining emission strength and less influenced by meteorological conditions.

The diurnal pattern of this source contribution was well correlated to the methylcyclohexane ($r = 0.76$, $p < 0.01$) and characterized by a double wave profile with the first peak at 02:00 LT and second peak at 06:00 LT (Fig. 13a). A small peak occurred at 02:00 was due to the increasing of ethylbenzene and *m, p*-xylene (Fig. 9) and the second peak occurred at 06:00 LT resulted from the low BLH. After sunrise, the contribution continuously decreased owing to the increasing of BLH and photochemical reactions and the minimum value occurred at 14:00 LT.

3.5.2 Natural gas

Ethane and propane are the most abundant non-methane hydrocarbon compounds in natural gas (Xiao et al., 2008; McCarthy et al., 2013). The ratio of *i*-butane / *n*-butane indicated the butanes were from the natural gas (section 3.4). By the PMF analysis, a NG source was identified through the high weights on ethane (81 ± 2.4%), propane (85 ± 5.3%), *n*-butane (62 ± 7.5%), and *i*-butane (54 ± 6.4%). As an important oil and gas resources base in China, the export amount of natural gas from this region was 4.4×10^9 m^3 and the loss rate was 1.4% in 2014 (Chen, 2015). The leakage from the exploiting, storing, transporting and processing cannot be ignored, suggesting that it was reasonable to attribute this factor to natural gas source.

The annual contribution of the NG leakage source was 53 ppbv, accounting for 62.6 ± 3.04% of the total VOCs averagely. The highest contribution presented in spring (65.2%), followed by summer (63.6%), autumn (63.0%) and winter (60.4 %). The daily variation of this source was influenced by meteorological parameters such as the BLH ($r = –0.42$, $p < 0.01$) (Table S3). The significant positive correlations between NO_2 and CO and the source contribution were also found with Pearson coefficients being 0.45 and 0.44, respectively (Fig. 12b), indicating that the daily variation of NG source was influenced by meteorological conditions and photochemical activities. The diurnal variation of the NG leakage was significantly correlated ($p < 0.01$) with the diurnal pattern of propane *n*-butanes and *i*-butane with Pearson coefficients as 0.94, 0.87 and 091, respectively (Fig. 13b), which was also reported by Baudic et al. (2016). The diurnal behaviors of this source were characterized by a nighttime high and mid-afternoon low pattern, which can be interpreted as the diurnal evolution of BLH (Bon et al., 2011; Baudic et al., 2016).

3.5.3 Combustion source

This source was dominantly weighted by ethylene (95 ± 3.5 %), acetylene (97 ± 2.6%) and moderately influenced by BTEX. These species are key markers of combustion (Fujita, 2001; Watson et al., 2001; Jobson, 2004) or from petrochemical source (Brocco et al., 1997; Song et al., 2007). However, the independent combustion tracers such as CO, NO_2 and $PM_{2.5}$ were well correlated to this source contribution with Pearson correlation coefficients of 0.59, 0.49 and 0.77, respectively (Fig. 12c and Table S3). Therefore, this factor was attributed to combustion source. This source exhibited obvious seasonal differences with highest contribution in winter (14.9%) and lowest contribution in summer (6.9%). The seasonal difference was due to the temperature change and was proved by the significant negative correlation with ambient temperature (r = –0.57, p < 0.01). The diurnal variation of combustion source was in accordance with the diurnal pattern of ethylene and CO with Pearson correlation coefficients as 0.71 (p < 0.05) and 0.84 (p < 0.01), respectively. It was characterized by a double peak profile with an initial increasing from 03:00 to 08:00 LT and a second increasing at nighttime (20:00–24:00 LT) (Fig. 13c). The increase in the morning was related to the low BLH. Different from other researches, no increasing trend of this source was found during 07:00–10:00 LT here, while combustion source was reported to be increasing at the rush-hour period (Gaimoz et al., 2011; Baudic et al., 2016). On the contrary, the decreasing trends were found for independent combustion tracers (CO and NO_2) during this period (Fig. 7a). In another rush-hour of 18:00–20:00 LT, the enhancement of combustion source contributions and CO from 16:00 LT (Fig. 11c) may be related with the reduction of BLH. The reduction of NO_2 from 18:00 LT (Fig. 7a) were also observed, which indicated that the diurnal variation of combustion source was less affected by vehicle exhaust in the present study.

3.5.4 Asphalt

Asphalt released predominantly C_8–C_{11} alkanes including *n*-octane, *n*-nonane, *n*-decane and *n*-undecane, totally contributed to over 50% of VOC emissions from asphalt application (Brown et al., 2007; Liu et al., 2008; Deygout, 2011), with *n*-undecane along accounting for 17% (Liu et al., 2008). Benzene, toluene and xylenes are also enriched for asphalt VOC emissions (Chong et al., 2014). High loadings of C_9–C_{12} VOCs including *n*-nonane, *n*-decane, *n*-undecane and *n*-dodecane were found in this factor, averaged as 46%, 64%, 72% and 85%, respectively. The annual processing capacity of heavy oil in this area was 9.0 × 10^6 t and the fugitive emission was inevitable. Therefore, this factor was attributed to asphalt.

The annual contribution of asphalt was the lowest among the five sources and only contributed 1.3 ± 0.69% to the total VOCs. The daily contributions of this source and temperature had a statistically reliable positive correlation (r = 0.19, p < 0.01). The seasonal variation of this source was influenced by temperature with the highest contributions occurred in autumn (2.1%) and the lowest in winter (0.5%). However, the influence of BLH on the contribution of asphalt was not significant (r = 0.04, p > 0.05). The correlations between this source and O_3 was found insignificant (r = –0.001, p > 0.05). However, significant positive correlation between asphalt and oil refinery source was observed (r = 0.47, p < 0.01) (Fig. 12d), indicating they shared the same origin, which should be oil refining processes in current study.

The diurnal variation of asphalt was different from other sources and well followed the diurnal patterns of decane (r = 0.76, p < 0.01) and undecane (r = 0.86, p < 0.01). It continuously decreased from 2:00 to 6:00 LT, slowly increased from 6:00 to 10:00 LT and subsequently decreased (Fig. 13d). A minimum source contribution occurred when the BLH was low in the morning, which was contrary to the other sources. In addition, no significant correlation between this source and O_3 (r = –0.02, p > 0.05) was found. Therefore, the temporal variation of asphalt was less controlled by BLH and photochemical reaction, but was more influenced by the emission strength.

3.5.5 Fuel evaporation

The gasoline evaporation profile holds high proportions of *i*-pentane, *trans*-2-pentene, *cis*-2-pentene, benzene and toluene (Liu et al., 2008; Zhang et al., 2013b). The *i*-pentane is a key tracer of gasoline evaporation due to its high abundance (Gentner et al., 2009; Zhang et al., 2013b). The ratio of *i*-pentane to *n*-pentane is useful to identify the potential sources including NG (0.82–0.89), liquid gasoline (1.5–3.0), fuel evaporation (1.8–4.6) and vehicle emission (2.2–3.8) (Harley et al., 2001; McGaughey et al., 2004; Russo et al., 2010; Gilman et al., 2013). As discussed above, the ratio of *i*-pentane to *n*-pentane indicated a mixed source in this region. From the PMF modeling results, high loadings on *i*-pentane and *n*-pentane were present in the factor profile, which accounted for 85 ±5.3% and 71 ±6.4% of total species, respectively. Additionally, this factor was influenced by hexane (60 ±5.6%), cyclohexane (45 ±10%), methylcyclohexane (52 ±7.5%), benzene (23 ±3.0%) and toluene (19 ±4.0%), which were related to diesel fuel evaporation (Liu et al., 2008). As shown in Fig. S1, the products of oil refinery included the gasoline and diesel, with the annual production of 9.5×10^5 t and 1.9×10^5 t, respectively (Chen, 2015). Therefore, this factor represented the fuel evaporation.

The fuel evaporation is controlled by temperature, leading to higher contributions in summer. The highest contribution was found in summer (22.9%) in this study. Same results were also observed previously (Baudic et al., 2016; Liu et al., 2016). A significant correlation between the contributions of NG and fuel evaporation was observed (r = 0.65, p < 0.01), indicating this two sources were influenced by the similar factors. The diurnal distribution pattern of fuel evaporation source was different from former studies in urban area (with an increasing trend from 7:00 to 10:00 LT due to the morning rush traffic) (Baudic et al., 2016). On the contrary, the source contribution followed the diurnal variations of fuel evaporation tracers such as *i*-pentane, *n*-pentane and methycyclohexane, with Pearson correlation coefficients being 0.86 (p < 0.01), 0.87 (p < 0.01) and 0.67 (p <0.05), respectively.

3.5.6 Contribution to OFP

The contributions of five identified VOC sources to OFP were also evaluated using F matrix and MIR method. The fuel evaporation showed the highest contribution (41.9%, 41.6 ppbv), followed by NG (29.6%, 29.4 ppbv), combustion (14.2%, 14.1 ppbv), oil refinery (11.3%, 11.2 ppbv) and asphalt (3.0%, 3.0 ppbv). Therefore, more attention should be paid to the fuel evaporation due to its high ozone formation potential. It should be noted that the source contributions to OFP were calculated by 20 selected VOC species in PMF modeling and the actual contributions to OFP were higher than the results.

3.6 Source contributions compared with previous studies

The source apportionment results showed that the dominant source in this study was the natural gas source, contributing 62.6 ± 3.04% to the total VOCs on the annual average, followed by fuel evaporation (21.5 ± 2.99%), combustion source (10.9 ± 1.57%), oil refinery (3.80 ± 0.50%) and asphalt emission (1.30 ± 0.69%). Each identified PMF factor exhibited obvious temporal variations due to the emission strength, photochemical reaction and meteorological conditions. The source apportionment results in this study were compared with formers based on long- term monitoring (Table 2).

The contributors to VOCs in urban areas were complex with at least five different sources including fuel evaporation, LPG/NG, industrial emission, vehicle emission and solvent usage (Table 2). While the number of VOC sources apportioned in industrial areas was less compared to the cities. For example, only three sources including vehicle emission (58.3%), solvent usage (22.2%) and industrial activities (19.5%) were apportioned by principle component analysis-multiple linear regression (PCA-MLR) in Lanzhou, a petrochemical industrialized city in northwest China (Jia et al., 2016). Same result was also found in Houston that only fuel evaporation, industrial emission and vehicle emission were identified (Leuchner and Rappenglück, 2010). In these studies, the vehicle emission was an important source both in urban and industrial areas and contributed about 11–58.3% to the total VOCs (Table 2). However, the vehicle emission source was not identified in this study due to several reasons. Firstly, despite there was similarity between the source profile of combustion or fuel evaporation in this study and the vehicle emission (i.e., high loadings on acetylene, ethylene, BTEX, butanes and pentanes), the temporal variations of these species did not show a distinct increasing during the traffic rush-hour. In fact, the identified combustion source in this study represented the characteristics of coal burning and torch burning in oil refinery (to eliminate the hazardous gases). Secondly, differences existed in sampling location and vehicle amounts. In previous urban studies, the sampling location was in megacities with huge vehicle flows. For example, in the research of Wuhan (Lyu et al., 2016), the sampling site located in the city center and the car population was 2.2×10^6 by the end of 2015. While the sampling location here was about 11 km away from the urban areas and the car ownership was only 1.1×10^5. Therefore, the factor with higher loadings of these species was not likely to be contributed by vehicle emission in this study.

LPG and NG sources are usually apportioned both in urban and industrial areas. These sources contribute 10%–32% to the total VOCs and are mainly from household or industrial fugitive emission. However, in this study, the NG source was mainly from the NG exploitation and NG chemical industry due to its abundance in this area and accounted for 62.6 ± 3.04% on average to VOCs, which was higher than many other areas as summarized in Table 2.

Solvent usage also accounts for a large proportion of total VOCs in urban areas (4.7–36.4%). In this study, a similar source related to asphalt was identified with heavy weights on C_9C_{12} compounds. The solvent usage in urban areas is usually from painting or coating. However, the asphalt in this study originated from oil refinery (Fig. S1) and fugitive emission from a black oil hill located at the northwest of the sampling site. Due to its high boiling point, the seasonal contribution of asphalt was distinct with highest contribution in July (7.2%) and lowest contribution in January (1.4%). Despite the source contribution of asphalt was low, it was unique in this study.

3.7 Geographic origins of VOCs sources: local vs. regional contributions

The possible geographic origins of five identified VOC sources were explored by CPF, PSCF and CWT as shown in Fig. 14, Fig. 15 and Fig. 16, respectively. These methods aimed at providing insights on the potential geographic origins of VOC sources but did not claim to be precise at the cell level or pixel level.

The highest CPF value of oil refinery was found in east direction of the sampling site (Fig. 14a), which indicated the potential location of this source. However, the oil refineries are mainly located in the southwest of the sampling site (Fig. 1c) and high CPF value (0.95) was also found in southwest direction. Therefore, the CPF results was able to reflect the location the oil refineries. Similarly, high probabilities and concentrations of oil refinery were also found from the southeast to southwest area of the sampling site according to the PSCF (Fig. 15a) and CWT plots (Fig. 16a). As shown in Fig. 1a and 1b, the sampling site is located in the west of the Junggar Basin, which is the second largest oil and gas basin in China. Indeed, high values of CPF, PSCF and CWT were found in the east direction (Fig. 14b, Fig. 15b and Fig. 16b), which indicated that the potential geographic origins of NG. Given the fact that the NG source was composed by long atmospheric lifetime species (i.e., ethane, propane and butanes), the high probabilities and concentrations of this factor were likely resulted from aged air masses from each direction. The combustion source showed high potentials from ESE to SE direction according to the CPF, PSCF and CWT plots. There were no high values in the northwest direction of sampling site, where the urban area locates. This also indicated that the combustion from vehicle emission was insignificant in this study. For asphalt source, highest CPF value was found in the east direction while the PSCF and CWT plots showed high values in the northeast direction. As discussed above, the asphalt source in this study were from the natural source (black oil hill in the northwest of sampling site) and oil refinery (southwest direction). The CPF, PSCF and CWT results indicated that these methods failed to locate the natural source of asphalt. The potential geographic origins of fuel evaporation were widespread from ESE to W direction, which was similar to the oil refinery source.

Diversities of geographic origins were also found in different seasons (Fig. S6–S13). The potential source areas of the five sources spread from northeast to southwest in autumn. In winter, both PSCF and CWT methods indicated that the VOC sources were probably from the southeast and southwest. In spring, VOCs were mainly from long-range transport from west. However, high probabilities and contributions existed around the sampling site. In summer, high potential and contribution were from the west to the southeast direction. Overall, the five sources exhibited different local source areas proved by the CPF plots on the annual scale. Similar regional distributions of these sources were found on the seasonal scale. To quantify the contributions of local emission and long-range transport to the sampling site, raster analysis based on CWT was used and the results are summarized in Table 3. Annually, except for the combustion source, the identified VOC sources were mainly from the local emission, with contributions of 53.6% for oil refining, 54.5% for NG, 50.5% for asphalt and 50.6% for fuel evaporation, respectively. The seasonal patterns were same with the annual pattern, exhibiting higher contributions from local areas and the differences only existed in the proportions. Highest local contributions of oil refining (69.4%) and combustion (69.2%) were

observed in summer, while the local sources contributed most to the NG (74.6%), asphalt (65.4%) and fuel evaporation (68.3%) in autumn.

4 Summary

Based on one-year continuously online monitoring of VOCs in an oil and gas field, and on the use of PMF receptor model, back trajectory, PSCF and CWT dispersion models, this study compared the VOC levels and compositions with other studies, identify the VOCs source and explored the potential geographic origins of five identified VOC sources. The main findings are summarized as follows:

1. The total VOC concentrations in this study were not only higher than those in urban areas, but also higher than those measured in petrochemical areas. Alkanes contributed most to the total VOCs (accounting for 87.5% and 128 ±82.4 ppbv on average), followed by alkenes (6.81% and 9.1 ±5.6 ppbv), aromatic hydrocarbons (3.37% and 4.8 ±6.5 ppbv) and acetylene (2.32% and 3.1 ±5.1ppbv).

2. Five sources with local characteristics were identified. The NG contributed most to the VOCs (62.6 ±3.04 %), followed by fuel evaporation (21.5 ±2.99%), combustion source (10.9 ±1.57%), oil refining (3.80 ±0.50%) and asphalt (1.30 ±0.69%). The NG and fuel evaporation source contributions showed positive correlation with each other and shared the same diurnal variation pattern, exhibiting a single peak profile. The diurnal variation of oil refining and combustion source exhibited similar double wave with peaks occurred 06:00–08:00 LT. Different from other sources, the diurnal profile of asphalt exhibited a decreasing trend from nighttime to its minimum before sunrise (06:00 LT).

3. The geographic origins of five VOC sources were the same during the whole period. The differences existed in the seasonal variations of them. For instance, VOCs were mainly from northeast and southwest in autumn, while they originated from southeast and southwest in winter. The raster analysis indicated that the VOCs in this study were mainly from local emission with contributions ranging from 48.4% to 74.6% in different seasons.

In summary, this study found that the VOC concentrations, compositions, ozone formation potential and sources were different from those in urban and industrial areas and similar to oil and gas rich areas. It will be helpful for the VOCs control in these type of regions around the world.

Data availability

The VOC concentrations during the whole sampling period are available on request from Shaofei Kong (kongshaofei@cug.edu.cn).

Acknowledgements. This study was financially supported by the Key Program of Ministry of Science and Technology of the People's Republic of China (2016YFA0602002; 2017YFC0212602). The research was also supported by the Fundamental Research Funds for the Central Universities, China University of Geosciences, Wuhan. The authors are grateful to the local Environmental Monitoring Center Station for their sampling work. We thank Qingyue Open Environmental Data Center (https://data.epmap.org) for providing air quality data.

Reference

An, J., Zhu, B., Wang, H., Li, Y., Lin, X. and Yang, H.: Characteristics and source apportionment of VOCs measured in an industrial area of Nanjing, Yangtze River Delta, China, Atmos. Environ., 97, 206–214, doi:10.1016/j.atmosenv.2014.08.021, 2014.

Atkinson, R. and Arey, J.: Atmospheric Degradation of Volatile Organic Compounds, Chem. Rev., 103(12), 4605–4638, doi:10.1021/cr0206420, 2003.

Baker, A. K., Beyersdorf, A. J., Doezema, L. A., Katzenstein, A., Meinardi, S., Simpson, I. J., Blake, D. R. and Sherwood Rowland, F.: Measurements of nonmethane hydrocarbons in 28 United States cities, Atmos. Environ., 42(1), 170–182, doi:10.1016/j.atmosenv.2007.09.007, 2008.

Baltrėnas, P., Baltrėnaitė, E., Šerevičienė, V. and Pereira, P.: Atmospheric BTEX concentrations in the vicinity of the crude oil refinery of the Baltic region, Environ. Monit. Assess., 182(1–4), 115–127, doi:10.1007/s10661-010-1862-0, 2011.

Bari, A., Dutkiewicz, V. A., Judd, C. D., Wilson, L. R., Luttinger, D. and Husain, L.: Regional sources of particulate sulfate, SO_2, $PM_{2.5}$, HCl, and HNO_3, in New York, NY, Atmos. Environ., 37(20), 2837–2844, doi:10.1016/S1352-2310(03)00200-0, 2003.

Bari, M. A., Kindzierski, W. B. and Spink, D.: Twelve-year trends in ambient concentrations of volatile organic compounds in a community of the Alberta Oil Sands Region, Canada, Environ. Int., 91, 40–50, doi:10.1016/j.envint.2016.02.015, 2016.

Baudic, A., Gros, V., Sauvage, S., Locoge, N., Sanchez, O., Sarda-Estève, R., Kalogridis, C., Petit, J.-E., Bonnaire, N., Baisnée, D., Favez, O., Albinet, A., Sciare, J. and Bonsang, B.: Seasonal variability and source apportionment of volatile organic compounds (VOCs) in the Paris megacity (France), Atmos. Chem. Phys., 16(18), 11961–11989, doi:10.5194/acp-16-11961-2016, 2016.

Bon, D. M., Ulbrich, I. M., de Gouw, J. A., Warneke, C., Kuster, W. C., Alexander, M. L., Baker, A., Beyersdorf, A. J., Blake, D., Fall, R., Jimenez, J. L., Herndon, S. C., Huey, L. G., Knighton, W. B., Ortega, J., Springston, S. and Vargas, O.: Measurements of volatile organic compounds at a suburban ground site (T1) in Mexico City during the MILAGRO 2006 campaign: measurement comparison, emission ratios, and source attribution, Atmos. Chem. Phys., 11(6), 2399–2421, doi:10.5194/acp-11-2399-2011, 2011.

Bressi, M., Sciare, J., Ghersi, V., Mihalopoulos, N., Petit, J.-E., Nicolas, J. B., Moukhtar, S., Rosso, A., Féron, A., Bonnaire, N., Poulakis, E. and Theodosi, C.: Sources and geographical origins of fine aerosols in Paris (France), Atmos. Chem. Phys., 14(16), 8813–8839, doi:10.5194/acp-14-8813-2014, 2014.

Brocco, D., Fratarcangeli, R., Lepore, L., Petricca, M. and Ventrone, I.: Determination of aromatic hydrocarbons in urban air of Rome, Atmos. Environ., 31(4), 557–566, doi:10.1016/S1352-2310(96)00226-9, 1997.

Brown, S. G., Frankel, A. and Hafner, H. R.: Source apportionment of VOCs in the Los Angeles area using positive matrix factorization, Atmos. Environ., 41(2), 227–237, doi:10.1016/j.atmosenv.2006.08.021, 2007.

Buzcu, B. and Fraser, M. P.: Source identification and apportionment of volatile organic compounds in Houston, TX, Atmos. Environ., 40(13), 2385–2400, doi:10.1016/j.atmosenv.2005.12.020, 2006.

Buzcu-Guven, B. and Fraser, M. P.: Comparison of VOC emissions inventory data with source apportionment results for Houston, TX, Atmos. Environ., 42(20), 5032–5043, doi:10.1016/j.atmosenv.2008.02.025, 2008.

Cai, C., Geng, F., Tie, X., Yu, Q. and An, J.: Characteristics and source apportionment of VOCs measured in Shanghai, China, Atmos. Environ., 44(38), 5005–5014, doi:10.1016/j.atmosenv.2010.07.059, 2010.

Callén, M. S., Iturmendi, A. and López, J. M.: Source apportionment of atmospheric $PM_{2.5}$-bound polycyclic aromatic hydrocarbons by a PMF receptor model. Assessment of potential risk for human health, Environ. Pollut., 195, 167–177, doi:10.1016/j.envpol.2014.08.025, 2014.

Carter, W. P. L.: Development of Ozone Reactivity Scales for Volatile Organic Compounds, Air Waste, 44(7), 881–899, doi:10.1080/1073161X.1994.10467290, 1994.

Cetin, E., Odabasi, M. and Seyfioglu, R.: Ambient volatile organic compound (VOC) concentrations around a petrochemical complex and a petroleum refinery, Sci. Total Environ., 312(1), 103–112, doi:10.1016/S0048-9697(03)00197-9, 2003.

Chan, Y., Hawas, O., Hawker, D., Vowles, P., Cohen, D. D., Stelcer, E., Simpson, R., Golding, G. and Christensen, E.: Using multiple type composition data and wind data in PMF analysis to apportion and locate sources of air pollutants, Atmos. Environ., 45(2), 439–449, doi:10.1016/j.atmosenv.2010.09.060, 2011.

Chen, Y.-C., Chiang, H.-C., Hsu, C.-Y., Yang, T.-T., Lin, T.-Y., Chen, M.-J., Chen, N.-T. and Wu, Y.-S.: Ambient $PM_{2.5}$-bound polycyclic aromatic hydrocarbons (PAHs) in Changhua County, central Taiwan: Seasonal variation, source apportionment and cancer risk assessment, Environ. Pollut., 218, 372–382, doi:10.1016/j.envpol.2016.07.016, 2016.

Cheng, H., Guo, H., Saunders, S. M., Lam, S. H. M., Jiang, F., Wang, X., Simpson, I. J., Blake, D. R., Louie, P. K. K. and Wang, T. J.: Assessing photochemical ozone formation in the Pearl River Delta with a photochemical trajectory model, Atmos. Environ., 44(34), 4199–4208, doi:10.1016/j.atmosenv.2010.07.019, 2010.

Cheng, I., Zhang, L., Blanchard, P., Dalziel, J. and Tordon, R.: Concentration-weighted trajectory approach to identifying potential sources of speciated atmospheric mercury at an urban coastal site in Nova Scotia, Canada, Atmos. Chem. Phys., 13(12), 6031–6048, doi:10.5194/acp-13-6031-2013, 2013.

Chong, D., Wang, Y., Guo, H. and Lu, Y.: Volatile Organic Compounds Generated in Asphalt Pavement Construction and Their Health Effects on Workers, J. Constr. Eng. Manag., 140(2), 4013051, doi:10.1061/(ASCE)CO.1943-7862.0000801, 2014.

Colman Lerner, J. E., Sanchez, E. Y., Sambeth, J. E. and Porta, A. A.: Characterization and health risk assessment of VOCs in occupational environments in Buenos Aires, Argentina, Atmos. Environ., 55, 440–447, doi:10.1016/j.atmosenv.2012.03.041, 2012.

Conner, T. L., Lonneman, W. A. and Seila, R. L.: Transportation-Related Volatile Hydrocarbon Source Profiles Measured in Atlanta, J. Air Waste Manag. Assoc., 45(5), 383–394, doi:10.1080/10473289.1995.10467370, 1995.

Deygout, F.: Volatile emissions from hot bitumen storage tanks, Environ. Prog. Sustain. Energy, 30(1), 102–112, doi:10.1002/ep.10444, 2011.

Dumanoglu, Y., Kara, M., Altiok, H., Odabasi, M., Elbir, T. and Bayram, A.: Spatial and seasonal variation and source apportionment of volatile organic compounds (VOCs) in a heavily industrialized region, Atmos. Environ., 98, 168–178, doi:10.1016/j.atmosenv.2014.08.048, 2014.

Dunlea, E. J., Herndon, S. C., Nelson, D. D., Volkamer, R. M., San Martini, F., Sheehy, P. M., Zahniser, M. S., Shorter, J. H., Wormhoudt, J. C., Lamb, B. K., Allwine, E. J., Gaffney, J. S., Marley, N. A., Grutter, M., Marquez, C., Blanco, S., Cardenas, B., Retama, A., Ramos Villegas, C. R., Kolb, C. E., Molina, L. T. and Molina, M. J.: Evaluation of nitrogen dioxide chemiluminescence monitors in a polluted urban environment, Atmos. Chem. Phys., 7(10), 2691–2704, doi:10.5194/acp-7-2691-2007, 2007.

Durana, N., Navazo, M., Gómez, M. C., Alonso, L., García, J. A., Ilardia, J. L., Gangoiti, G. and Iza, J.: Long term hourly measurement of 62 non-methane hydrocarbons in an urban area: Main results and contribution of non-traffic sources, Atmos. Environ., 40(16), 2860–2872, doi:10.1016/j.atmosenv.2006.01.005, 2006.

Fujita, E. M.: Hydrocarbon source apportionment for the 1996 Paso del Norte Ozone Study, Sci. Total Environ., 276(1–3), 171–184, doi:10.1016/S0048-9697(01)00778-1, 2001.

Gaimoz, C., Sauvage, S., Gros, V., Herrmann, F., Williams, J., Locoge, N., Perrussel, O., Bonsang, B., d'Argouges, O., Sarda-Estève, R. and Sciare, J.: Volatile organic compounds sources in Paris in spring 2007. Part II: source apportionment using positive matrix factorisation, Environ. Chem., 8(1), 91, doi:10.1071/EN10067, 2011.

Garzón, J. P., Huertas, J. I., Magaña, M., Huertas, M. E., Cárdenas, B., Watanabe, T., Maeda, T., Wakamatsu, S. and Blanco, S.: Volatile organic compounds in the atmosphere of Mexico City, Atmos. Environ., 119, 415–429, doi:10.1016/j.atmosenv.2015.08.014, 2015.

Geng, F., Tie, X., Xu, J., Zhou, G., Peng, L., Gao, W., Tang, X. and Zhao, C.: Characterizations of ozone, NO_x, and VOCs measured in Shanghai, China, Atmos. Environ., 42(29), 6873–6883, doi:10.1016/j.atmosenv.2008.05.045, 2008.

Gentner, D. R., Harley, R. A., Miller, A. M. and Goldstein, A. H.: Diurnal and Seasonal Variability of Gasoline-Related Volatile Organic Compound Emissions in Riverside, California, Environ. Sci. Technol., 43(12), 4247–4252, doi:10.1021/es9006228, 2009.

Gilman, J. B., Burkhart, J. F., Lerner, B. M., Williams, E. J., Kuster, W. C., Goldan, P. D., Murphy, P. C., Warneke, C., Fowler, C., Montzka, S. A., Miller, B. R., Miller, L., Oltmans, S. J., Ryerson, T. B., Cooper, O. R., Stohl, A. and de Gouw, J. A.: Ozone variability and halogen oxidation within the Arctic and sub-Arctic springtime boundary layer, Atmos. Chem. Phys., 10(21), 10223–10236, doi:10.5194/acp-10-10223-2010, 2010.

Gilman, J. B., Lerner, B. M., Kuster, W. C. and de Gouw, J. A.: Source Signature of Volatile Organic Compounds from Oil and Natural Gas Operations in Northeastern Colorado, Environ. Sci. Technol., 47(3), 1297–1305, doi:10.1021/es304119a, 2013.

Guo, H., Wang, T. and Louie, P. K. K.: Source apportionment of ambient non-methane hydrocarbons in Hong Kong: Application of a principal component analysis/absolute principal component scores (PCA/APCS) receptor model, Environ. Pollut., 129(3), 489–498, doi:10.1016/j.envpol.2003.11.006, 2004.

Guo, H., Zou, S. C., Tsai, W. Y., Chan, L. Y. and Blake, D. R.: Emission characteristics of nonmethane hydrocarbons from private cars and taxis at different driving speeds in Hong Kong, Atmos. Environ., 45(16), 2711–2721, doi:10.1016/j.atmosenv.2011.02.053, 2011.

Harley, R. A., McKeen, S. A., Pearson, J., Rodgers, M. O. and Lonneman, W. A.: Analysis of motor vehicle emissions during the Nashville/Middle Tennessee Ozone Study, J. Geophys. Res. Atmos., 106(D4), 3559–3567, doi:10.1029/2000JD900677, 2001.

He, Z., Li, G., Chen, J., Huang, Y., An, T. and Zhang, C.: Pollution characteristics and health risk assessment of volatile organic compounds emitted from different plastic solid waste recycling workshops, Environ. Int., 77, 85–94, doi:10.1016/j.envint.2015.01.004, 2015.

Helmig, D., Thompson, C. R., Evans, J., Boylan, P., Hueber, J. and Park, J.-H.: Highly Elevated Atmospheric Levels of Volatile Organic Compounds in the Uintah Basin, Utah, Environ. Sci. Technol., 48(9), 4707–4715, doi:10.1021/es405046r, 2014.

Ho, K. F., Lee, S. C., Guo, H. and Tsai, W. Y.: Seasonal and diurnal variations of volatile organic compounds (VOCs) in the atmosphere of Hong Kong, Sci. Total Environ., 322(1–3), 155–166, doi:10.1016/j.scitotenv.2003.10.004, 2004.

Hoque, R. R., Khillare, P. S., Agarwal, T., Shridhar, V. and Balachandran, S.: Spatial and temporal variation of BTEX in the urban atmosphere of Delhi, India, Sci. Total Environ., 392(1), 30–40, doi:10.1016/j.scitotenv.2007.08.036, 2008.

Hsieh, L., Yang, H. and Chen, H.: Ambient BTEX and MTBE in the neighborhoods of different industrial parks in Southern Taiwan, J. Hazard. Mater., 128(2–3), 106–115, doi:10.1016/j.jhazmat.2005.08.001, 2006.

Jia, C., Mao, X., Huang, T., Liang, X., Wang, Y., Shen, Y., Jiang, W., Wang, H., Bai, Z., Ma, M., Yu, Z., Ma, J. and Gao, H.: Non-methane hydrocarbons (NMHCs) and their contribution to ozone formation potential in a petrochemical industrialized city, Northwest China, Atmos. Res., 169, 225–236, doi:10.1016/j.atmosres.2015.10.006, 2016.

Jobson, B. T.: Hydrocarbon source signatures in Houston, Texas: Influence of the petrochemical industry, J. Geophys. Res., 109(D24), doi:10.1029/2004JD004887, 2004.

Leuchner, M. and Rappenglück, B.: VOC source–receptor relationships in Houston during TexAQS-II, Atmos. Environ., 44(33), 4056–4067, doi:10.1016/j.atmosenv.2009.02.029, 2010.

Li, L., Xie, S., Zeng, L., Wu, R. and Li, J.: Characteristics of volatile organic compounds and their role in ground-level ozone formation in the Beijing-Tianjin-Hebei region, China, Atmos. Environ., 113, 247–254, doi:10.1016/j.atmosenv.2015.05.021, 2015.

Li, L., An, J. Y., Shi, Y. Y., Zhou, M., Yan, R. S., Huang, C., Wang, H. L., Lou, S. R., Wang, Q., Lu, Q. and Wu, J.: Source apportionment of surface ozone in the Yangtze River Delta, China in the summer of 2013, Atmos. Environ., 144, 194–207, doi:10.1016/j.atmosenv.2016.08.076, 2016.

Lin, T., Sree, U., Tseng., Chiu, K., Wu, C. and Lo, J.: Volatile organic compound concentrations in ambient air of Kaohsiung petroleum refinery in Taiwan, Atmos. Environ., 38(25), 4111–4122, doi:10.1016/j.atmosenv.2004.04.025, 2004.

Ling, Z., Guo, H., Cheng, H. and Yu, Y. F.: Sources of ambient volatile organic compounds and their contributions to photochemical ozone formation at a site in the Pearl River Delta, southern China, Environ. Pollut., 159(10), 2310–2319, doi:10.1016/j.envpol.2011.05.001, 2011.

Liu, B., Liang, D., Yang, J., Dai, Q., Bi, X., Feng, Y., Yuan, J., Xiao, Z., Zhang, Y. and Xu, H.: Characterization and source apportionment of volatile organic compounds based on 1-year of observational data in Tianjin, China, Environ. Pollut., 218, 757–769, doi:10.1016/j.envpol.2016.07.072, 2016.

Liu, Y., Shao, M., Fu, L., Lu, S., Zeng, L. and Tang, D.: Source profiles of volatile organic compounds (VOCs) measured in China: Part I, Atmos. Environ., 42(25), 6247–6260, doi:10.1016/j.atmosenv.2008.01.070, 2008b.

Lyu, X., Chen, N., Guo, H., Zhang, W., Wang, N., Wang, Y. and Liu, M.: Ambient volatile organic compounds and their effect on ozone production in Wuhan, central China, Sci. Total Environ., 541, 200–209, doi:10.1016/j.scitotenv.2015.09.093, 2016.

McCarthy, M. C., Aklilu, Y.-A., Brown, S. G. and Lyder, D. A.: Source apportionment of volatile organic compounds measured in Edmonton, Alberta, Atmos. Environ., 81, 504–516, doi:10.1016/j.atmosenv.2013.09.016, 2013.

McGaughey, G. R., Desai, N. R., Allen, D. T., Seila, R. L., Lonneman, W. A., Fraser, M. P., Harley, R. A., Pollack, A. K., Ivy, J. M. and Price, J. H.: Analysis of motor vehicle emissions in a Houston tunnel during the Texas Air Quality Study 2000, Atmos. Environ., 38(20), 3363–3372, doi:10.1016/j.atmosenv.2004.03.006, 2004.

Miller, L., Xu, X., Grgicak-Mannion, A., Brook, J. and Wheeler, A.: Multi-season, multi-year concentrations and correlations amongst the BTEX group of VOCs in an urbanized industrial city, Atmos. Environ., 61, 305–315, doi:10.1016/j.atmosenv.2012.07.041, 2012.

Mo, Z., Shao, M., Lu, S., Qu, H., Zhou, M., Sun, J. and Gou, B.: Process-specific emission characteristics of volatile organic compounds (VOCs) from petrochemical facilities in the Yangtze River Delta, China, Sci. Total Environ., 533, 422–431, doi:10.1016/j.scitotenv.2015.06.089, 2015.

Mo, Z., Shao, M., Lu, S., Niu, H., Zhou, M. and Sun, J.: Characterization of non-methane hydrocarbons and their sources in an industrialized coastal city, Yangtze River Delta, China, Sci. Total Environ., 593–594, 641–653, doi:10.1016/j.scitotenv.2017.03.123, 2017.

Monod, A., Sive, B. C., Avino, P., Chen, T., Blake, D. R. and Sherwood Rowland, F.: Monoaromatic compounds in ambient air of various cities: a focus on correlations between the xylenes and ethylbenzene, Atmos. Environ., 35(1), 135–149, doi:10.1016/S1352-2310(00)00274-0, 2001.

Na, K. and Kim, Y.: Seasonal characteristics of ambient volatile organic compounds in Seoul, Korea, Atmos. Environ., 35, 2603–2614, doi:10.1016/S1352-2310(00)00464-7, 2001.

Pétron, G., Frost, G., Miller, B. R., Hirsch, A. I., Montzka, S. A., Karion, A., Trainer, M., Sweeney, C., Andrews, A. E., Miller, L., Kofler, J., Bar-Ilan, A., Dlugokencky, E. J., Patrick, L., Moore, C. T., Ryerson, T. B., Siso, C., Kolodzey, W., Lang, P. M., Conway, T., Novelli, P., Masarie, K., Hall, B., Guenther, D., Kitzis, D., Miller, J., Welsh, D., Wolfe, D., Neff, W. and Tans, P.: Hydrocarbon emissions characterization in the Colorado Front Range: A pilot study, J. Geophys. Res. Atmos., 117(D4), doi:10.1029/2011JD016360, 2012.

Qin, Y., Walk, T., Gary, R., Yao, X. and Elles, S.: C_2–C_{10} nonmethane hydrocarbons measured in Dallas, USA—Seasonal trends and diurnal characteristics, Atmos. Environ., 41(28), 6018–6032, doi:10.1016/j.atmosenv.2007.03.008, 2007.

Ran, L., Zhao, C., Geng, F., Tie, X., Tang, X., Peng, L., Zhou, G., Yu, Q., Xu, J. and Guenther, A.: Ozone photochemical production in urban Shanghai, China: Analysis based on ground level observations, J. Geophys. Res., 114(D15), doi:10.1029/2008JD010752, 2009.

Rodolfo Sosa, E., Humberto Bravo, A., Violeta Mugica, A., Pablo Sanchez, A., Emma Bueno, L. and Krupa, S.: Levels and source apportionment of volatile organic compounds in southwestern area of Mexico City, Environ. Pollut., 157(3), 1038–1044, doi:10.1016/j.envpol.2008.09.051, 2009.

Russo, R. S., Zhou, Y., White, M. L., Mao, H., Talbot, R. and Sive, B. C.: Multi-year (2004-2008) record of nonmethane hydrocarbons and halocarbons in New England: seasonal variations and regional sources, Atmos. Chem. Phys., 10(10), 4909–4929, doi:10.5194/acp-10-4909-2010, 2010.

Rutter, A. P., Griffin, R. J., Cevik, B. K., Shakya, K. M., Gong, L., Kim, S., Flynn, J. H. and Lefer, B. L.: Sources of air pollution in a region of oil and gas exploration downwind of a large city, Atmos. Environ., 120, 89–99, doi:10.1016/j.atmosenv.2015.08.073, 2015.

Shao, P., An, J., Xin, J., Wu, F., Wang, J., Ji, D. and Wang, Y.: Source apportionment of VOCs and the contribution to photochemical ozone formation during summer in the typical industrial area in the Yangtze River Delta, China, Atmos. Res., 176–177, 64–74, doi:10.1016/j.atmosres.2016.02.015, 2016.

Simpson, I. J., Blake, N. J., Barletta, B., Diskin, G. S., Fuelberg, H. E., Gorham, K., Huey, L. G., Meinardi, S., Rowland, F. S., Vay, S. A., Weinheimer, A. J., Yang, M. and Blake, D. R.: Characterization of trace gases measured over Alberta oil sands mining operations: 76 speciated C_2–C_{10} volatile organic compounds (VOCs), CO_2, CH_4, CO, NO, NO_2, NO_y, O_3 and SO_2, Atmos. Chem. Phys., 10(23), 11931–11954, doi:10.5194/acp-10-11931-2010, 2010.

Song, Y., Shao, M., Liu, Y., Lu, S., Kuster, W., Goldan, P. and Xie, S.: Source Apportionment of Ambient Volatile Organic Compounds in Beijing, Environ. Sci. Technol., 41(12), 4348–4353, doi:10.1021/es0625982, 2007.

Squizzato, S. and Masiol, M.: Application of meteorology-based methods to determine local and external contributions to particulate matter pollution: A case study in Venice (Italy), Atmos. Environ., 119, 69–81, doi:10.1016/j.atmosenv.2015.08.026, 2015.

Tang, J., Chan, L., Chan, C., Li, Y., Chang, C., Liu, S., Wu, D. and Li, Y.: Characteristics and diurnal variations of NMHCs at urban, suburban, and rural sites in the Pearl River Delta and a remote site in South China, Atmos. Environ., 41(38), 8620–8632, doi:10.1016/j.atmosenv.2007.07.029, 2007.

US EPA: Positive Matrix Factorization Model for environmental data analyses, [online] Available from: https://www.epa.gov/air-research/positive-matrix-factorization-model-environmental-data-analyses (Accessed 20 September 2016), 2014.

Vega, E., Sánchez-Reyna, G., Mora-Perdomo, V., Iglesias, G. S., Arriaga, J. L., Limón-Sánchez, T., Escalona-Segura, S. and Gonzalez-Avalos, E.: Air quality assessment in a highly industrialized area of Mexico: Concentrations and sources of volatile organic compounds, Fuel, 90(12), 3509–3520, doi:10.1016/j.fuel.2011.03.050, 2011.

Wang, B., Shao, M., Lu, S., Yuan, B., Zhao, Y., Wang, M., Zhang, S. and Wu, D.: Variation of ambient non-methane hydrocarbons in Beijing city in summer 2008, Atmos. Chem. Phys., 10(13), 5911–5923, doi:10.5194/acp-10-5911-2010, 2010.

Wang, H.: Source Profiles and Chemical Reactivity of Volatile Organic Compounds from Solvent Use in Shanghai, China, Aerosol Air Qual. Res., doi:10.4209/aaqr.2013.03.0064, 2014.

Wang, L., Liu, Z., Sun, Y., Ji, D. and Wang, Y.: Long-range transport and regional sources of $PM_{2.5}$ in Beijing based on long-term observations from 2005 to 2010, Atmos. Res., 157, 37–48, doi:10.1016/j.atmosres.2014.12.003, 2015.

Wang, Q., Liu, M., Yu, Y. and Li, Y.: Characterization and source apportionment of $PM_{2.5}$-bound polycyclic aromatic hydrocarbons from Shanghai city, China, Environ. Pollut., 218, 118–128, doi:10.1016/j.envpol.2016.08.037, 2016.

Wang, Y., Zhang, X. and Draxler, R. R.: TrajStat: GIS-based software that uses various trajectory statistical analysis methods to identify potential sources from long-term air pollution measurement data, Environ. Model. Softw., 24(8), 938–939, doi:10.1016/j.envsoft.2009.01.004, 2009.

Warneke, C., Geiger, F., Edwards, P. M., Dube, W., Pétron, G., Kofler, J., Zahn, A., Brown, S. S., Graus, M., Gilman, J. B., Lerner, B. M., Peischl, J., Ryerson, T. B., de Gouw, J. A. and Roberts, J. M.: Volatile organic compound emissions from the oil and natural gas industry in the Uintah Basin, Utah: oil and gas well pad emissions compared to ambient air composition, Atmos Chem Phys, 14(20), 10977–10988, doi:10.5194/acp-14-10977-2014, 2014.

Watson, J. G., Chow, J. C. and Fujita, E. M.: Review of volatile organic compound source apportionment by chemical mass balance, Atmos. Environ., 35(9), 1567–1584, doi:10.1016/S1352-2310(00)00461-1, 2001.

Wei, W., Lv, Z., Cheng, S., Wang, L., Ji, D., Zhou, Y., Han, L. and Wang, L.: Characterizing ozone pollution in a petrochemical industrial area in Beijing, China: a case study using a chemical reaction model, Environ. Monit. Assess., 187(6), 377, doi:10.1007/s10661-015-4620-5, 2015.

Xiao, Y., Logan, J. A., Jacob, D. J., Hudman, R. C., Yantosca, R. and Blake, D. R.: Global budget of ethane and regional constraints on U.S. sources, J. Geophys. Res., 113(D21), doi:10.1029/2007JD009415, 2008.

Yu, C. H., Zhu, X. and Fan, Z.: Spatial/Temporal Variations and Source Apportionment of VOCs Monitored at Community Scale in an Urban Area, edited by Y. Zhang, PLoS ONE, 9(4), e95734, doi:10.1371/journal.pone.0095734, 2014.

Yuan, B., Shao, M., Lu, S. and Wang, B.: Source profiles of volatile organic compounds associated with solvent use in Beijing, China, Atmos. Environ., 44(15), 1919–1926, doi:10.1016/j.atmosenv.2010.02.014, 2010.

Zhang, R., Jing, J., Tao, J., Hsu, S.-C., Wang, G., Cao, J., Lee, C. S. L., Zhu, L., Chen, Z., Zhao, Y. and Shen, Z.: Chemical characterization and source apportionment of $PM_{2.5}$ in Beijing: seasonal perspective, Atmos. Chem. Phys., 13(14), 7053–7074, doi:10.5194/acp-13-7053-2013, 2013a.

Zhang, Y., Wang, X., Zhang, Z., Lü, S., Shao, M., Lee, F. S. C. and Yu, J.: Species profiles and normalized reactivity of volatile organic compounds from gasoline evaporation in China, Atmos. Environ., 79, 110–118, doi:10.1016/j.atmosenv.2013.06.029, 2013b.

Zhao, M., Huang, Z., Qiao, T., Zhang, Y., Xiu, G. and Yu, J.: Chemical characterization, the transport pathways and potential sources of $PM_{2.5}$ in Shanghai: Seasonal variations, Atmos. Res., 158–159, 66–78, doi:10.1016/j.atmosres.2015.02.003, 2015.

Zou, Y., Deng, X. J., Zhu, D., Gong, D. C., Wang, H., Li, F., Tan, H. B., Deng, T., Mai, B. R., Liu, X. T. and Wang, B. G.: Characteristics of 1 year of observational data of VOCs, NO_x and O_3 at a suburban site in Guangzhou, China, Atmos. Chem. Phys., 15(12), 6625–6636, doi:10.5194/acp-15-6625-2015, 2015.

Table 1. Concentrations (mean ± standard deviation) during the sampling period and the photochemical properties of VOCs.

Species	r^2 [a]	MDL [b]	MIR [c]	k_{OH} ($\times 10^{12}$) [d]	ppbv	ppbC
Alkanes					129 ± 173	387 ± 439
Ethane [e]	0.997	0.05	0.25	0.25	39.7 ± 57.3	53.2 ± 76.7
Propane	0.999	0.021	0.48	1.09	22.6 ± 33.5	44.5 ± 65.7
***i*-Butane**	0.994	0.012	1.21	2.12	12.5 ± 17.5	32.3 ± 45.2
***n*-Butane**	0.994	0.03	1.02	2.36	15.8 ± 21.4	40.8 ± 55.5
Cyclopentane	0.997	0.026	2.4	4.97	8.64 ± 16.0	26.7 ± 50.1
***i*-Pentane**	0.998	0.012	1.38	3.6	8.96 ± 13.3	28.8 ± 42.7
***n*-Pentane**	0.984	0.026	1.04	3.94	8.81 ± 12.4	28.3 ± 39.7
2,2-Dimethylbutane	0.998	0.007	0.82	2.23	0.24 ± 0.60	0.92 ± 2.30
2,3-Dimethylbutane	0.999	0.005	1.07	5.78	1.81 ± 2.91	6.94 ± 11.2
2-Methylpentane	0.984	0.005	1.5	5.2	3.36 ± 5.96	12.9 ± 22.9
3-Methylpentane	0.998	0.007	1.5	5.2	1.40 ± 2.77	5.38 ± 10.6
***n*-Hexane**	0.997	0.016	0.98	5.2	3.08 ± 4.88	11.8 ± 18.7
2,4-Dimethylpentane	0.999	0.005	1.5	4.77	0.13 ± 0.56	0.56 ± 2.49
Methylcyclopentane	0.999	0.008	2.8	–	1.28 ± 3.30	4.81 ± 12.4
2-Methylhexane	0.996	0.008	1.08	–	0.57 ± 1.19	2.56 ± 5.31
Cyclohexane	0.997	0.004	1.28	6.97	1.23 ± 2.03	4.61 ± 7.61
2,3-Dimethylpentane	0.997	0.016	1.31	–	0.85 ± 1.37	3.79 ± 6.11
3-Methylhexane	0.995	0.006	1.4	–	0.77 ± 1.39	3.42 ± 6.21
2,2,4-Trimethylpentane	0.994	0.003	0.93	3.34	0.05 ± 0.42	0.23 ± 2.14
n-Heptane	0.994	0.007	0.81	6.76	4.07 ± 30.5	18.2 ± 136
Methylcyclohexane	0.995	0.008	1.8	9.64	1.43 ± 2.47	7.28 ± 12.6
2,3,4-Trimethylpentane	0.994	0.008	1.6	6.6	0.06 ± 0.48	0.32 ± 2.42
2-Methylheptane	0.99	0.008	0.96	–	0.87 ± 1.41	4.43 ± 7.20
3-Methylheptane	0.991	0.009	0.99	–	0.20 ± 0.53	1.03 ± 2.71
***n*-Octane**	0.989	0.121	0.6	8.11	0.97 ± 1.28	4.93 ± 6.51
***n*-Nonane**	0.998	0.021	0.54	9.7	0.29 ± 0.59	1.68 ± 3.37
***n*-Decane**	0.995	0.03	0.46	11	0.23 ± 0.32	1.44 ± 2.05
***n*-Undecane**	0.992	0.02	0.42	12.3	0.19 ± 0.24	1.29 ± 1.66
***n*-Dodecane**	0.993	0.01	0.38	13.2	39.7 ± 57.3	53.2 ± 76.7
Alkenes					9.52 ± 14.5	30.6 ± 41.7
Ethylene	0.997	0.003	7.4	8.52	1.42 ± 1.69	1.78 ± 2.11
Propylene	0.998	0.025	9.4	26.3	1.88 ± 10.2	3.53 ± 19.1
trans-2-butene	0.997	0.031	10	64	0.60 ± 1.34	1.50 ± 3.34
1-Butene	0.994	0.03	8.9	31.4	0.63 ± 1.04	1.58 ± 2.59
cis-2-butene	0.999	0.023	10	56.4	0.70 ± 2.06	1.75 ± 5.15
1-Pentene	0.993	0.03	6.2	31.4	4.47 ± 6.72	14.0 ± 21.0
trans-2-Pentene	0.998	0.009	8.8	67	0.19 ± 0.67	0.60 ± 2.11
Isoprene	0.998	0.008	9.1	101	0.20 ± 0.75	0.62 ± 2.28
cis-2-Pentene	0.998	0.015	8.8	65	0.09 ± 0.25	0.28 ± 0.80
1-Hexene	0.984	0.008	4.4	37	1.36 ± 2.74	5.09 ± 10.3
Acetylene	0.998	0.048	0.5		3.03 ± 5.55	3.52 ± 6.44
Aromatics					4.28 ± 8.24	22.2 ± 25.3
Benzene	0.997	0.007	0.42	1.22	1.13 ± 1.62	3.95 ± 5.66
Toluene	0.995	0.005	2.7	5.63	1.06 ± 1.91	4.34 ± 7.84
Ethylbenzene	0.992	0.003	2.7	7	0.30 ± 2.40	1.41 ± 11.4
***m*, *p*-Xylene**	0.986	0.002	7.4	18.7	0.72 ± 1.94	3.42 ± 9.19

_o_-Xylene	0.989	0.003	6.5	13.6	0.20 ± 0.59	0.95 ± 2.79
Styrene	0.991	0.013	2.2	58	0.40 ± 2.60	1.87 ± 12.1
iso-Propylbenzene	0.986	0.02	2.2	6.3	0.06 ± 0.24	0.32 ± 1.28
n-Propylbenzene	0.986	0.016	2.1	5.8	0.05 ± 0.16	0.29 ± 0.85
m-ethyltoluene	0.989	0.02	–	18.6	0.09 ± 0.20	0.49 ± 1.06
p-ethyltoluene	0.992	0.02	–	11.8	0.08 ± 0.18	0.44 ± 0.98
1,3,5-Trimethylbenzene	0.989	0.004	10.1	56.7	0.09 ± 0.18	0.47 ± 0.96
o-ethyltoluene	0.999	0.02	–	11.9	0.07 ± 0.18	0.35 ± 0.95
1,2,4-Trimethylbenzene	0.991	0.003	8.8	32.5	0.14 ± 0.25	0.74 ± 1.34
1,2,3-Trimethylbenzene	0.993	0.002	8.9	32.7	0.09 ± 0.16	0.46 ± 0.85
m-diethylbenzene	0.991	0.02	–	–	0.07 ± 0.08	0.40 ± 0.50
p-diethylbenzene	0.993	0.03	–	–	0.10 ± 0.10	0.57 ± 0.61

[a] Correlation coefficients of calibration curves

[b] Method detection limit

[c] Units: g O_3/g VOCs (Carter, 1994).

[d] Units: $\times 10^{-12}$ cm^3 $molecule^{-1}$ s^{-1} (Atkinson and Arey, 2003)

[e] Species in bold were used in source apportionment

Table 2 Comparison of VOCs source apportionment results with formers

Area	Sampling period	Model	Sources								
			Fuel evaporatio	LPG/NG	Industrial emission	Vehicle emission	Solvent usage	Coal or biomass	Stationery +mobile	Biogenic	others
Tianjing, urban [a]	Nov. 2014–Oct. 2015	PMF	8.7	18.6	19.9	39.1	4.7	10.6			9.0
Wuhan, urban [b]	Feb. 2013–Oct. 2014	PMF		19.8 ±0.9	14.4 ±0.9	27.8 ±0.9	16.2 ±0.4	21.8 ±0.9			
Lanzhou, downtown [c]	Jan.–Dec. 2013	PCA-MLR			19.5	58.3	22.2				
Paris, urban [d]	Jan. 2010–Dec. 2010	PMF	5	16		23	26	13		17	
Nanjing, industrial area [e]	Mar. 2011–Feb. 2012	PCA/APCS		15–48	15–23	29–50	6–15			1–4	15–23
Hong Kong, urban [f]	Jan. 2001–Dec. 2001	PCA/APCS		11–19.4	5.2–9	38.9–48	32–36.4			0.1	
Paterson, urban [g]	Nov. 2005–Dec. 2006	PMF			16	31	19		12		22
Houston，industrial area [h]	Aug. 2006–Sep. 2006	PMF		20–37	39–58	11–16					
Los Angeles, urban [i]	2001–2003 (Jul.–Sep.)	PMF	47–58	13	15	22–24				1–3	
This study	Sep. 2014–Aug. 2015	PMF	21.5 ±2.99	62.6 ±3.04	3.8 ±0.50			10.9 ±1.57			1.3 ±0.69

[a] Liu et al. (2016); [b] Lyu et al. (2016); [c] Jia et al. (2016); [d] Baudic et al. (2016); [e] An et al., (2014); [f] Guo et al. (2007); [g] Yu et al. (2014); [h] Leuchner and Rappenglück, (2010); [i] Brown et al. (2007)

Table 3 Contributions (%) of local sources and regional transport of five sources in different seasons.

Seasons	Oil refining		NG		Combustion source		Asphalt		Fuel evaporation	
	Local	Regional	Local	Regional	Local	Regional	Local	Regional	Local	Regional
Autumn	64.5	35.5	74.6	25.4	68.6	31.4	65.4	34.6	68.3	31.7
Winter	60.1	39.9	60.0	40.0	58.5	41.5	60.3	39.7	59.0	41.0
Spring	66.5	33.5	66.0	34.0	59.7	40.3	64.0	36.0	60.0	40.0
Summer	69.4	30.6	71.9	28.1	69.2	30.8	62.7	37.3	65.9	34.1
Annual	53.6	46.4	54.5	45.5	48.8	51.2	50.5	49.5	50.6	49.4

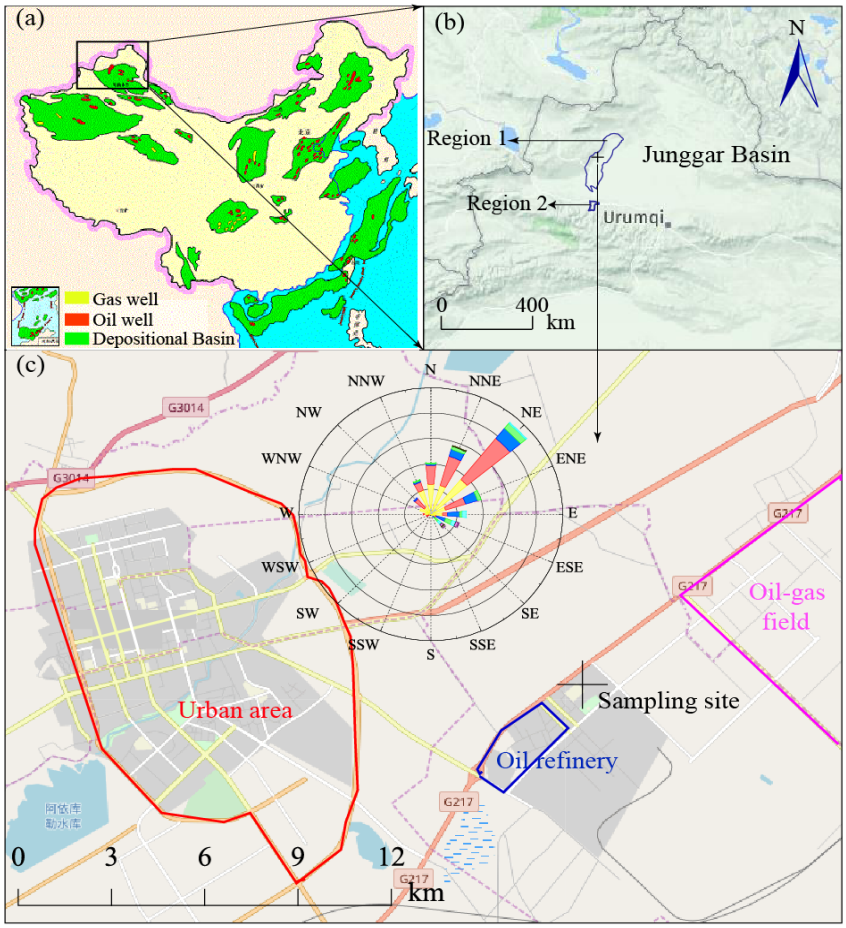

Figure 1. The spatial distribution of oil and gas bearing basins in China (a) and the terrain of the study area (b). The sampling site is about 11 km away from the urban area and located in the northeast of an oil refinery plant and southwest of an oil and gas field. The northeasterly winds prevailed during the sampling periods (c).

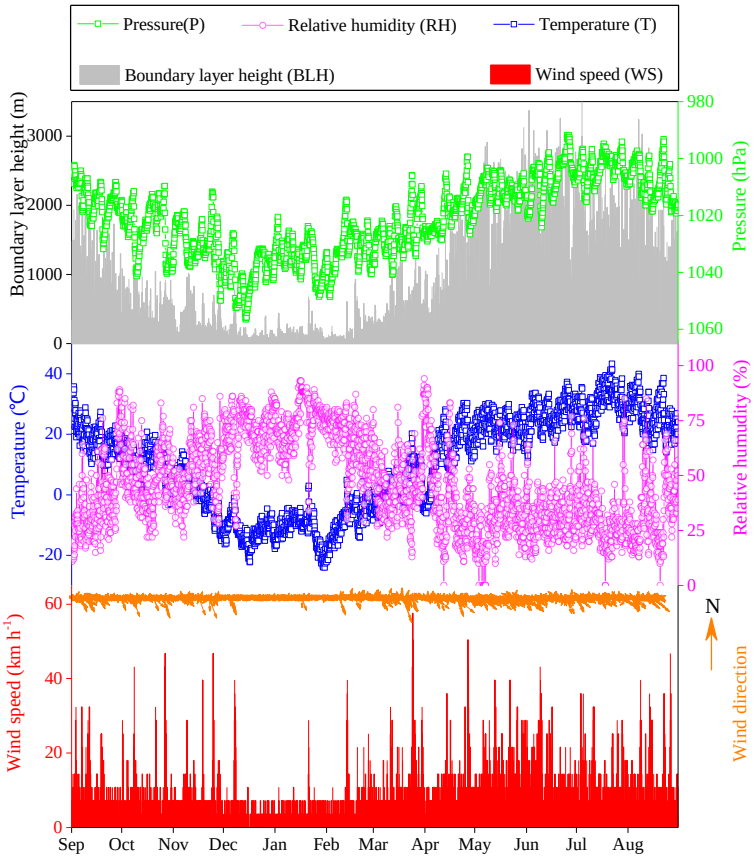

Figure 2. Every three-hour meteorological parameters at the observation site from September 2014 to August 2015.

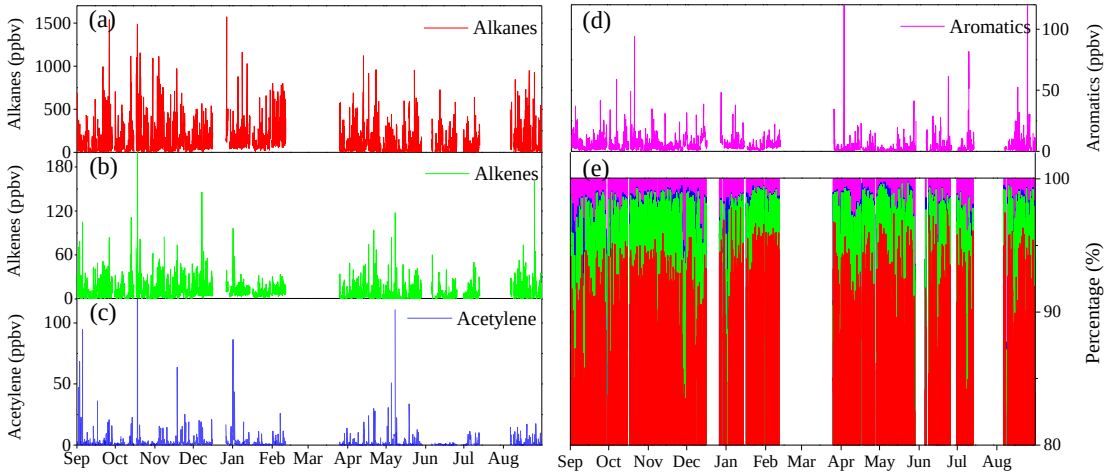

Figure 3. Time series of every two hours' concentrations (expressed in ppbv) for the four of VOCs categories including alkanes (a), alkenes (b), acetylene (c), aromatics (d) and their fractions (e) during the sampling period.

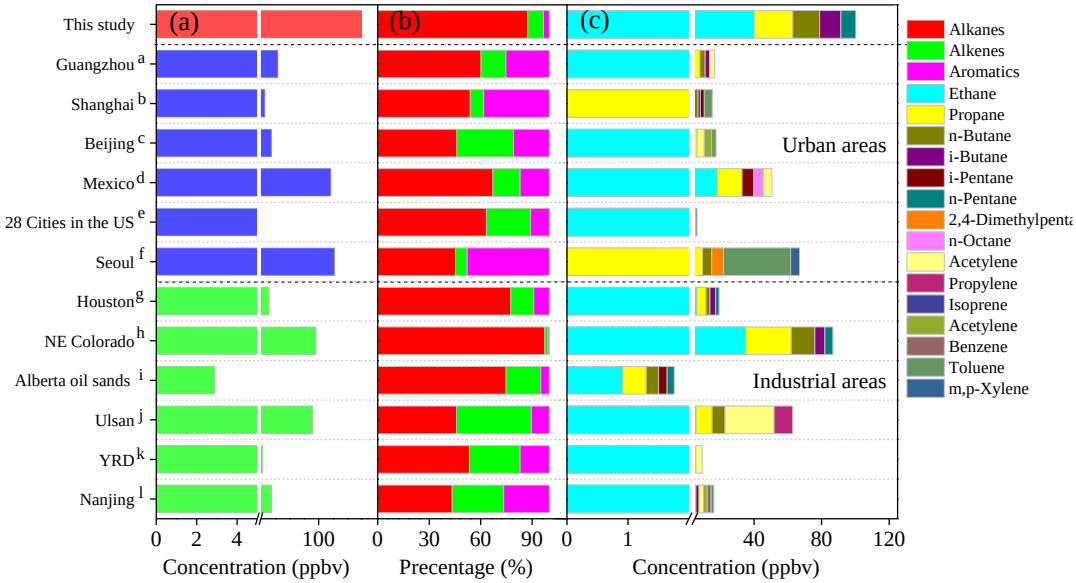

Figure 4. Comparison of the VOCs concentrations (a), compositions (b) and the top five VOCs species (c) in this study with former studies concerning the VOCs in ambient air of urban and industrial areas.
[a] Zou et al. (2015); [b] Cai et al. (2010); [c] Wang et al. (2010); [d] Garzón et al. (2015); [e] Baker et al. (2008); [f] Na and Kim, (2001); [g] Leuchner and Rappenglück, (2010); [h] Gilman et al. (2013); [i] Simpson et al. (2010); [j] Na et al. (2001); [k] An et al. (2014); [l] Shao et al. (2016)

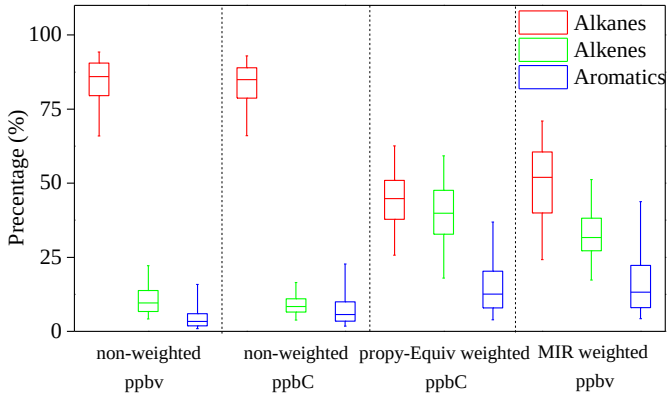

Figure 5. Box and whisker plots of VOCs profiles based on different scales during the whole sampling period. Box and Whisker plots are constructed according to 25th–75th and 5th–95th percentile of the calculation results.

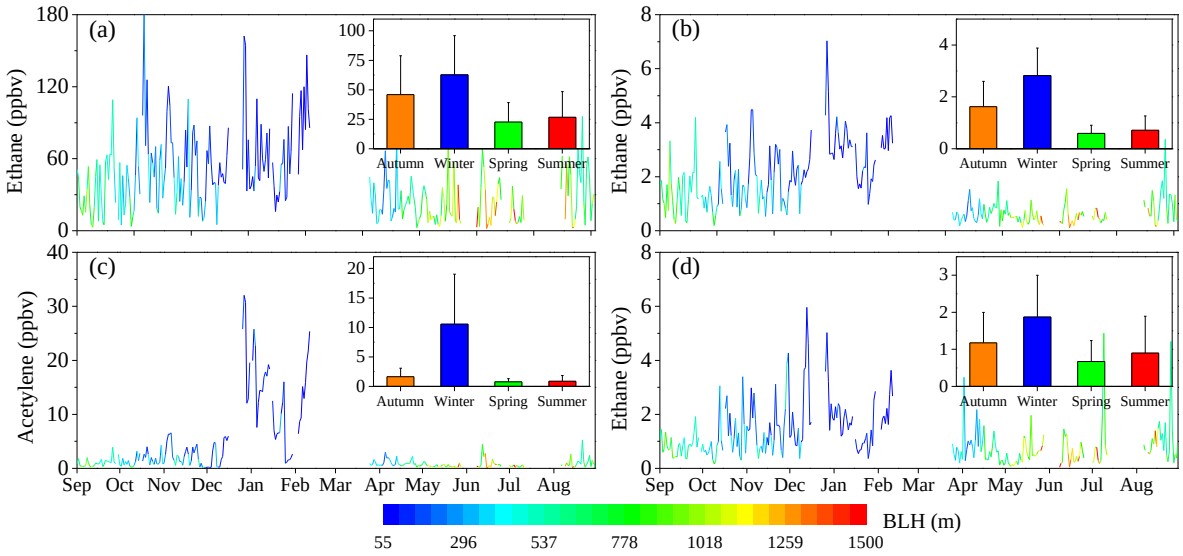

Figure. 6 Seasonal and daily variations of ethane (a), ethylene (b), acetylene (c) and benzene (d) during the sampling period.

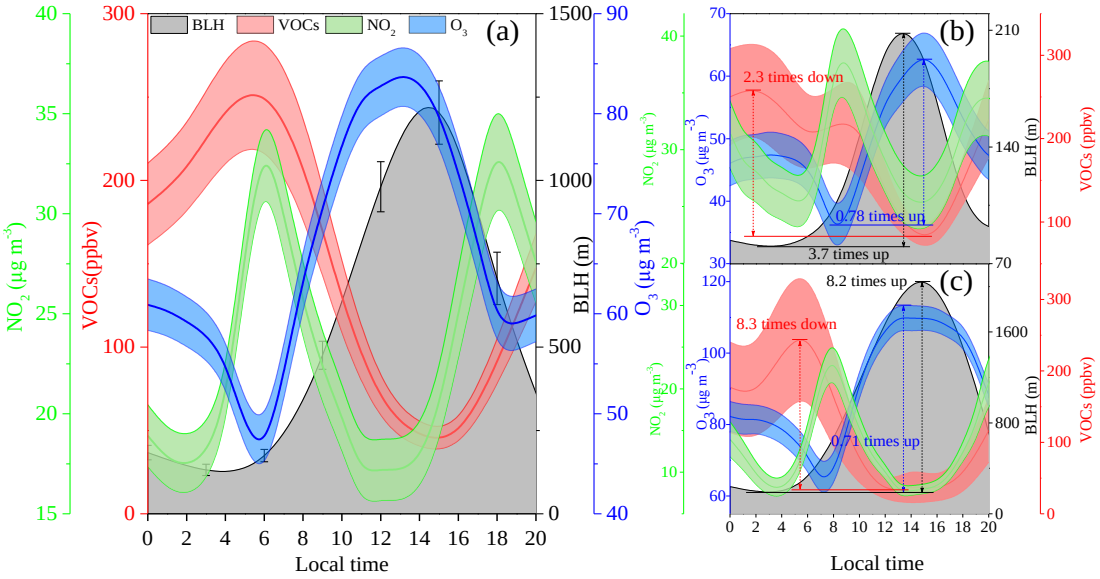

Figure 7. Diurnal variation of boundary layer height (BLH), VOCs, NO_2, and O_3 concentrations in different timescale: annual (a), winter (b) and summer (c). Solid line represents the average value and filled area indicates the 95th confidence intervals of the mean.

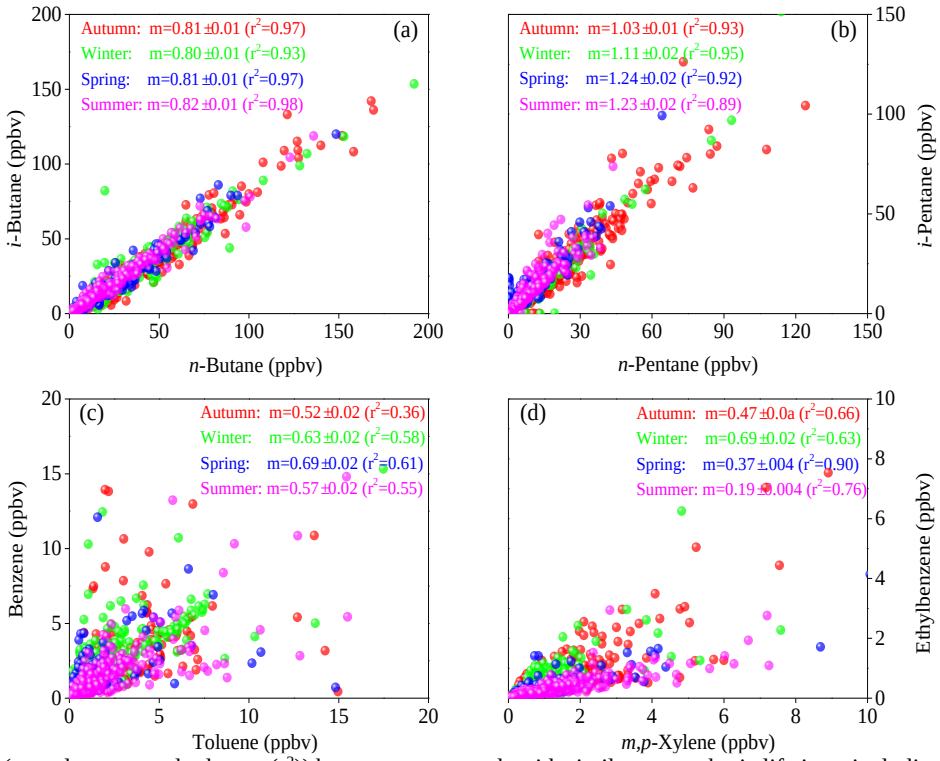

Figure 8. Correlations (m = slope ± standard error (r^2)) between compounds with similar atmospheric lifetimes including *i*-butane/*n*-butane (a) and *i*-pentane/*n*-pentane (b), and compounds with different lifetimes including benzene/toluene (c) and ethylbenzene/ *m*, *p*-xylenes (d).

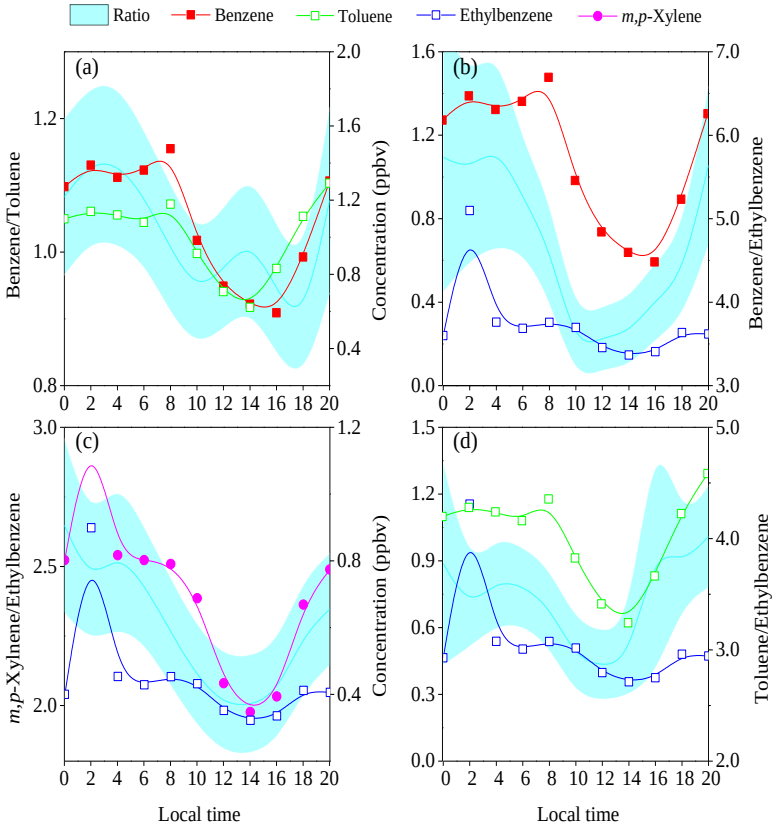

Figure 9. Diurnal variations of benzene, toluene, ethylbenzene and *m*, *p*-xylene and their ratios: benzene/toluene (a), benzene/ethylbenzene (b), *m*, *p*-xylene/ethylbenzene (c), and toluene/ethylbenzene (d).

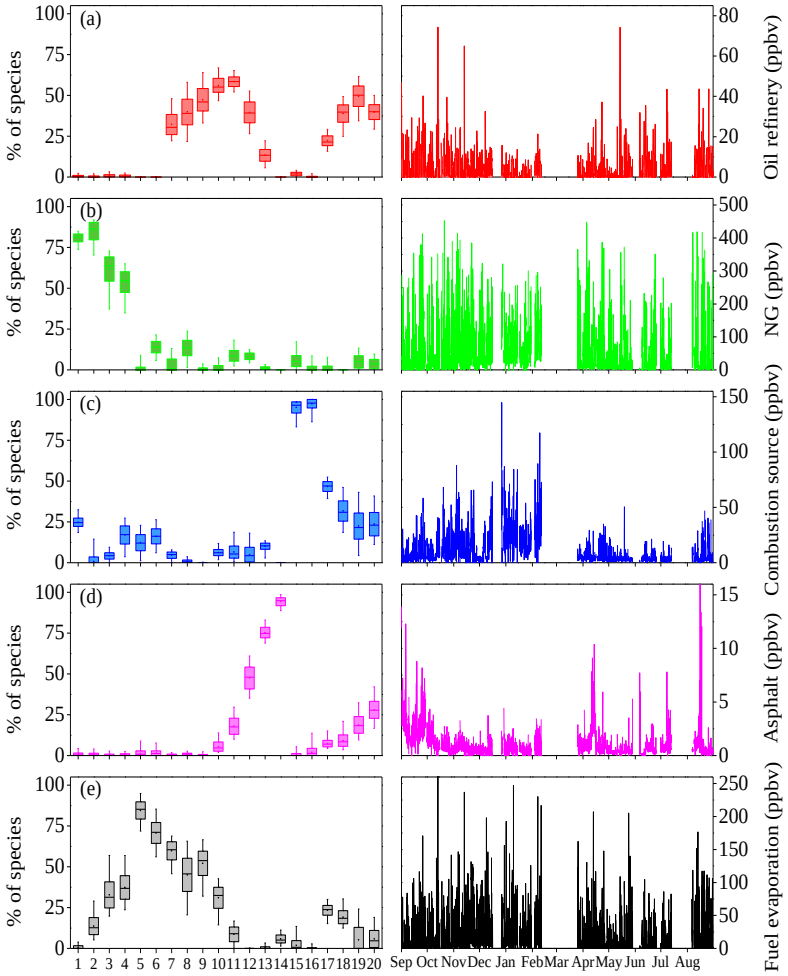

Figure 10. Source profiles of five factors resolved by PMF modeling including oil refinery (a), NG (b), combustion source (c), asphalt (d) and fuel evaporation (e), and their corresponding hourly source contributions. Box and whisker plots are constructed according to the 5th-95th percentiles of the F-peak bootstrap runs (n = 100).

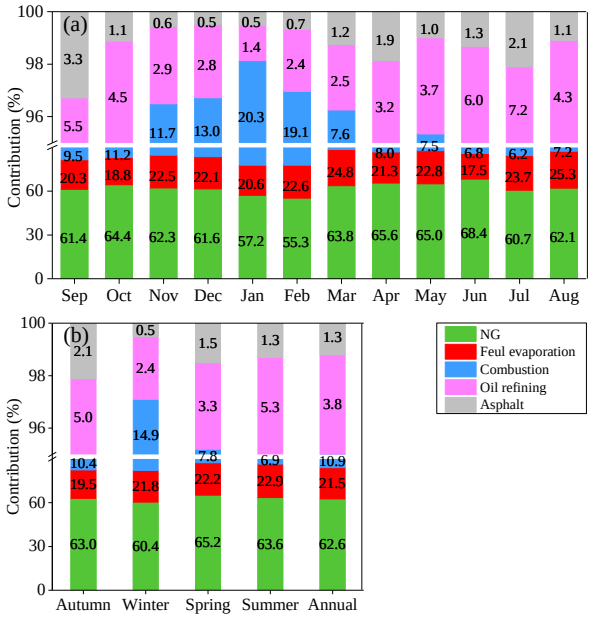

Figure 11. Variation of monthly averaged (a) and seasonal averaged (b) contributions of five identified VOCs sources (expressed in %).

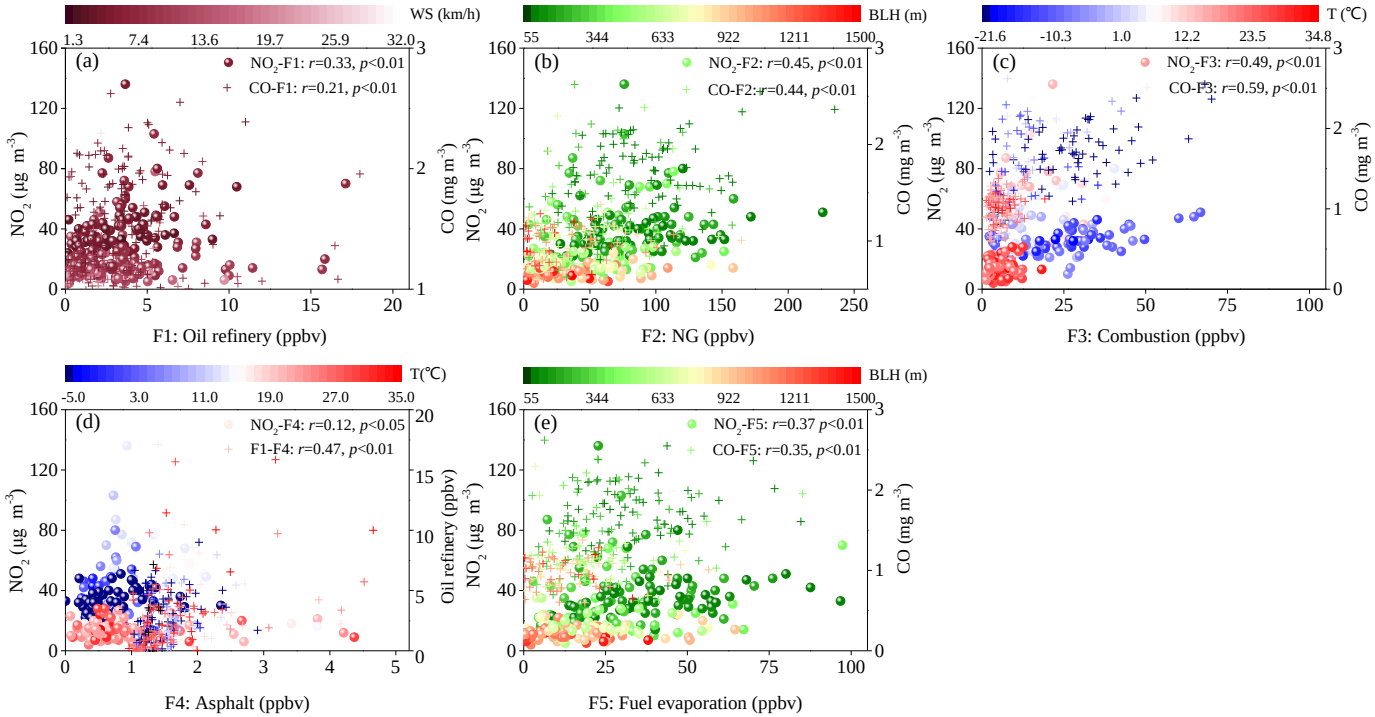

Figure 12. Scatter plots of daily concentrations of trace gas and source contributions including oil refinery (a), NG (b), combustion (c), asphalt (d) and fuel evaporation (e) under different meteorological conditions (wind speed (WS), boundary layer height (BLH) and temperature (T)).

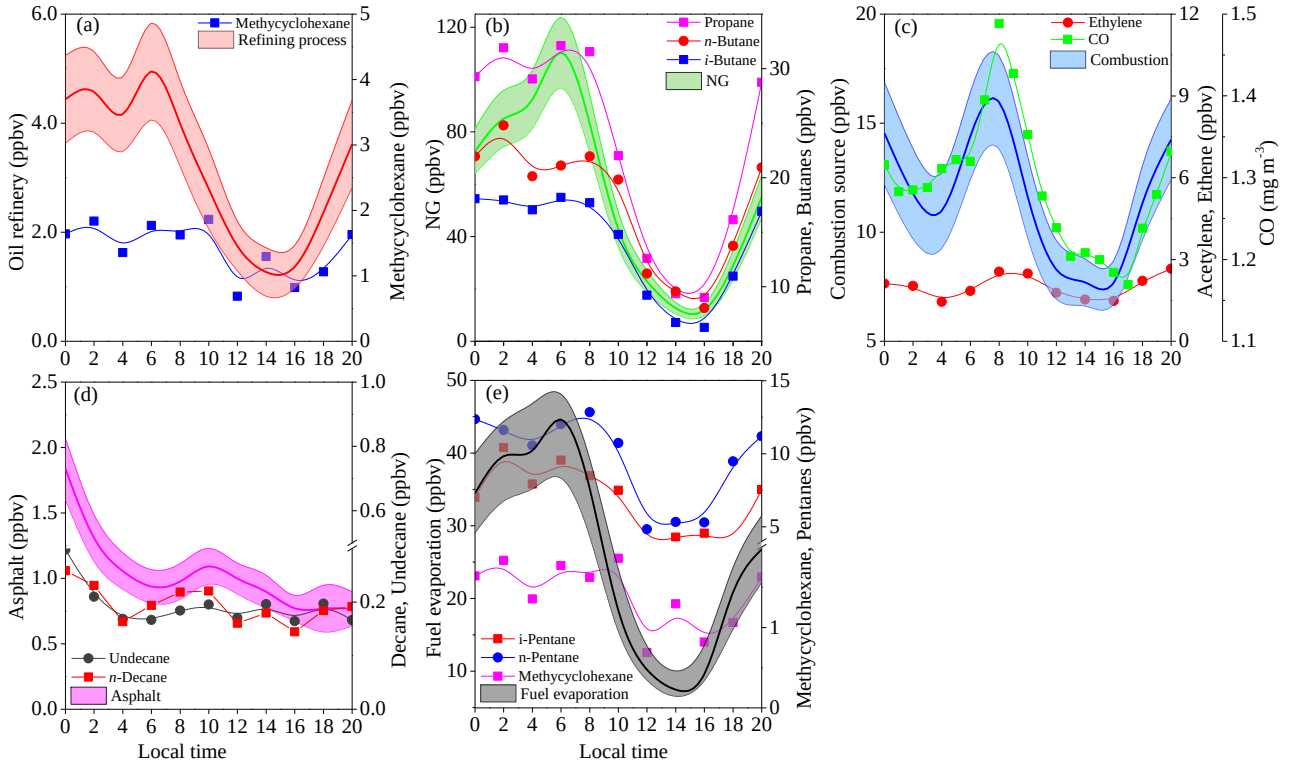

Figure 13. Diurnal variation of the contributions (expressed in ppbv) of five identified sources including oil refining process (a), NG (b), combustion source (c), asphalt (d) and fuel evaporation (e), and specific compounds with high loadings in each source profile. Note that the CO in combustion source was expressed in mg m^{-3}.

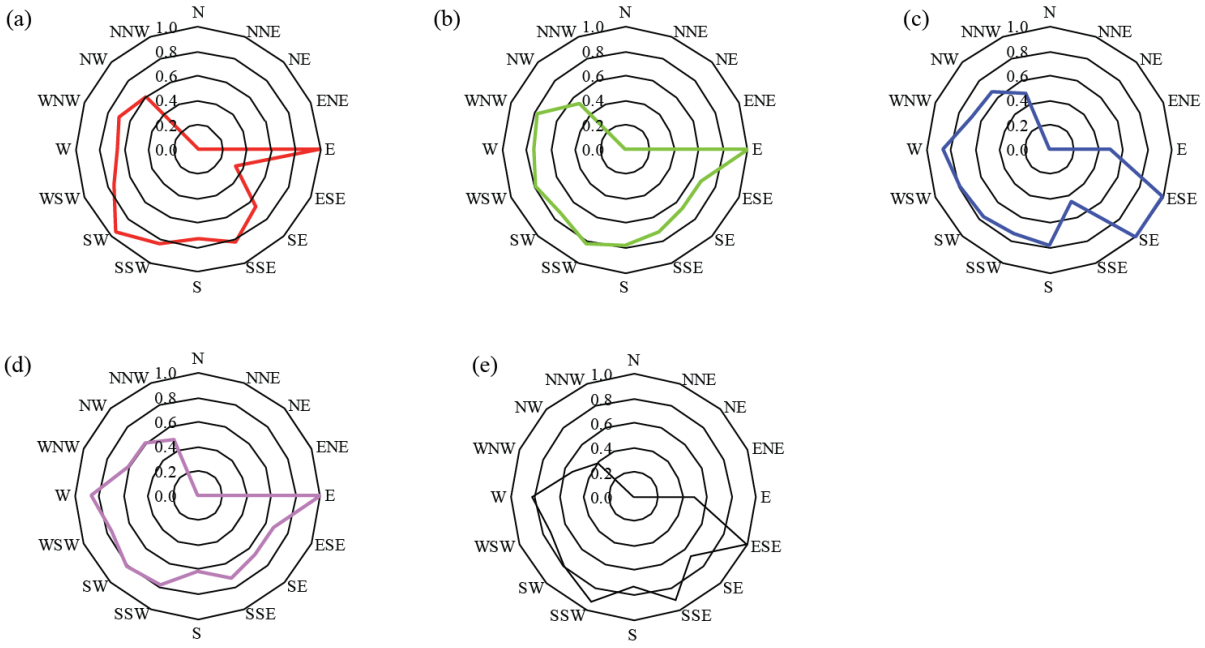

Figure 14. Annual conditional probability function (CPF) plots of five identified VOCs source including oil refinery (a), NG (b), combustion (c), asphalt (d) and fuel evaporation (e).

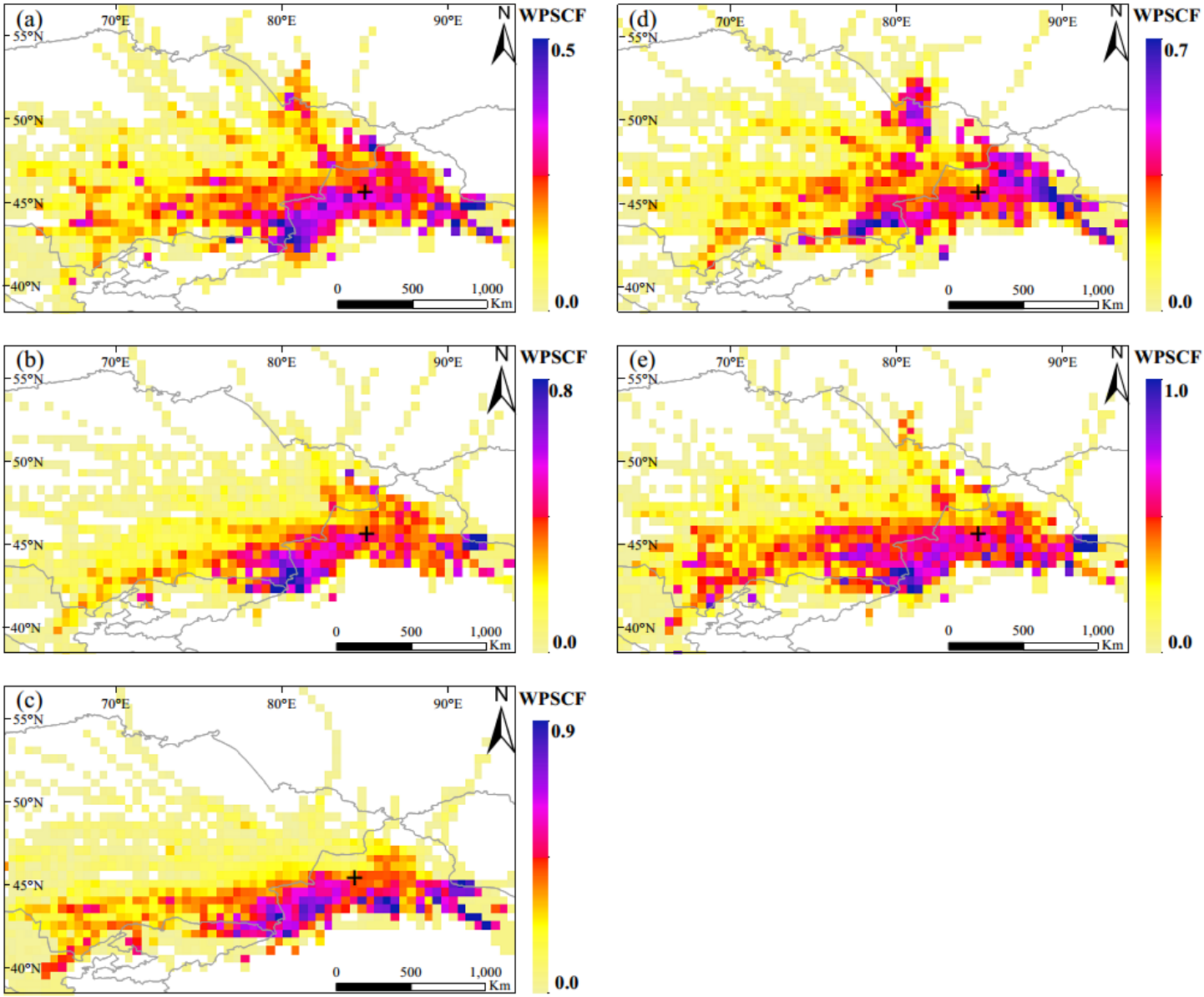

Figure 15. Annual weight potential source contribution function (WPSCF) maps for five identified sources derived from PMF analysis including oil refining (a), NG (b), combustion source (c), asphalt (d) and fuel evaporation (e). The black cross represents the sampling site.

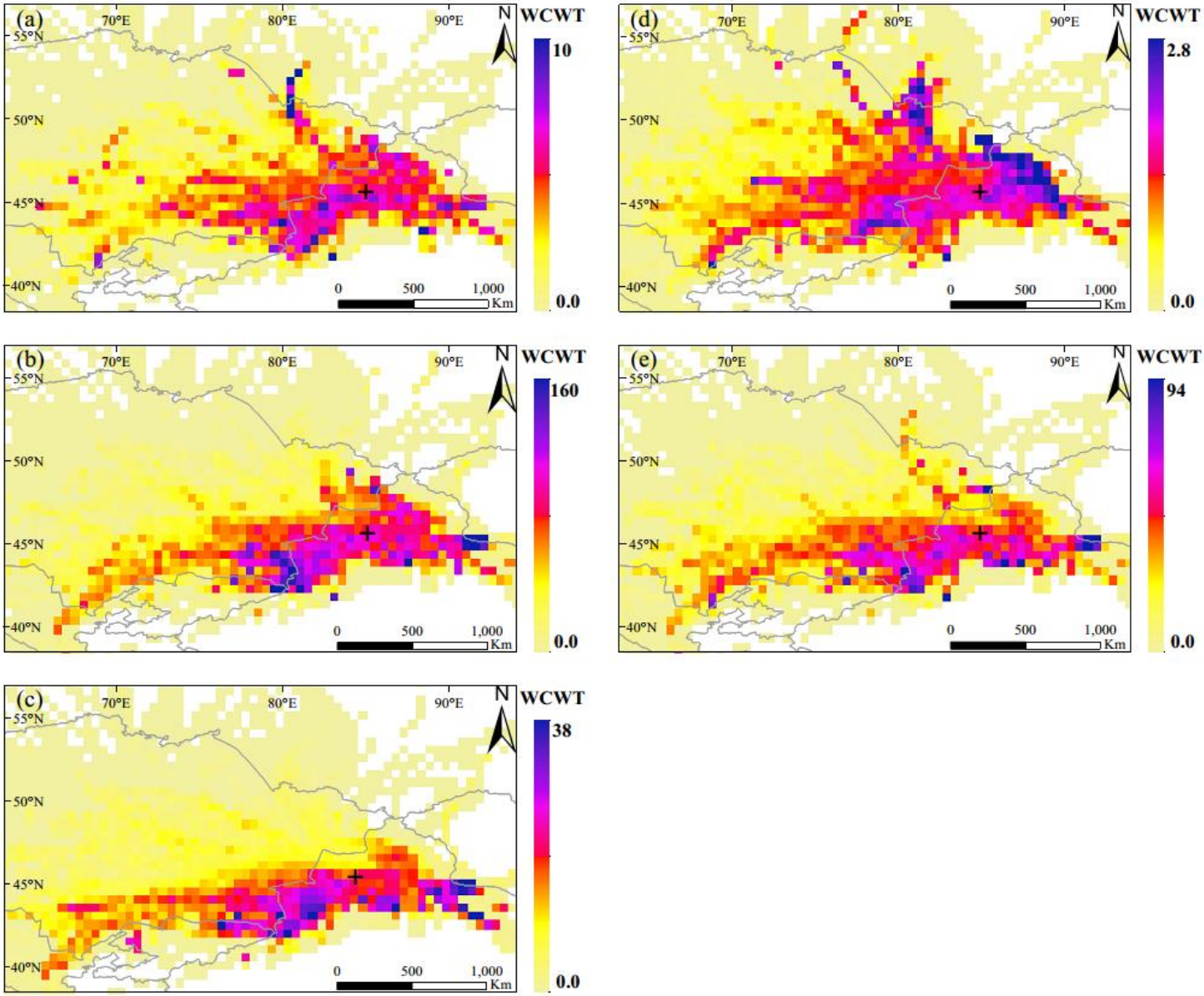

Figure 16. Annual weight concentration weighted trajectory (WCWT) maps for five identified sources derived from PMF analysis including oil refining (a), NG (b), combustion source (c), asphalt (d) and fuel evaporation (e). The black cross represents the sampling site.

Appendix A Detail operation of positive matrix factorization (PMF) in source apportionment of VOCs dataset.

A1 Data preparation

Two files including species concentration and uncertainty are required to be introduced into the EPA PMF 5.0 model. The concentration file is a i (number of samples) ×j dimension (number of species) matrix (**X** matrix), i.e., 2743×20 in this study. There are two types of uncertainty files: sample-specific and equation-based. The sample-specific uncertainty file is also a matrix with the same dimension as the concentration matrix. The equation-based uncertainty dataset is constructed according to the method detection limit (MDL) and error fraction (%):

$$U_{ij} = \frac{5}{6} \times \mathrm{MDL} \qquad \text{Cix.} \tag{1}$$

$$U_{ij} = \sqrt{(\mathrm{ER} \times \mathrm{concentration})^2 + (\mathrm{MDL})^2} \qquad \text{C>MDL} \tag{2}$$

Not all fifty-seven VOCs are introduced into the PMF model, there are some rules to decide which species should be included or excluded from the PMF model: (1) highly collinear species, such as propane & *n*-butane, benzene & toluene are included (Fig. A1); (2) species indicating VOC sources (i.e., acetylene is the marker of combustion sources) are retained; (3) species that are highly reactive are excluded (i.e., *i*-pentene), since they are rapidly reacted away in the ambient atmosphere (Guo et al., 2011b; Shao et al., 2016). Prior to the PMF model base run, the retained species were firstly classified into strong, weak, and bad based on their signal-noise-ratios (S/N). Species with S/N ratios less than 0.5 were grouped into bad and grouped into weak if S/N ratios are in the range of 0.5–1.0 (US EPA, 2014). However, the S/N ratios were not useful to categorize species because all species have S/N ratios greater than 2.0 in this study. Therefore, the percentage of samples below the detection limit (BDL), residual scale and priority knowledge of VOC source tracers are used. The species with BDL greater than 60% were categorized as bad and were excluded from the model (i.e., *trans*/*cis*-2-pentene, isoprene); species with BDL > 50% were characterized as weak (Callén et al., 2014). Finally, nine species (ethane, propane, *n*-hexane, cyclohexane, methylcyclohexane, *n*-octane, *n*-nonane, *n*-decane and *u*-undecane) were categorized as strong and eleven species (*i*-butane, *n*-butane, *i*-pentane, *n*-pentane, *n*-dodecane, ethylene, acetylene, benzene, toluene, *m*, *p*-xylene and *o*-xylene) were characterized as weak due to its residual scale beyond 3.

A2 The optimal number of factors.

Choosing the optimal number of factors (P-value) is a critical question in PMF analysis. Too many factors will result in meaningless factor profiles, while too few factors will make it difficult to segregate the mixing sources (Bressi et al., 2014). Factors ranging from 3 to 8 were tested in this study. Each model was run for 20 times with a random seed. All the *Q* values (Q_{true}, Q_{robust}, Q_{except} and Q_{ture}/Q_{except}), observed verse predicted (O/P) concentrations and scaled residuals were evaluated. In theory, if the number of sources is estimated properly, the Q_{true} value should be approximately to Q_{except}. If the number of sources is not well determined, the *Q* value may deviate from the theoretical value

(Bressi et al., 2014; Baudic et al., 2016). However, the Q_{true} always deviates from the Q_{except} in many cases especially for large dataset (Liu et al., 2016, 2017; Shao et al., 2016). The variation of the Q values to the number of factors is shown in Fig. A2a and the correlation coefficients between O/P values in each factor number solution is shown in Fig. A2b.

As shown in Fig. A2a, Q_{true}/Q_{except} decreased substantially between 2-, 3- and 4-factor solutions, indicating that a substantial amount of the variability in the dataset was accounted for each additional factor; for P = 5, the Q_{ture}/Q_{except} exhibited the minimum value; as the factor number changed from 6 to 8, the Q_{ture}/Q_{except} value increased again. The Pearson correlation coefficients between the observed and predicted total VOC concentrations for different factor numbers are shown in Fig. A2b, which indicating that the total VOC concentrations were well reproduced by PMF model. In addition, for the 20 individual VOC species, the PMF model also well reproduced the predicted concentrations, with the r^2 ranged from 0.42 to 0.96 (Table A1). Therefore, we considered that the 5-factor solution was the optimum solution for this PMF analysis (Fig. A3).

A3 Bootstrap run (BS)

After choosing the 5-factor solution, the bootstrap (BS) method was used to detect and estimate disproportionate effects of a small set of observations on the solution and also, to lesser extent, effects of rotational ambiguity. BS datasets are constructed by randomly sampling blocks of observations from the original data set (US EPA, 2014). The base run with the lowest Q_{robust} is provided to map with each BS run in minimum Pearson correlation coefficient being 0.6. The number of BS is set as 100 to ensure the robustness of the statistics. In this study, the base and boot factors were matched except for factor 3 (combustion) and factor 5 (fuel evaporation) (Table A2). Mapping over 80% of the factors indicates that BS uncertainties can be interpreted and the number of factors may be appropriate. Seen from the Table A2, the BS results indicated a rotational ambiguity and F-peak should be further applied.

A4 BS-DISP error estimation

BS-DISP estimates the errors associated with both random and rotational ambiguity. A key file containing the number of cases accepted, largest decrease in Q, number of swaps in best fit and DISP was generated (Table A3). Swaps by factor was used to assess the error fraction. There were 99 bootstrap cases accepted and one resample was rejected. The decrease of Q was less than 1%, which indicated that the test of BS was validated and no more testing was required. It suggested that the solution was well constrained and the BS-DISP results can be reported.

Finally, the F-Peak values from –1 to 1 at 0.1 interval were used to remove the rotational ambiguity as discussed above. The F-peak bootstrap was also used to test the mapping between the base model and the F-peak runs. Results indicated that the F-peak = 0.2 was the optimal solutions with all factors mapping 100% and the base run of each species was within the inter quartile range (IQR) of the BS run.

Appendix B The calculation of PSCF and CWT

B1 PSCF

The PSCF values were calculated to explore the potential geographic origins of VOC sources using the source contributions apportioned from the PMF model and the backward-trajectory. The PSCF is defined as:

$$PSCF_{ij} = \frac{m_{ij}}{n_{ij}} \quad (4)$$

where i and j are the latitude and longitude, n_{ij} is the total number of endpoints that fall in the ij-th cell, and m_{ij} is defined as the number of endpoints in the same cell that exceeded the threshold criterion. The 75th percentile of each identified source contribution (i.e., 3.9 ppbv for oil refinery, 75.1 ppbv for NG, 15.3 ppbv for combustion, 1.2 ppbv for asphalt and 33.4 ppbv for fuel evaporation) was used as the criterion value. When each grid average number of trajectory endpoint (n_{ave}) is three times larger than the n_{ij}, the uncertainty of cell is reduced by multiplying a weight function (W_{ij}) into the PSCF value. The weight function is expressed as:

$$\mathrm{WPSCF} = \frac{m_{ij}}{n_{ij}} \times W(n_{ij}) \quad (5)$$

$$\mathrm{W}(n_{ij}) = \begin{cases} 1.00, & n_{ij} > 3n_{ave} \\ 0.70, & 3n_{ave} > n_{ij} > 1.5n_{ave} \\ 0.40, & 1.5n_{ave} > n_{ij} > n_{ave} \\ 0.20, & n_{ij} > n_{ave} \end{cases} \quad (6)$$

B2 CWT

Since PSCF value just gives the proportion of potential sources in a grid with struggles to distinguish the pollution levels of different potential regions, a concentration-weighted-trajectory (CWT) model was employed in this study. The geographical-domain was sliced into grid cell with a resolution of 0.5 º × 0.5 º. The CWT was calculated according to:

$$C_{\mathrm{ij}} = \frac{1}{\sum_{l=1}^{M} \tau_{ijl}} \sum_{l=1}^{M} c_l \tau_{ijl} \quad (7)$$

where C_{ij} represents the average weight concentrations in the grid cell (i, j), C_l is the measured VOCs concentration observed on the arrival of trajectory l, τ_{ijl} is the number of trajectory end points in the grid cell (i, j) associated with the C_l sample. The weighting function described in equation (6) was also used in the CWT analysis.

Appendix C Detailed calculating of local emission and regional transport contribution

C1 Choosing the radius to distinguish the local and regional area.

Grids within a given radius were excluded in case they could be too close to the origin. Usually, the radius was set according to the area covered by first 6 h of trajectories (Bari et al., 2003; Wang et al., 2015; Wang et al., 2016). In this study, we referred to it with modification. 12 h backward trajectories were set as the radius to distinguish the local and regional areas due to the following reasons:

(1) The duration of *m, p*-xylene decreasing from highest concentrations (at 02:00 LT) to lowest (14:00 LT) was 12 h (section 3.4). The atmospheric lifetime of *m, p*-xylene is about 11.8 h assuming the OH radical equals to 10^6 rad cm^{-3}. The compounds with atmospheric lifetime longer than *m, p*-xylene can be transported from long distance.

(2) The endpoint of each backward trajectories in the first 24 h was tested to find the optimum range of "local and nearby" area (Fig. C1). As the backward time increasing from 1 h to 24 h, the area covered by the long air mass increased significantly. Before the first 5 h, the air masses were mainly from the northwest and the east of the sampling site. From 7 h to 12 h, air masses from the northeast, southeast and southwest of the sampling site and the "shape of local area" formed. After 12h, the air masses, especial for the trajectories from the west transported for long distance and reached the sampling site, indicating the regional contributions.

It can be seen from Fig. C1 that the farthest endpoint of 12 h backward was about 7 degrees away from the sampling site. Therefore, the local and regional area was divided by a circle with the radius being 7 degrees (sampling site as the origin).

C2 Raster analysis

The CWT results obtained by TrajStat software were stored in shapefile format and then were introduced into the Arc GIS software (10.1, Esri, US). The first step was to remove the negative value from the shapefile and then convert the shapefile into raster format. The local and regional area was extracted by a circle with radius being 7 degrees as discussed above. The inner of the circle was defined as the local while the external area was set as the regional transport and an example is shown in Fig. C2. The statistics (count, minimum, maximum, sum, mean and standard deviation) of the extracted raster were shown in its layer properties in Arc Map. The statistics of each VOC sources in each season are summarized in Table C1. The percentage contributions of local source and regional transport of the five VOC sources in different seasons were calculated according to Eq. (2) in section 2.4.3.

Table A1. Pearson coefficients between the observed and predicted VOCs concentrations for 5-factor solution

Species	Intercept	Slope	SE [a]	r^2
ethane	1.09	0.90	7.69	0.96
propane	0.25	0.98	5.96	0.94
i-butane	0.42	0.86	4.27	0.88
n-butane	0.76	0.82	6.06	0.84
i-pentane	2.50	0.71	7.57	0.50
n-pentane	0.69	0.75	4.20	0.77
n-hexane	0.24	0.84	1.05	0.91
cyclohexane	0.11	0.82	0.37	0.92
methylcyclohexane	0.14	0.81	0.43	0.93
n-octane	0.06	0.84	0.21	0.93
n-nonane	0.04	0.71	0.10	0.85
n-decane	0.05	0.68	0.08	0.79
n-undecane	0.04	0.70	0.09	0.71
n-dodecane	0.17	0.27	0.29	0.49
ethylene	0.25	0.50	0.75	0.53
acetylene	0.57	0.15	0.93	0.42
benzene	0.36	0.33	0.56	0.42
toluene	0.32	0.24	0.48	0.37
m, *p*-xylene	0.14	0.48	0.32	0.57
o-xylene	0.05	0.36	0.07	0.55

[a] SE = standard error

Table A2. Mapping of bootstrap factors to base factors

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Unmapped
Boot Factor 1	100	0	0	0	0	0
Boot Factor 2	0	100	0	0	0	0
Boot Factor 3	2	3	95	0	0	0
Boot Factor 4	0	0	0	100	0	0
Boot Factor 5	14	20	0	0	66	0

Table A3. BS-DISP Diagnostics.

# of Cases Accepted:	99				
% of Cases Accepted:	99%				
Largest Decrease in Q:	–14.10				
%d*Q*:	–0.014				
# of Decreases in *Q*:	1				
# of Swaps in Best Fit:	0				
# of Swaps in DISP:	0				
Swaps by Factor:	0	0	0	0	0

Table C1. The statistics of VOCs sources in each season obtained by raster analysis

		Oil refinery		NG		Combustion		Asphalt		Fuel evaporation	
		Local	Regional	Local	Region	Local	Region	Local	Region	Local	Region
Autumn	Minimum	0.016	0.00055	0.083	0.005	0.021	0.019	0.022	0.0032	0.012	0.0027
	Mean	2.68	0.48	41.15	4.63	7.6	1.16	0.98	0.17	17.43	2.9
	Sum	816	449	12758	4347	2356	1098	305	160	5315	2468
	Number	304	928	310	939	310	943	310	944	305	851
Winter	Minimum	0.017	0.0007	0.62	0.089	0.092	0.092	0.0052	0.00043	0.13	0.0025
	Mean	1.41	0.32	36.1	7.82	12.34	2.93	0.35	0.076	18.51	4.21
	Sum	492	319	12707	8411	4333	3147	121	79	6406	4421
	Number	348	1010	352	1076	351	1073	345	1036	346	1049
Spring	Minimum	0.0027	0.000094	0.11	0.064	0.48	0.033	0.0066	0.0041	0.002	0.0015
	Mean	1.05	0.19	16.71	2.86	2.6	0.5	0.49	0.093	7.66	1.91
	Sum	360	180	6068	3175	945	553	177	103	2764	1840
	Number	343	954	363	1112	363	1111	363	1109	361	964
Summer	Minimum	0.00044	0.0011	0.036	0.021	0.01	0.0021	0.0082	0.00022	0.0097	0.00028
	Mean	1.92	0.29	19.57	2.6	2.63	0.4	0.51	0.1	9.98	1.87
	Sum	556	242	5714	2250	776	343	149	89	2923	1513
	Number	289	845	292	865	295	863	293	876	293	808
Annual	Minimum	0.0036	0.000094	0.0356	0.011	0.0645	0.0019	0.0023	0.00043	0.0036	0.0015
	Mean	2.16	0.41	35.26	6.52	8.08	1.87	0.73	0.16	16.44	3.8
	Sum	953	821	15973	13338	3653	3805	328	330	7399	7225
	Number	442	1959	453	2047	452	2035	450	2014	450	1899

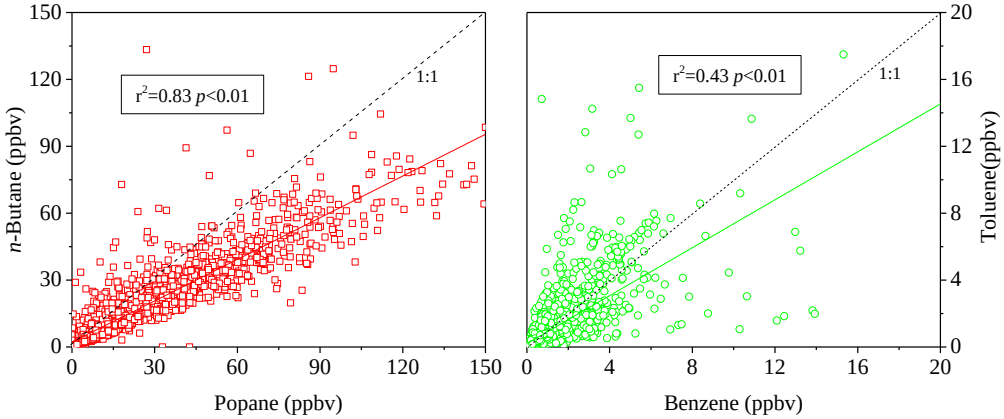

Figure A1. Samples of highly collinear species.

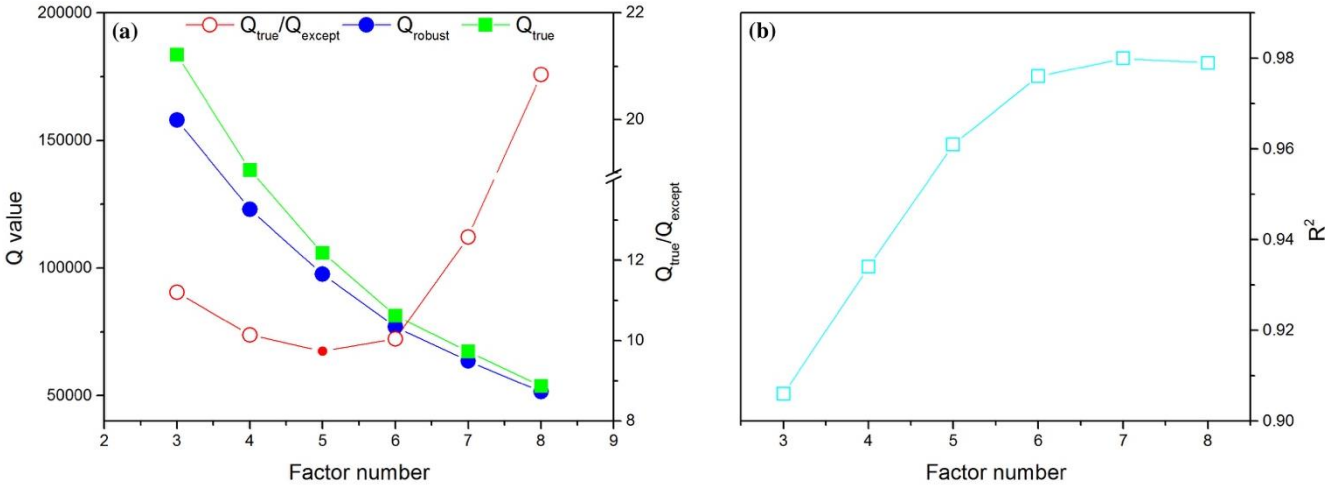

Figure A2. Q_{true}/Q_{except}, Q_{robust} and Q_{true} plotted against the number of factors used in the positive matrix factorization (PMF) solution.

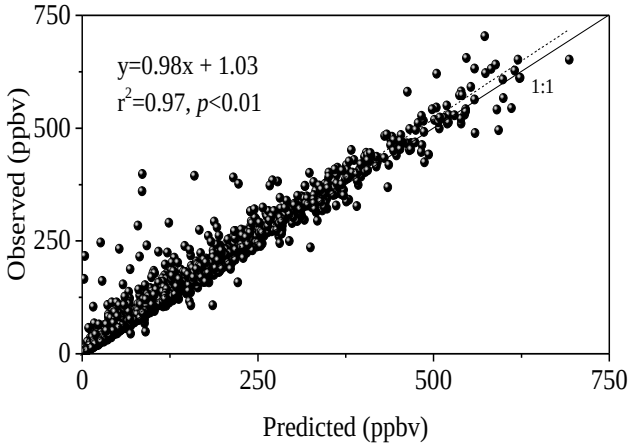

Figure A3. Scatter plots between the total predicted and observed VOC concentrations based on the 5-factor PMF solution.

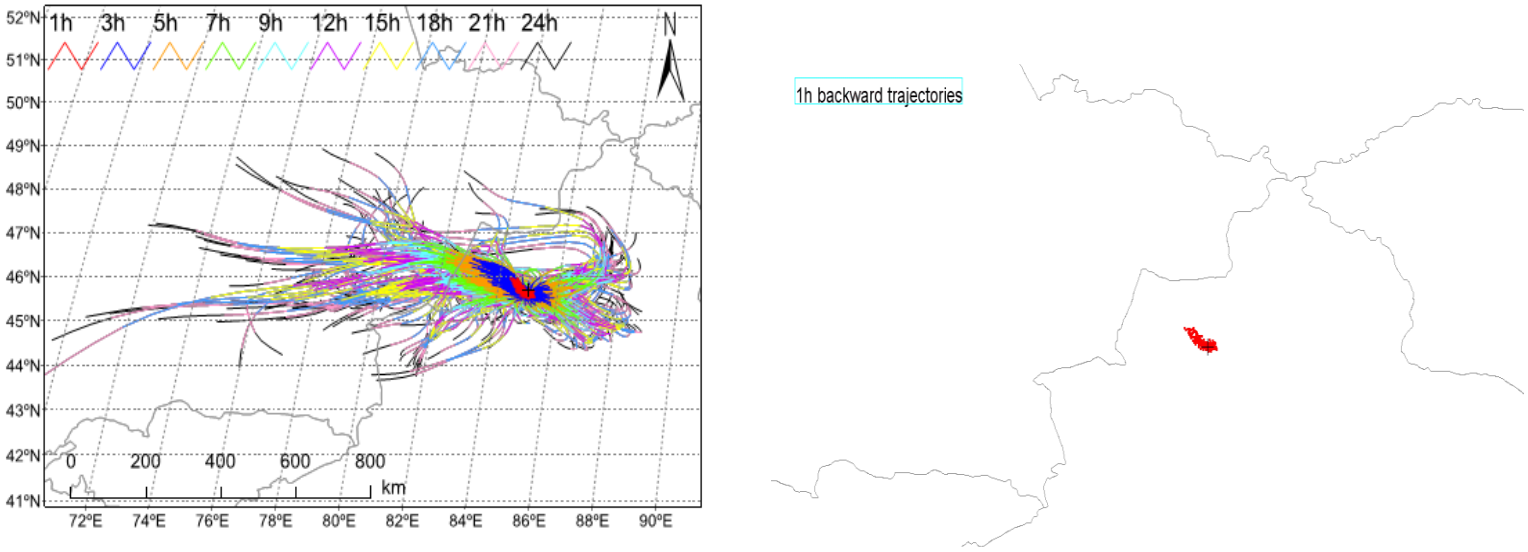

Figure C1. The region covered by the first 24 h backward trajectories. Left: the domain covered by the trajectories. Right: the dynamic variation of backward trajectories from 1h to 24h. (Double click to view the GIF)

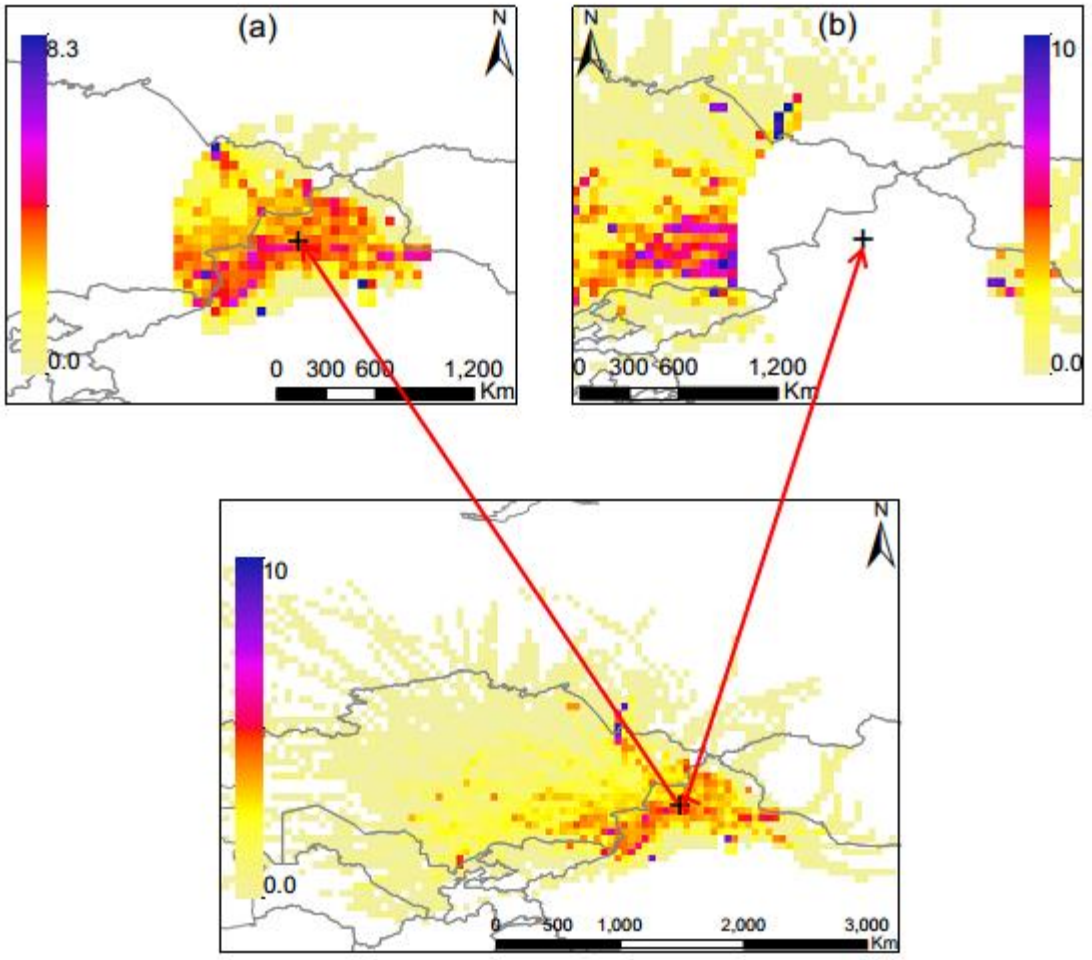

Figure C2. Example of (a) local area and (b) domain of regional transport range extracted by a circle with the radius being seven degrees. The black cross represents the sampling point and is set as the original point.