Atmospheric Environment 279 (2022) 119114

Contents lists available at ScienceDirect

Atmospheric Environment

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

Analysis of aerosol particle number size distribution and source attribution at three megacities in China

Dongmei Zhang [a], Zhanqing Li [b,*], Hao Wu [c], Tong Wu [a], Rongmin Ren [a], Zhaoxin Cai [a], Chen Liang [a], Lu Chen [a]

[a] College of Global Change and Earth System Science, Beijing Normal University, Beijing, 100875, China
[b] Department of Atmospheric and Oceanic Science, University of Maryland, College Park, MD, 20742, USA
[c] School of Electrical Engineering, Chengdu University of Information Technology, Chengdu, 610225, China

HIGHLIGHTS

- Particle sources were identified and quantified in three Chinese megacities by K-means clustering.
- Frequent Nocturnal Nucleation Events have been observed in Beijing and Guangzhou.
- Traffic emissions were the main contributor in Beijing and Guangzhou.

ARTICLE INFO

Keywords:
Particle number size distributions
Ultrafine particles
k-means clustering
Traffic emissions

ABSTRACT

Air pollution has been a major concern in China, especially in densely populated megacities. Particulate pollutants, generally known as aerosols, have sizes that range widely and complex physical and chemical compositions with high spatiotemporal variabilities. We have conducted a series of experiments using the same comprehensive observation system, measuring various aerosol properties in the three megacities of Beijing, Shanghai, and Guangzhou. Although there are differences in the total particle number concentration and particle number size distribution among the three sites due to differences in emission sources and sampling seasons, a high proportion of ultrafine particles was observed in all cities, i.e., 86.33%, 88.34%, and 81.1%, respectively. The high particle number concentration at the Beijing and Guangzhou sites was caused by an increase in nucleation-mode particles at nighttime. The concentration of nucleation-mode particles at nighttime was significantly higher than during the daytime. We call this phenomenon a nocturnal nucleation event, and combined with analyses of factors involved, attribute this event to traffic emissions. To investigate the pollution sources at each location, we identified and quantified the pollution contributions from different sources by applying the k-means clustering analysis to observational datasets of environmental quantities. Clustering results showed that traffic emissions were the main contributor in Beijing and Guangzhou, with direct and significant impacts ranging from 37% to 45%, respectively.

1. Introduction

Air pollution has become a major risk to public health, with about seven million people dying prematurely each year from exposure to high concentrations of atmospheric particulate matter (WHO, 2018). Small aerosol particles are more readily inhaled into lungs, spreading to other body parts and causing a serious health concern (Kumar et al., 2014; R. Zhang et al., 2015; Perera, 2017; Garate et al., 2020). The particle number concentration (PNC) in an urban atmospheric environment is mostly dominated by ultrafine particles (UFP) with sizes less than 100 nm (Cheung et al., 2013; Agudelo et al., 2019). This indicates the necessity of studying atmospheric fine particles, especially UFP, and the importance of clarifying the source of UFP and minimizing the number concentration of atmospheric UFP.

Megacities in developing countries such as China and India have complex mixtures of air pollutants and high atmospheric oxidation capacity associated with such intensive human activities as coal burning, vehicle exhaustion, and industrial activities resulting from rapid

* Corresponding author.
E-mail address: zhanqing@umd.edu (Z. Li).

https://doi.org/10.1016/j.atmosenv.2022.119114
Received 15 January 2022; Received in revised form 9 April 2022; Accepted 11 April 2022
Available online 20 April 2022

economic development and urbanization over the past several decades (Z. Li et al., 2019; Y. Wang et al., 2019; Q. Zhang et al., 2021). Aerosols emitted from these activities are usually characterized by high number concentrations of UFP, complex chemical compositions, and strong temporal and spatial variabilities (Brines et al., 2015; Birmili et al., 2015; Chu et al., 2021; Q. Zhang et al., 2021). Road traffic emissions are the main source of UFP in urban areas, contributing 71–94% to the total PNC (Rönkkö et al., 2017; Rivas et al., 2020; Wu et al., 2021). In addition to a large number of primary soot particles with an average particle size of about 30–100 nm, unexpected nucleation and growth of nucleation-mode particles was found in automobile exhaust gas (Brines et al., 2015; Agudelo et al., 2019). Because automobile exhaust gas also contains a large number of gaseous components, such as volatile organic compounds and sulfuric acid, they will immediately condense or nucleate due to the change in saturation during the dilution and cooling process of hot exhaust gas (Olin et al., 2020; S. Guo et al., 2020). In addition to road traffic contributions, new particle formation (NPF) also contributes to UFP in urban atmospheric environments (R. Zhang et al., 2012; S. Guo et al., 2014; Kulmala et al., 2016). Although high concentrations of preexisting particles in megacities may inhibit homogeneous nucleation and NPF, NPF frequently occurs in megacities (Kanawade et al., 2020). It not only strongly contributes to the number concentration of particles but also has an important influence on the formation of winter haze and particle mass concentrations in northern China (F. Zhang et al., 2019; R. Zhang et al., 2015; Sun et al., 2016).

In this study, comprehensive field observation experiments were carried out in Beijing (BJ), Guangzhou (GZ), and Shanghai (SH), located in the hearts of the three urban agglomerations of China's North China Plain (NCP), Pearl River Delta (PRD), and Yangtze River Delta (YRD), respectively. The purpose of this work is to observe the particle number size distribution (PNSD) and PNC of three densely populated and economically developed megacities in China using the same set of instruments and to use the same methodologies to determine aerosol processes and possible sources related to the emission, formation, and transformation of UFP in megacities. The size distribution of aerosols is one of the most important properties of particles, contributing to our understanding of aerosol dynamics and providing information about particle sources and atmospheric processes (Vu et al., 2015; Liang et al., 2020). We applied cluster analysis techniques to the PNSD datasets of the three megacities to quantify the contribution of different sources to PNC in urban atmospheric environments.

2. Sampling and methods

2.1. Sampling sites

There are three major urban agglomerations in China, namely the NCP in northern China, the YRD in eastern China, and the PRD in southern China. Their rapid economic development and dense population made them the largest sources of air pollution in China. We chose BJ (in the NCP region), SH (in the YRD region), and GZ (in the PRD region) as our research sites (Table 1), which are megacities with permanent populations exceeding 10 million. The three megacities have large differences in weather, climate, geographical location, and industrial structure, all contributing to significant differences in air quality.

2.2. Instrumentation and quality assurance

Field experiments aimed at acquiring aerosol physical and chemical properties together with general meteorology and ambient conditions were carried out using the same observation system in the three megacities (Z. Li et al., 2019). Particle number concentrations and size distributions were monitored using an assembled Scanning Mobility Particle Sizer (SMPS, TSI 3938) system. The aerosol size distribution describes the distribution of aerosol number concentrations in different particle-size bins, an essential physical characteristic of particulate matter. The system first uses a differential mobility analyzer (DMA) to classify particles according to their migration characteristics in the electric field. A condensation particle counter (CPC, TSI 3772) is then used to measure number concentrations in different size bins. The sample intake flow rate was set at 1.0 L/min. The temporal resolution of the measured aerosol size distribution was 5 min, and the particle size range was 11.3–552.3 nm. Corrections of the particle size spectra include the multiple charge correction (treating large particles with multiple charges as smaller particles) and the diffusion loss correction (correcting for the loss of particles $<$100 nm in diameter through diffusion at the inner wall of the pipeline).

By integrating the size distribution of particulate matter, the number concentration values of different modes of particulate matter can be obtained, providing more insightful information. In this study, aerosols are divided into four categories according to the size of the particle: nucleation mode (NU, diameters less than 25 nm), small-Aitken mode (SA, diameters from 25 nm to 50 nm), large-Aitken mode (LA, diameters from 50 nm to 100 nm), and accumulation mode (AC, diameters greater than 100 nm).

In addition, the system was equipped with a vorticity covariance LI-COR 7500A analyzer and an aerosol chemical speciation monitor (ACSM) of the Capture Vaporizer (CV) to obtain meteorological parameters and the aerosol mass and chemical composition of non-refractory submicron particulate matter in the atmospheric environment in real time. The ACSM is used for online observations of chemical components of aerosol particles, including organic matter (Org), nitrate (NO_3^-), sulfate (SO_4^{2-}), ammonium salt (NH_4^+), and chloride (Chl), with a data resolution of 15 min. The vorticity covariance LI-COR 7500A system can measure the thermodynamic parameters of air pressure, air temperature, water vapor density, and three-dimensional wind vector. Particulate matter with diameters less than 2.5 μm ($PM_{2.5}$), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3) concentrations used in this study were obtained from the nearest EPA observatory operated by the National Environmental Monitoring Center of China (last access date: October 30, 2021, URL: http://www.cnemc.cn/).

2.3. Clustering particle number size distribution data

Cluster analysis was applied to group objects according to similarities in data feature vectors (Jain et al., 1999), a technique widely used in many studies of air pollution, source identification, long-distance

Table 1
Information about the monitoring stations selected for the study in the three megacities.

Megacity	Coordinates	Geographical region	Site description	Observation period
BJ	39.82°N,116.48°E	NCP	The site is located in the meteorological bureau compound in the southern suburbs of Beijing, surrounded by three roads, including the Southern Fifth Ring Road to the south. It is mainly affected by surrounding traffic emissions. It is an urban traffic emission site.	2019/01/01–2019/04/27
GZ	23.01°N,113.33°E	PRD	Located in Panyu District, south of Guangzhou City. Sources of pollution are complex, including road, catering, and industrial emission sources. It is an urban multi-source compound pollution site.	2019/11/04–2020/01/20
SH	31.39°N,121.45°E	YRD	This station is located in the Shanghai Baoshan Meteorological Bureau, far away from the city center, with less traffic flow on surrounding roads and no obvious industrial park. It is a city background site.	2020/04/05–2020/06/05

transport of pollutants, and the formulation of effective control and mitigation strategies (Salimi et al., 2014; Brines et al., 2015; Agudelo et al., 2019; Govender and Sivakumar, 2020; Chen et al., 2021). Typically, the goal of cluster analysis is to identify groups of similar objects, with objects in the same cluster grouping more similar to each other than in different clusters. It was first proposed in 1930, became popular in the 1960s, and was widely used in atmospheric sciences in the 1980s (Gong and Richman, 1995; Govender and Sivakumar, 2020). Given the amount of data to analyze, k-means clustering was applied to the three megacity observational datasets, which include simultaneous measurements of PNSD, major aerosol chemical components (SO_4^{2-}, NO_3^-, and Org), gaseous pollutants (SO_2, NO_2, CO, and O_3), and meteorological elements (temperature, humidity, and wind speed). The reason for choosing k-means clustering technology is that it can find clusters with the smallest size, the farthest interval, and highest degree of internal cluster similarity compared with other clustering technologies, such as fuzzy, k-median, and model-based clustering (Beddows et al., 2009). K-means is an iterative algorithm that minimizes the sum of square distances between all points and clustering centers to find a given number of clusters (Salimi et al., 2014). In other words, k-means clustering defines a homogeneous set by minimizing the clustering error, defined as the square of the Euclidean distance between each dataset point and the corresponding clustering center (Agudelo et al., 2019).

Determining the appropriate number of clusters is an additional challenge. The ratio between the minimum inter-cluster distance and the maximum intra-cluster distance, i.e., the Dunn index, is used to select the optimal number of clusters (Beddows et al., 2009; Wegner et al., 2012). Although there currently exist standards to help researchers determine the number of clusters representing different sources before conducting a cluster analysis, the optimal number of clusters must be selected in combination with the data structure and sampling background conditions to determine whether the most meaningful and representative number of clusters will be generated (Govender and Sivakumar, 2020). In view of the large difference in PNSD between a nocturnal nucleation cluster and clusters of other emission sources in BJ and GZ, repeated clustering results showed that other datasets without nocturnal nucleation cluster could not be well clustered. We thus chose re-clustering. First, nocturnal nucleation clusters were obtained by the clustering technique, then the datasets excluding nocturnal nucleation clusters were clustered for the second time. The origin of each cluster was explained by calculating the mean concentrations of PNC and other pollutants and the mean values of meteorological parameters, as well as their diurnal patterns and frequencies of occurrence.

3. Results and discussion

3.1. Characteristics of aerosol particle size distributions in the three megacities

Fig. 1 shows average diurnal variations of PNSD and mean particle diameter (mean D_p) at the three observational sites in China. The average particle number concentrations (PNCs) in BJ, GZ and SH were 24,868 cm^{-3}, 24,120 cm^{-3} and 7,540 cm^{-3}, respectively. Although the total number of particles at the three sites is different due to different driving factors such as sampling season and emission sources and meteorological conditions at the sampling sites, a very high proportion of UFP is present at all three sites. The proportion of particles consisting of UFP in BJ, GZ, and SH was 86.33%, 88.34%, and 81.1%, respectively. The PNSD and mean D_P at BJ and GZ had similar diurnal variations. The concentration of nucleation-mode particles were significantly higher at night than in the daytime. Accordingly, the overall mean D_P at night was lower than in the daytime. We call this nighttime phenomenon in BJ and GZ a nocturnal nucleation event (NNE), characterized by a continuous nucleation-mode particle burst, resulting in the reduction of average particle diameters. NNE is different from the photochemically driven NPF, characterized by the conversion of gaseous precursors to particles in the atmosphere due to nucleation, followed by the growth process through physicochemical interactions. Therefore, there is no banana-shaped increase in PNSD for an NNE. Nucleation is defined by the fact that during the occurrence of an NNE, nucleation-mode particles continue to explode at a high concentration level for several hours. This is like having an unknown source continuously release high levels of nucleation-mode particles into the atmospheric environment. This source may be one of several sources, such as increased emissions of primary particles, enhanced secondary nucleation from the gas phase to particles, a reduced boundary layer, or a combination of sources. The occurrence of NNEs in BJ and GZ will be explained in later sections. The lowest value of mean D_P at the BJ site (45 nm) appeared at ~06:00 local time (LT), then gradually increased, ultimately remaining at a high

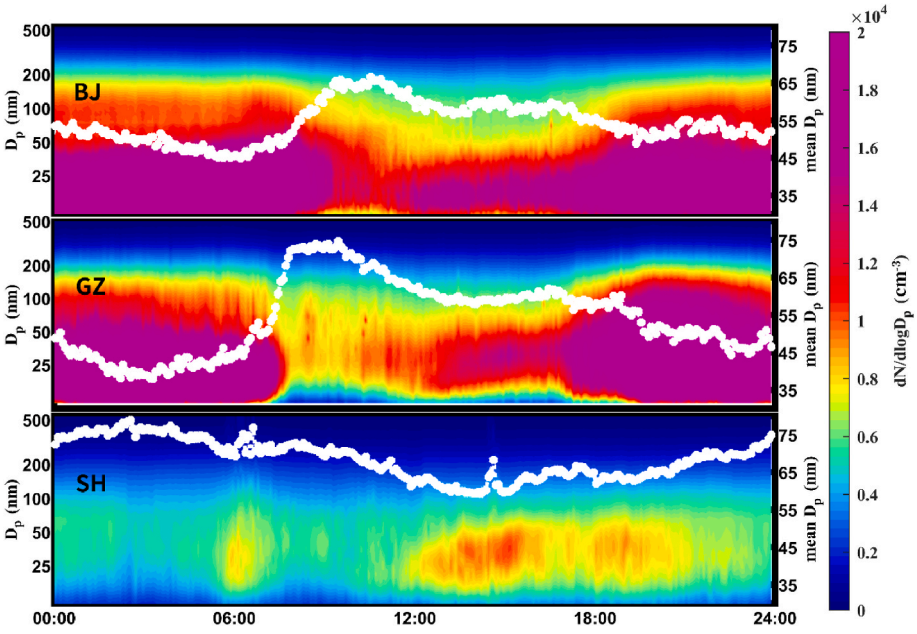

Fig. 1. Average diurnal variations of the PNSD (unit: cm^{-3}) and mean D_p (unit: nm, white dots) at the sites in BJ, GZ, and SH.

value of ~65 nm between 09:00 and 10:00 LT. The mean D_P at the GZ site remained small between 02:00 and 04:00 LT (~38 nm), then rose in the early morning, reaching its maximum value between 07:30 and 09:30 LT (74 nm). The mean D_P increased significantly in the early morning, indicating the modal evolution process of particles from small particles to larger particle sizes. The particle concentration at the SH site was significantly lower than at the BJ and GZ sites, with little difference in particle concentration between night and day.

Similar to the mean D_p, there were significant differences in the mean PNSD among the three cities, with PNSDs differing between day and night (Fig. 2). In general, changes in PNSD are driven by the dilution process and the transformation process (including nucleation, condensation, coagulation, and deposition, among others) in the atmosphere, as well as changes in emissions (Ketzel and Berkowicz, 2005; Lorelei de Jesus et al., 2020). Peak number concentrations at the BJ and GZ sites at night were higher than in the daytime. Peaks at night reached 35,000 cm^{-3} in BJ at $D_P = 18.8$ nm and 15,000 cm^{-3} in GZ at $D_P = 18.1$ nm. The former is much higher than reported by Du et al. (2017) in BJ in 2015. Note that aside from synoptically driven changes and seasonal differences, our site is located in the very polluted southern part of BJ with denser road traffic, while the site chosen by Du et al. (2017) was in the cleaner northern part of BJ. A second peak occurred in BJ at ~70 nm, while the distributions were unimodal in SH and GZ. The peak in GZ appeared between 25 and 30 nm and was slightly larger than that in BJ, suggesting more aging of particulate matter in GZ. This may be related to the influence of surrounding cooking emissions from food stalls in the Chimelong Paradise amusement park in GZ. In SH, the peak of the PNSD occurred during the day at ~30 nm.

Fig. 3 compares the mean number concentration for each mode and the mean mass concentration for each chemical component of particles during the daytime and nighttime. As mentioned before, mean total PNCs in BJ and GZ were significantly different during the daytime and nighttime, mainly driven by particles in the nucleation mode. The average concentration of nucleation-mode particles in BJ at nighttime was about twice as high as that in the daytime, increasing from 7117 cm^{-3} to 14,199 cm^{-3}, accompanied by an increase in the proportion of total particle numbers from 39.26% to 46.23%. The proportion of nucleation-mode particles in GZ increased from 28.36% during the daytime to 43.84% at night and increased the mean PNC by 4.4 times. The increase in nucleation-mode PNC at night may be due to a change in the emission source structure or a decrease in the boundary layer. However, the increase in the proportion of nucleation-mode particles suggests that local emission source is the main influence. This rules out the possibility of boundary layer factors alone, which can only increase particle concentrations and not have a significant effect on the proportion of newly emitted nucleation-mode particles. From the changes in chemical components, only the concentration of Org components at the GZ site increased significantly at night, from 21.26 $\mu g/m^3$ to 27.6 $\mu g/m^3$. The concentrations of other chemical components did not change much at night in GZ. As the proportion of Org compounds increased (up to 62.47%), the proportion of other components at night decreased. Overall, the Org at the GZ site was the highest among the three urban sites during both daytime and nighttime, indicating the contribution of cooking emissions. As for the BJ site, in addition to the significant increase in Org, there was also a minor increase in SO_4^{2-} at night.

3.2. Dependency on wind speed and direction

Wind transport plays an important role in air quality. Figs. 4 and 5 show the relationships between the four PNC modes and wind conditions at the BJ, GZ, and SH sites. Fig. 4 shows how the air quality of the urban sites changed with wind direction during the course of a 24-h period, while Fig. 5 reveals how it changed with both wind speed and direction during daytime and nighttime for the four aerosol modes at the three urban sites. They are helpful to understand and explain differences between day and night and identify sources of pollutants. The rose plot in Fig. S1 shows which wind direction contributed most to overall mean concentrations.

A higher PNC occurred during nighttime in BJ, associated with northeast and southwest winds under weak wind speed conditions (<5 m/s), presumably caused by local emissions. For small particles in the nucleation and Aitken modes, high concentrations appeared predominantly between 03:00 and 06:00 LT under northeasterly wind conditions. Particles emitted from diesel engines fall mainly in the range of 20–130 nm, while that of gasoline is in the range of 40–80 nm (Agudelo et al., 2019). During daytime with strong northwest winds, nucleation- and small-Aitken-mode particles dominate over accumulation-mode particles. For nucleation-mode particles, the mean number concentration at nighttime was nearly twice that in the daytime, mainly from the northeast, with a contribution of nearly 30%, followed by south and southwest winds. The contribution of small-Aitken-mode particles was similar to that of the nucleation mode, indicating that they may have the same source of pollution. Strong winds blown in from a cleaner background cleared out preexisting particles, creating favorable conditions for the formation of new particles.

In GZ, heavy pollution also occurred during nighttime, typical for most urban environments (e.g., von Bismarck et al., 2013; Weber et al., 2013; Hama et al., 2017). Note that there are two independent local

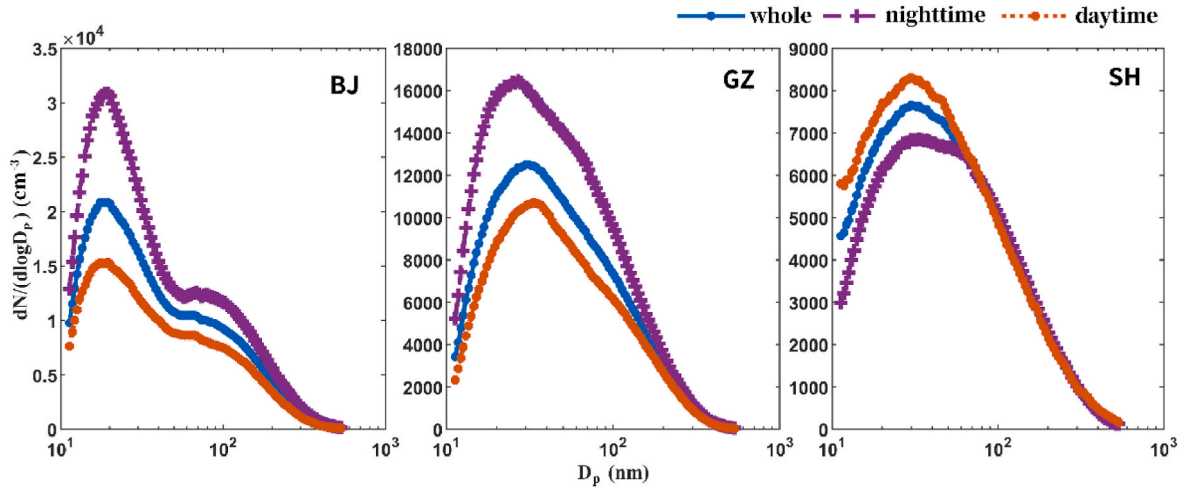

Fig. 2. Average particle number size distributions at the sites in BJ, GZ, and SH during the whole day (blue curves), at nighttime (purple curves), and during the daytime (orange curves). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

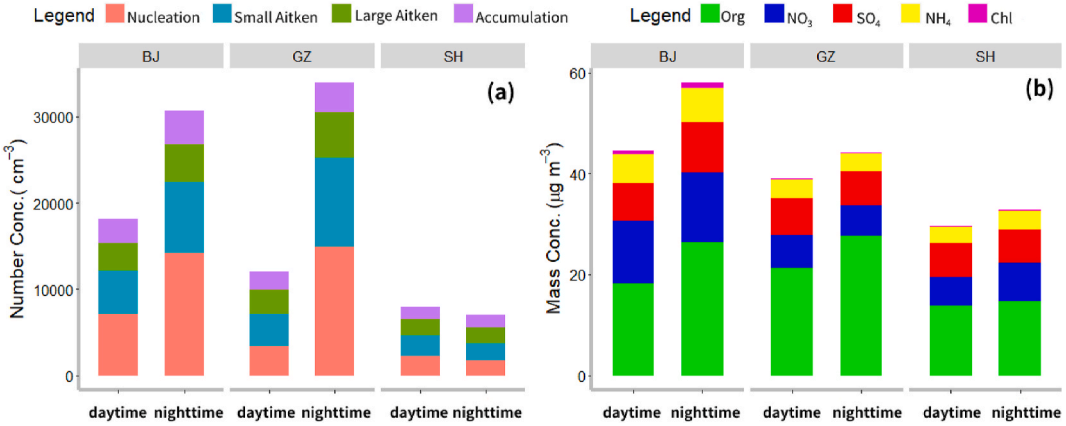

Fig. 3. (a) Average daytime and nighttime particle number concentrations divided by mode and (b) average particle mass concentrations for each urban site divided by chemical composition.

emission sources located in opposite directions, i.e., in the southwest and the northeast (Fig. 5). They correspond to the Xinguang Expressway and an industrial production area with a night market and barbecues set up on the street. The mean number concentration of nucleation and large-Aitken-mode particle at nighttime was nearly four times that of the daytime.

Compared to BJ and GZ, there are major local emission sources in SH. The observation site is far away from the city center, and there are no industrial parks and busy roads around it. When northwest winds blew in from a clean background at speeds greater than 5 m/s, the concentration of nucleation-mode particles increased significantly, which may be related to NPF, whose further growth may have led to high concentrations of small-Aitken- and large-Aitken-mode particles as well. When winds came from the northeast, the concentration of large-Aitken and accumulation-mode particles increased significantly, likely related to the Wusongkou International Cruise Terminal located 6 km away.

3.3. Clustering PNSD data using k-means clustering

Through continuous optimization, the aerosol datasets of BJ, GZ, and SH were divided into 6, 7, and 4 representative clusters, respectively, by selecting the best spectral number. Fig. 6 shows the average PNSDs of dN/dlog (D_p) of each cluster in the three megacities. The potential origins of these particular clusters are based on the mean values of gaseous pollutants, chemical composition, and meteorological variables during their occurrences. The mean values of relevant variables are shown in Fig. 7. The mean diurnal variation of PNSDs for each cluster of three megacities sites are shown in Fig. S2. Table 2 summarizes the frequencies of the clusters and their associated sources of pollutants.

3.3.1. Sources in BJ

Six clusters were obtained at the BJ site. Cluster 1, which only appeared at night, was an NNE cluster with the lowest mean D_p among the six clusters (45.35 nm). It had the highest PNC value of 57,755 cm^{-3} at 19.5 nm, of which UFP accounted for 90.7%. Accounting for 14.71% of the total observed PNSD data, low wind speeds, high humidity, and high traffic-related gas concentrations (NO_2 and CO) characterized the occurrence period of this cluster. The fresh traffic cluster (Cluster 2), with a frequency of 7.06%, occurred at night and during daytime rush hours. The peak particle size was 26.9 nm, and UFP contributed 92.9%, accounting for the highest proportion of particles. Cluster 3 was severe pollution caused by fossil fuel or biomass combustion, accounting for 15.21% of the total, with the highest average mass concentration of NO_3^- (34.12 μg/m^3), SO_4^{2-} (27.05 μg/m^3), and Org (59.38 μg/m^3) among the six clusters. This is comparable to concentrations measured during winter haze pollution events in northern Chinese cities (Zhao et al., 2017; Y. Zhang et al., 2018; S. Li et al., 2020). Average concentrations of SO_2, NO_2, and CO were also the highest, indicating that Cluster 3 was affected by the combustion of fuels such as coal and oil (Chen et al., 2021). Cluster 4 was associated with photochemically driven nucleation of NPF, accounting for 22.85% of the total PNSD. During the occurrence of this cluster, the mass concentration of each chemical component was at its lowest value, accompanied by high O_3 concentrations and low air humidity. The particle size distribution reached its peak at 17.5 nm, with a PNC of 15,015 cm^{-3}. Cluster 5, with a ratio of 20.87%, had the largest mean D_p value of 75.74 nm, presenting a bimodal distribution located in the nucleation mode and the accumulation mode. Particles in this cluster were classified as aerosols transported over long distances consisting mainly of accumulation-mode particles. Smaller particles in this cluster, such as nucleation- and Aitken-mode particles, are produced by coagulation and condensation during transport (Vu et al., 2015). Cluster 6 was a clean background cluster, accounting for 19.3%. High wind speeds removed many particles from the atmosphere, and the PNC concentration was the lowest of the six clusters at 5367 cm^{-3}.

3.3.2. Sources in GZ

The GZ site has seven pollution contributors, making it the most complex atmospheric environment among the three megacities. Six of the clusters were the same as in BJ, while the seventh (Cluster 5, aging aerosols) was unique to the GZ site. The first four clusters are NNE (Cluster 1), fresh traffic emissions (Cluster 2), fossil fuel or biomass burning (Cluster 3), and NPF (Cluster 4). Cluster 1 accounted for 17.17%, and the PNC reached a high value of 78,633 cm^{-3}, 38.14% higher than the mean value of the PNC for the NNE cluster in BJ. The mean D_p was 44.59 nm, with a peak value of 20.2 nm. The proportion of the fresh traffic cluster (Cluster 2) was 17.78%, significantly higher than the corresponding cluster in BJ. The mean D_p was 58.31 nm, also significantly higher than at the BJ site. This may be due to the influence of cooking emissions, whose main mode diameter is 40–50 nm, similar to the main mode particles from traffic emissions (Buonanno et al., 2009). Many barbecue catering sites were distributed within 1 km of the observation site in GZ. The high CO concentration of 0.96 mg/m^3 also confirmed the effect of cooking emissions. Emissions from fossil fuel or biomass combustion (Cluster 3) accounted for 9.9%, and the mean D_p was the largest among the seven clusters at 78.31 nm. Accordingly, the mass concentrations of SO_4^{2-}, NO_3^-, and Org during the occurrence period of this cluster were significantly higher than the mean values of the other six clusters but significantly lower than the mean value of a similar cluster in BJ. This is because the cluster in BJ included the cold winter haze weather. The proportion of the NPF cluster (Cluster 4) in GZ was 6.45%, much lower than in BJ. The O_3 concentration was as high as 101.03 μg/m^3, and its PNC and mean D_p were similar to those at the BJ

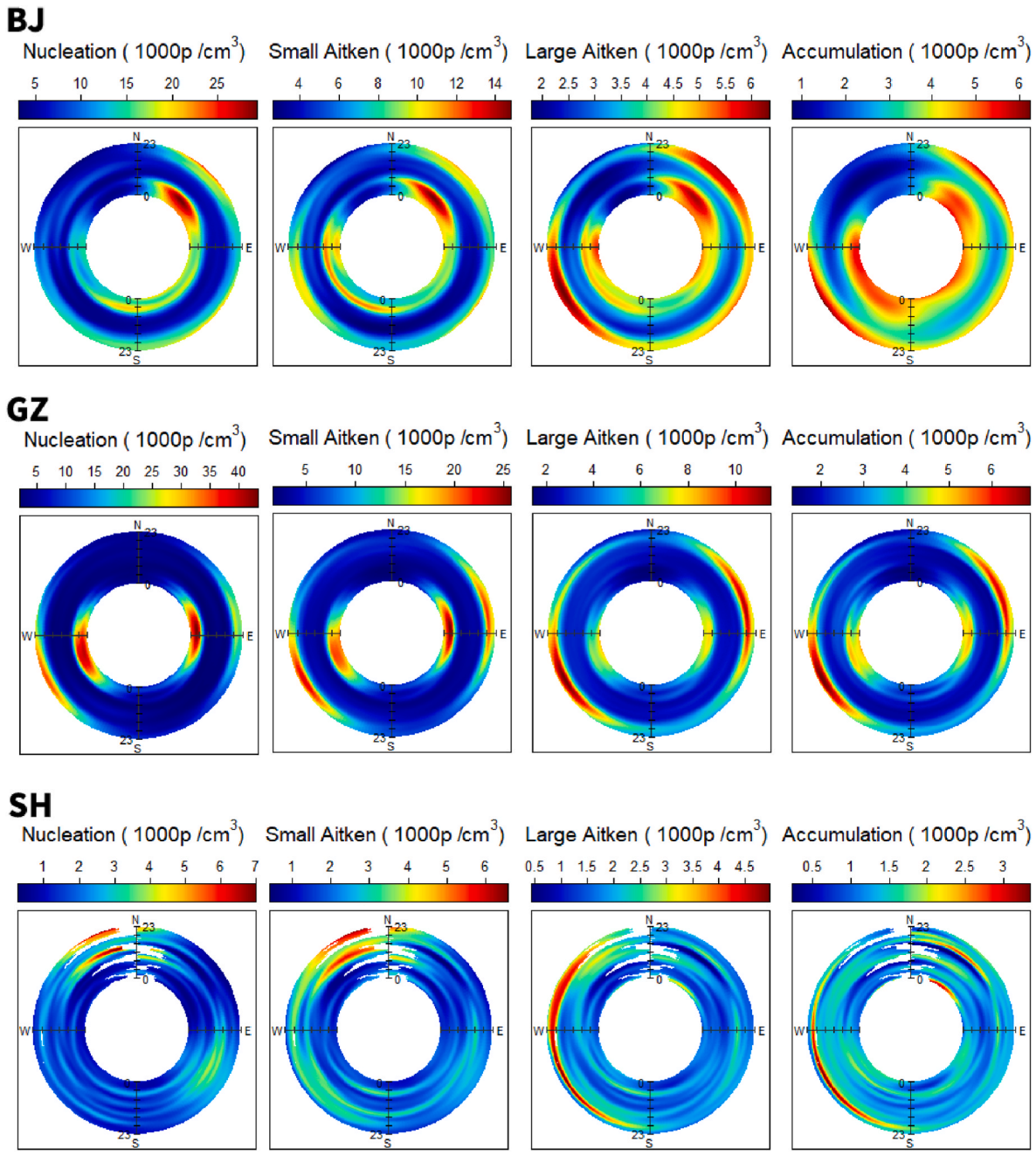

Fig. 4. Diurnal variations in particle number concentration (unit: 1000 particles/cm^3) in each mode according to wind direction at the BJ, GZ, and SH sites. Time intervals are shown inside each circle.

site. Considering that the average particle size distribution and PNC value of the fresh traffic cluster at the GZ site were close to that of the NPF cluster, some NPF at the growth stage may have made its way into the fresh traffic cluster during clustering, resulting in a low proportion of the NPF cluster and a high proportion of the fresh traffic cluster at GZ. However, the particle concentration of traffic emissions at the BJ site was significantly higher than that of NPF, so the significant difference in particle size distribution made the clustering result better. Cluster 5 was the aging aerosol cluster, accounting for 17.49%, representing the further growth stages of the fresh traffic and NPF clusters. Chen et al. (2021) have reported a similar cluster. Clusters 6 and 7 were long-distance regionally transported and clean background clusters, accounting for 14.73% and 16.48%, respectively.

3.3.3. *Sources in SH*

The emission sources of the SH site were relatively simple, with only four clusters that did not include the NNE cluster. At this site, the fresh traffic emission cluster and the NPF cluster could not be separated due to their similar characteristics. Brines et al. (2015) and Chen et al. (2021) also reported this. The NPF and fresh vehicle cluster (Cluster 1) accounted for 16.06% in total, reaching the peak mode at 25 nm, with a large mean D_p value (51.75 nm). The proportion of the fossil fuel and biomass burning cluster (Cluster 2) was 26.02%, likely due to its proximity to the Wusongkou wharf. In a study of passenger ship plumes (Jonsson et al., 2011), the mean D_p was 38 nm, with a unimodal distribution. Cluster 3 accounted for the highest proportion among the four clusters (36.19%) and was classified as regional transmission based on its characteristics. The proportion of the clean background cluster (Cluster 4) in the total PNSD at the SH site was 21.71%.

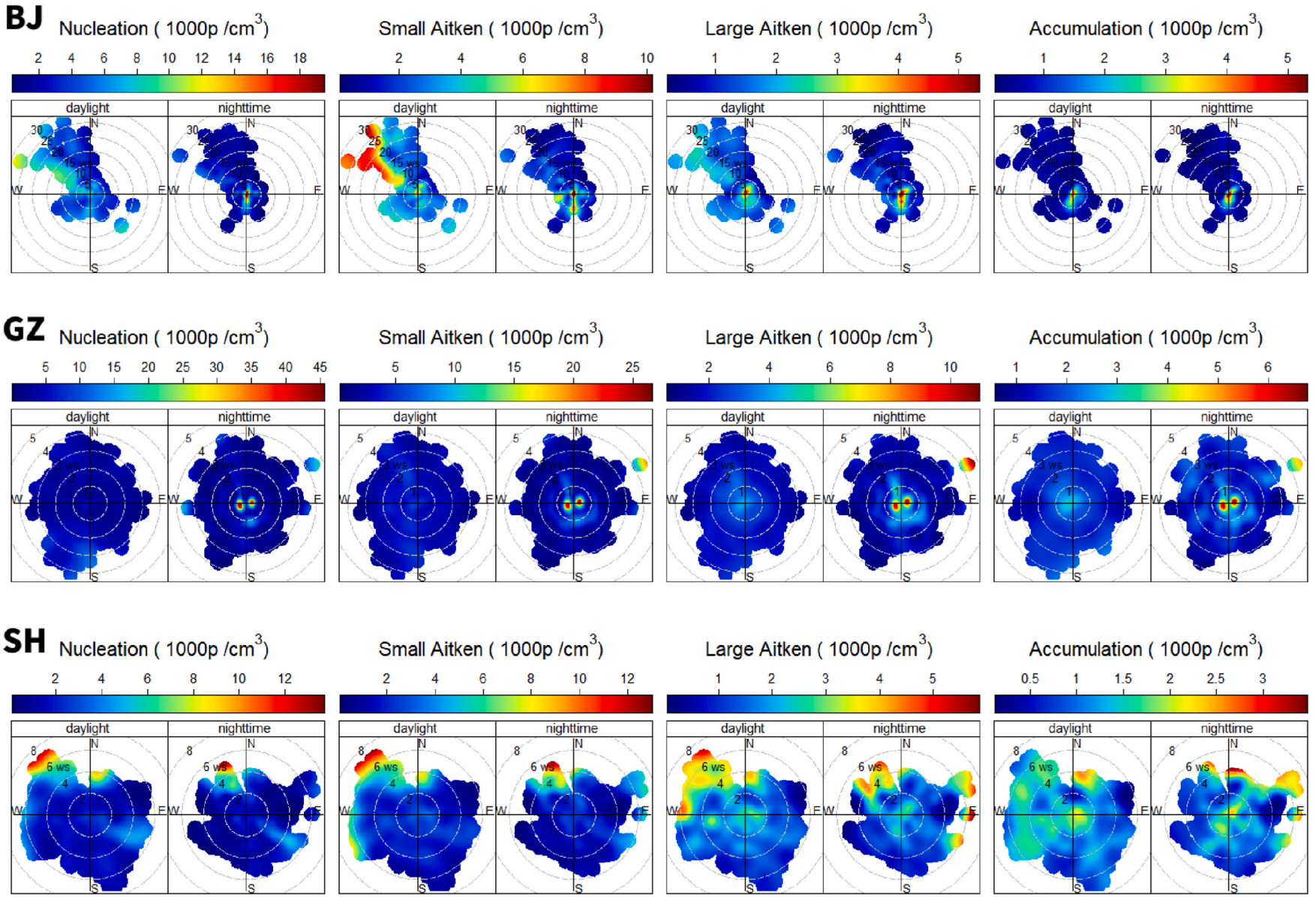

Fig. 5. Daytime and nighttime bivariate polar plots of nucleation (11.3–25 nm), small Aitken (25–50 nm), large Aitken (50–100 nm) and accumulation (100–532 nm) mode particulate concentrations at the BJ, GZ, and SH sites. The center of each plot represents a wind speed of zero, increasing radially outward. The color scale shows concentrations (unit: 1000 particles/cm^3). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

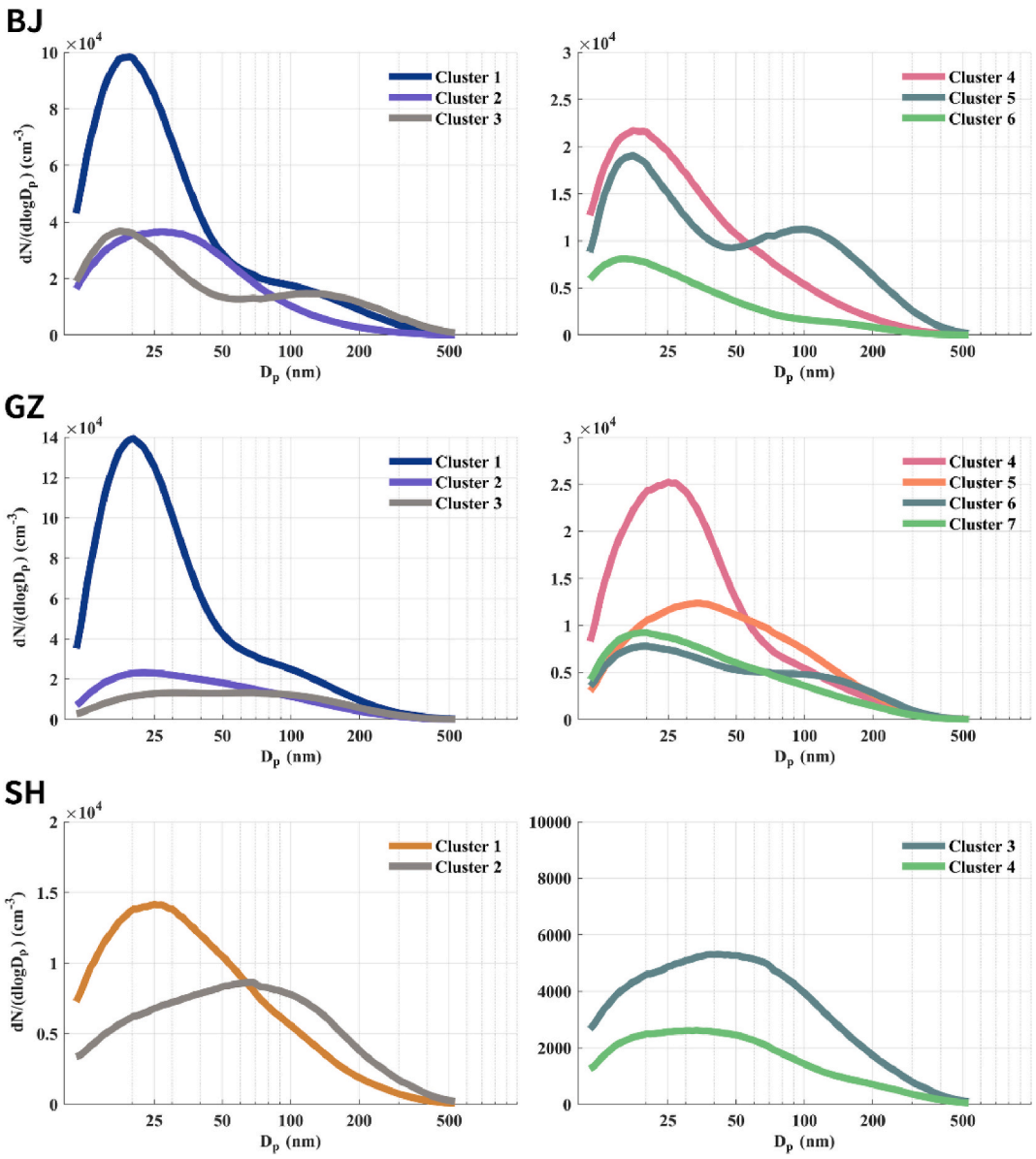

Fig. 6. Particle number size distributions of each cluster at the BJ, GZ, and SH sites.

3.4. Nocturnal nucleation events and traffic contribution

As mentioned in section 3.1, frequent NNEs were observed in aerosol observations at the BJ and GZ sites. Unlike NPF, which is driven by photochemical nucleation, NNEs consist of explosions of nucleation-mode particles lasting several hours, with no subsequent growth of particles. This is similar to the "transient" nucleation events described by Stanier et al. (2004), in which the number concentration of particles in the nucleation-mode size range increased rapidly but did not grow into larger particles. They divided atmospheric nucleation events into two categories, the other being "regional" nucleation events, often referred to as NPF events. Similar nocturnal nucleation has been observed in other studies (Ortega et al., 2012; Kecorius et al., 2015; Y. Guo et al., 2021). We thus proposed a novel approach that differentiates nucleation factors into two sources: photochemically driven NPF taking place during daytime and NNE taking place at night. In the analysis of section 3.1, we excluded the possibility that the reduction of the boundary layer had a significant impact on NNE. In section 3.2, combined with wind direction and speed information, we concluded that it was mainly caused by local emission sources. In section 3.3, NNE-related clusters were accompanied by high gas pollutant concentrations related to traffic emissions. Many studies have attributed such "transient" nucleation events to traffic emissions. For example, Rivas et al. (2020) used NO_x as an indicator of traffic emissions and divided nucleation factors into two sources: photo-nucleation and traffic-nucleation. Rodríguez and Cuevas (2007) divided the number of atmospheric ambient particles affected by fresh traffic discharge into aerosol particles directly emitted in the particle phase and new particles formed by nucleation during the dilution and cooling of vehicle exhaust.

There is also increasing evidence that nocturnal nucleation in urban atmospheric environments is closely related to intensive local transport emissions. In addition to primary particles, vehicle exhaust contains many gaseous precursors discharged into the atmospheric environment through hot exhaust air that quickly cools down and condenses with water vapor to form nucleation-mode particles (S. Guo et al., 2020; Olin et al., 2020). M. Wang et al. (2020) found that nitric acid and ammonia can rapidly condense and grow into new atmospheric particles by simulating real atmospheric environmental conditions at low temperatures in winter. When the temperature is cold enough (below −15 °C), nitric acid and ammonia gas can nucleate directly by an acid-base

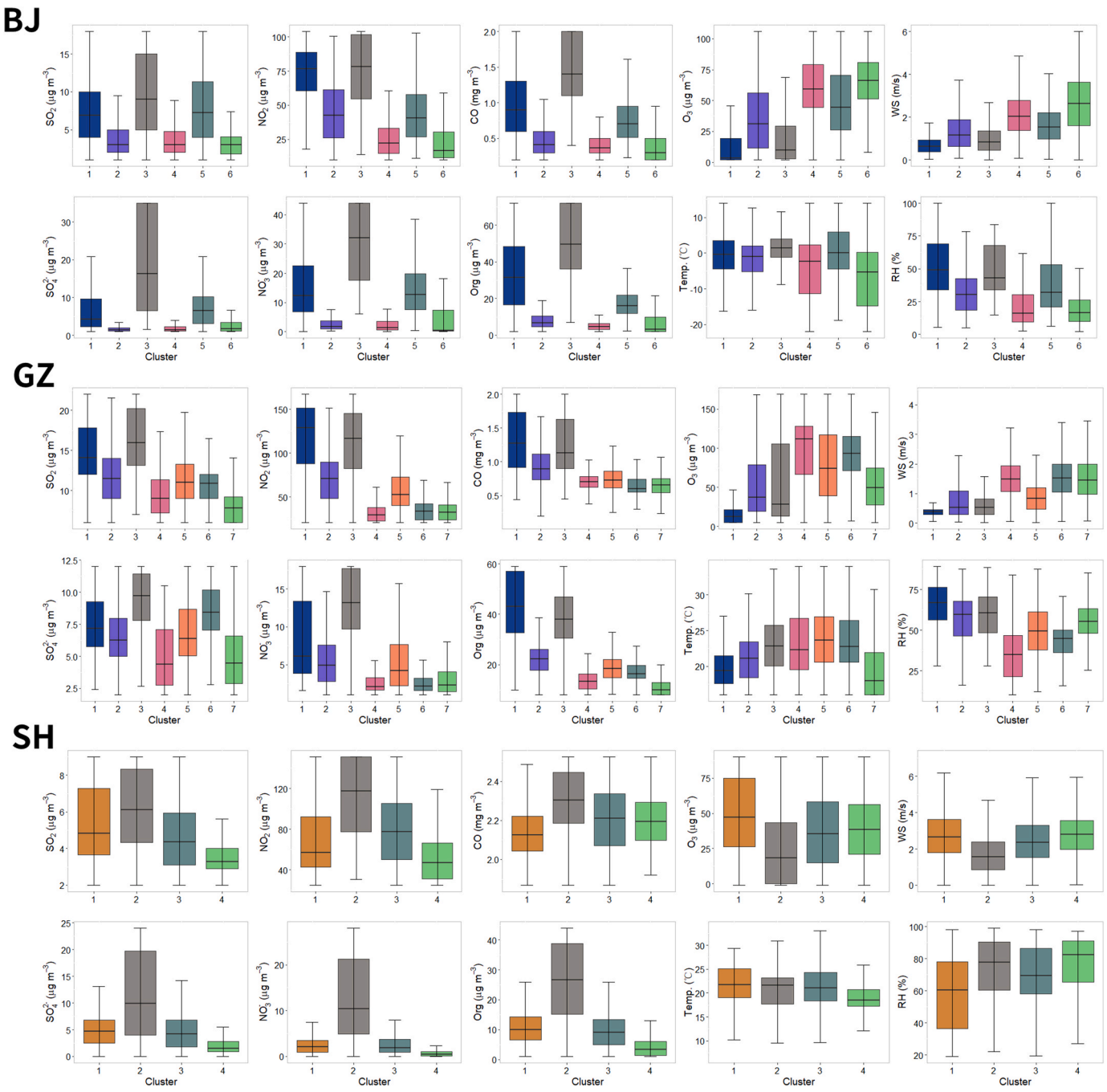

Fig. 7. SO_2, SO_4^{2-}, NO_2, NO_3^-, CO, Org, and O_3 concentrations and the temperature (Temp.), wind speed (WS), and relative humidity (RH) of each cluster at the BJ, GZ, and SH sites. The horizontal lines, boxes, and whiskers represent the median, 25–75th percentile, and 10–90th percentile, respectively.

stabilization mechanism to form ammonium nitrate particles. This nucleation mechanism may be even more pronounced during cold winter nights, driven by both vertical mixing and a strong local source (traffic emissions). Additionally, cars emit SO_2 from sulfur-containing fuels and lubricants (Olin et al., 2020). SO_2 can be photochemically oxidized in the atmosphere by an oxidizing agent (such as the hydroxyl radical, OH) from the production of solar radiation, which is an important source of H_2SO_4, the most important gaseous precursor for NPF in the atmosphere (Kulmala et al., 2014). Y. Guo et al. (2021) recently found that the reaction between SO_2 and the oxidant produced by the olefin-ozone reaction is responsible for the frequent formation of sulfates during winter nights, contributing a large amount of nanocluster aerosols. In addition to the formation mechanism of inorganic aerosol particles mentioned above, primary organic aerosols (POA) directly emitted from fossil fuels can also rapidly convert the water phase to secondary organic aerosols (SOA) under high relative humidity conditions (J. Wang et al., 2021). We also found that the concentration of Org and the proportion of Org in total particle mass concentrations increased noticeably at the same time as NNE at the BJ and GZ sites at night. There was also a minor increase in the SO_4^{2-} composition at the BJ site at night.

Overall, the exhaust gas emitted by urban traffic contains abundant gaseous precursors (NOx, SO_2, POA, etc.), which can form new secondary NO_3^-, SO_4^{2-} and Org particles at night in winter under low temperature and high humidity conditions. At the same time, driven by the reduced boundary layer height at night, these newly generated secondary particles and the primary particles directly discharged from the

Table 2
Characteristics of clusters identified from our clustering analysis.

BJ	Cluster	1	2	3	4	5	6
	Source/ Origin	nocturnal nucleation events	fresh vehicle emissions	diesel or biomass combustion	new particle formation	regional transport	clean background
	Rate	14.71%	7.06%	15.21%	22.85%	20.87%	19.30%
GZ	**Cluster**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
	Source/ Origin	nocturnal nucleation events	fresh vehicle emissions	diesel or biomass combustion	new particle formation	aging aerosol	regional transport	clean background
	Rate	17.17%	17.78%	9.90%	6.45%	17.49%	14.73%	16.48%
SH	**Cluster**	**1**	**2**	**3**	**4**
	Source/ Origin	new particle formation and fresh vehicle emissions	diesel or biomass combustion	regional transport	clean background
	Rate	16.06%	26.02%	36.19%	21.73%

tail gas cannot diffuse as rapidly as in the daytime. As a result, our instruments can detect high concentrations of nucleated particles at night. Of particular note, the preponderance of NO_3^-, SO_4^{2-} or Org particles generated at night in each megacity considered here may vary depending on the type of vehicle in the local traffic, as well as other emission sources. We conclude that traffic emissions were the main contributor in BJ and GZ, with the direct and significant impacts on the PNSD ranged from 37% to 45%.

4. Conclusions

We have conducted comprehensive field experiments at three megacities in China (BJ, GZ and SH) under different environmental, climate and weather conditions to characterize air pollution, and to identify and quantify major contributors by applying the k-clustering method. Compared to previous studies that were chiefly conducted on a single city with different observation instruments and methods, we employed the same sets of measurements following the same analysis method so that the analysis results were consistent and comparable among the three megacities.

The PNSD of the three sites showed higher total PNCs in BJ and GZ, mainly due to the increase of nucleation-mode particles at night. The average concentration of nucleation-mode particles in BJ at night was 14,199 cm^{-3}, about twice that in the daytime. Meanwhile, the proportion of nucleation-mode particles in the PNC increased from 39.26% in the daytime to 46.23% at night. The concentration of nucleation-mode particles in GZ at night was 4.4 times that in the daytime, and the contribution to the PNC increased from 28.36% to 43.84%. The high PNCs in BJ and GZ were mainly driven by frequent NNEs. The proportions changes of different modes PNCs and different chemical components during daytime and nighttime, respectively, indicate that NNEs are more influenced by emission source enhancement rather than meteorological factors. Using wind information, nighttime nucleation-mode particles in BJ were mainly related to local pollution source emissions, as was the case in GZ.

The k-means clustering technique was applied to datasets of three megacity sites with different source contributions driven by different meteorological conditions and emission patterns. Six, seven, and four city contribution sources were identified at the BJ, GZ and SH sites respectively. NNE-related clusters were identified for the BJ and GZ sites. The NNE-related clusters were all accompanied by higher traffic-related gaseous pollutants, indicating a strong association with traffic emissions. Traffic emissions are also major contributors to the BJ and GZ sites, with the significant and direct contribution of 37%–45%. This work describes in detail the different source contributions affecting aerosol PNC and PNSD in the atmospheric environment of three megacities in China, and will be helpful in evaluating the impact on human health in future studies.

Data availability

Data from the field experiment used in this study are available from the first author upon request (zhangdm0118@163.com).

Author contributions

Dongmei Zhang, Hao Wu and Zhanqing Li conceived the conceptual development of the manuscript. Dongmei Zhang directed and performed of the experiments with Hao Wu, Tong Wu, Rongmin Ren, Zhaoxin Cai, Chen Liang and Lu Chen. Dongmei Zhang conducted the data analysis and wrote the draft of the manuscript, and all authors edited and commented on the various sections of the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work was funded by the Natural Science Foundation of China (Grant No. 42030606), the National Key R&D Program of China (Grant No. 2017YFC1501702), the Guangdong Basic and Applied Basic Research Fund (2020B1515130003), and the Guangzhou Science and Technology Bureau (No. 202206010016). We thank all participants in the three field campaigns for their laborious work and cooperation.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.atmosenv.2022.119114.

References

Agudelo, D.M., Teixeira, E.C., Braga, M., et al., 2019. Cluster analysis of urban ultrafine particles size distributions. Atmos. Pollut. Res. 10 (1), 45–52. https://doi.org/10.1016/j.apr.2018.06.006.

Beddows, D.C.S., Dall'Osto, M., Harrison, R.M., 2009. Cluster analysis of rural, urban, and curbside atmospheric particle size data. Environ. Sci. Technol. 43 (13), 4694–4700. https://doi.org/10.1021/es803121t.

Brines, M., Beddows, D.C.S., Querol, X., et al., 2015. Traffic and nucleation events as main sources of ultrafine particles in high-insolation developed world cities. Atmos. Chem. Phys. 15 (10), 5929–5945. https://doi.org/10.5194/acp-15-5929-2015.

Buonanno, G., Morawska, L., Stabile, L., 2009. Particle emission factors during cooking activities. Atmos. Environ. 43 (20), 3235–3242. https://doi.org/10.1016/j.atmosenv.2009.03.044.

Chen, L., Qi, X., Nie, W., et al., 2021. Cluster analysis of submicron particle number size distributions at the SORPES station in the Yangtze River Delta of East China. J. Geophys. Res. Atmos. 126 (13) https://doi.org/10.1029/2020jd034004.

Cheung, H.C., Chou, C.C.K., Huang, W.R., et al., 2013. Characterization of ultrafine particle number concentration and new particle formation in an urban environment of Taipei, Taiwan. Atmos. Chem. Phys. 13 (17), 8935–8946. https://doi.org/10.5194/acp-13-8935-201.

Chu, B., Dada, L., Liu, Y., et al., 2021. Particle growth with photochemical age from new particle formation to haze in the winter of Beijing, China. Sci. Total Environ. 753, 142207. https://doi.org/10.1016/j.scitotenv.2020.142207.

Du, W., Zhao, J., Wang, Y., et al., 2017. Simultaneous measurements of particle number size distributions at ground level and 260 m on a meteorological tower in urban Beijing, China. Atmos. Chem. Phys. 17 (11), 6797–6811. https://doi.org/10.5194/acp-17-6797-2017.

Garate, L., Escudero-Lourdes, C., Salado, D., et al., 2020. Anthropogenic iron oxide nanoparticles induce damage to brain microvascular endothelial cells forming the blood-brain barrier. J. Alzheim. Dis. 76 (4), 1527–1539. https://doi.org/10.3233/jad-190929.

Gong, X., Richman, M.B., 1995. On the application of cluster Analysis to growing season precipitation data in north America east of the rockies. J. Clim. 8 (4), 897–931. https://doi.org/10.1175/1520-0442(1995)008<0897:OTAOCA>2.0.CO;2.

Govender, P., Sivakumar, V., 2020. Application of k-means and hierarchical clustering techniques for analysis of air pollution: a review (1980–2019). Atmos. Pollut. Res. 11 (1), 40–56. https://doi.org/10.1016/j.apr.2019.09.009.

Guo, S., Hu, M., Zamora, M.L., et al., 2014. Elucidating severe urban haze formation in China. Proc. Natl. Acad. Sci. U. S. A. 111 (49), 17. https://doi.org/10.1073/pnas.1419604111, 373–17,378.

Guo, S., Hu, M., Peng, J., et al., 2020. Remarkable nucleation and growth of ultrafine particles from vehicular exhaust. Proc. Natl. Acad. Sci. U. S. A. 117 (7), 3427–3432. https://doi.org/10.1073/pnas.1916366117.

Guo, Y., Yan, C., Li, C., et al., 2021. Formation of nighttime sulfuric acid from the ozonolysis of alkenes in Beijing. Atmos. Chem. Phys. 21 (7), 5499–5511. https://doi.org/10.5194/acp-21-5499-2021.

Hama, S.M.L., Cordell, R.L., Monks, P.S., 2017. Quantifying primary and secondary source contributions to ultrafine particles in the UK urban background. Atmos. Environ. 166, 62–78. https://doi.org/10.1016/j.atmosenv.2017.07.013.

Jain, A.K., Murty, M.N., Flynn, P.J., 1999. Data clustering: a review. ACM Comput. Surv. 31 (3), 264–323. https://doi.org/10.1145/331499.331504.

Jonsson, Å.M., Westerlund, J., Hallquist, M., 2011. Size-resolved particle emission factors for individual ships. Geophys. Res. Lett. 38 (13) https://doi.org/10.1029/2011gl047672.

Kanawade, V.P., Tripathi, S.N., Chakraborty, A., et al., 2020. Chemical characterization of sub-micron aerosols during new particle formation in an urban atmosphere. Aerosol Air Qual. Res. 20 (6), 1294–1305. https://doi.org/10.4209/aaqr.2019.04.0196.

Kecorius, S., Zhang, S., Wang, Z., et al., 2015. Nocturnal aerosol particle formation in the North China Plain. Lith. J. Phys. 55 (1), 44–53.

Ketzel, M., Berkowicz, R., 2005. Multi-plume aerosol dynamics and transport model for urban scale particle pollution. Atmos. Environ. 39 (19), 3407–3420. https://doi.org/10.1016/j.atmosenv.2005.01.058.

Kulmala, M., Petaja, T., Ehn, M., et al., 2014. Chemistry of atmospheric nucleation: on the recent advances on precursor characterization and atmospheric cluster composition in connection with atmospheric new particle formation. Annu. Rev. Phys. Chem. 65, 21–37. https://doi.org/10.1146/annurev-physchem-040412-110014.

Kulmala, M., Luoma, K., Ding, A.J., 2016. On the mode-segregated aerosol particle number concentration load: contributions of primary and secondary particles in Hyytiala and Nanjing. Boreal Environ. Res. 21 (3–4), 319–331.

Kumar, P., Morawska, L., Birmili, W., et al., 2014. Ultrafine particles in cities. Environ. Int. 66, 1–10. https://doi.org/10.1016/j.envint.2014.01.013.

Li, S., Zhang, F., Jin, X., et al., 2020. Characterizing the Ratio of Nitrate to Sulfate in Ambient Fine Particles of Urban Beijing during 2018–2019, vol. 237. Atmospheric Environment. https://doi.org/10.1016/j.atmosenv.2020.117662.

Li, Z., Wang, Y., Guo, J., et al., 2019. East asian study of tropospheric aerosols and their impact on regional clouds, precipitation, and climate (EAST-AIR$_{CPC}$). J. Geophys. Res. Atmos. 124 (23), 13. https://doi.org/10.1029/2019jd030758, 026–13,054.

Liang, C.S., Wu, H., Li, H.Y., et al., 2020. Efficient data preprocessing, episode classification, and source apportionment of particle number concentrations. Sci. Total Environ. 744, 140923. https://doi.org/10.1016/j.scitotenv.2020.140923.

Lorelei de Jesus, A., Thompson, H., Knibbs, L.D., et al., 2020. Long-term trends in $PM_{2.5}$ mass and particle number concentrations in urban air: the impacts of mitigation measures and extreme events due to changing climates. Environ. Pollut. 263 https://doi.org/10.1016/j.envpol.2020.114500.

Olin, M., Kuuluvainen, H., Aurela, M., et al., 2020. Traffic-originated nanocluster emission exceeds H2SO4-driven photochemical new particle formation in an urban area. Atmos. Chem. Phys. 20 (1), 1–13. https://doi.org/10.5194/acp-20-1-2020.

Ortega, I.K., Suni, T., Boy, M., et al., 2012. New insights into nocturnal nucleation. Atmos. Chem. Phys. 12 (9), 4297–4312. https://doi.org/10.5194/acp-12-4297-2012.

Perera, F.P., 2017. Multiple threats to child health from fossil fuel combustion: impacts of air pollution and climate change. Environ. Health Perspect. 125 (2), 141–148. https://doi.org/10.1289/EHP299.

Rivas, I., Beddows, D.C.S., Amato, F., et al., 2020. Source apportionment of particle number size distribution in urban background and traffic stations in four European cities. Environ. Int. 135, 105345. https://doi.org/10.1016/j.envint.2019.105345.

Rodríguez, S., Cuevas, E., 2007. The contributions of "minimum primary emissions" and "new particle formation enhancements" to the particle number concentration in urban air. J. Aerosol Sci. 38 (12), 1207–1219. https://doi.org/10.1016/j.jaerosci.2007.09.001.

Rönkkö, T., Kuuluvainen, H., Karjalainen, et al., 2017. Traffic is a major source of atmospheric nanocluster aerosol. Proc. Natl. Acad. Sci. USA 114 (29), 7549–7554. https://doi.org/10.1073/pnas.1700830114.

Salimi, F., Ristovski, Z., Mazaheri, M., et al., 2014. Assessment and application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment. Atmos. Chem. Phys. 14 (21), 11. https://doi.org/10.5194/acp-14-11883-2014, 883–11,892.

Stanier, C.O., Khlystov, A.Y., Pandis, S.N., 2004. Nucleation events during the Pittsburgh air quality study: description and relation to key meteorological, gas phase, and aerosol parameters - special issue of aerosol science and technology on findings from the fine particulate matter supersites program. Aerosol. Sci. Technol. 38 (Suppl. 1), 253–264. https://doi.org/10.1080/02786820390229570.

Sun, Y., Chen, C., Zhang, Y., et al., 2016. Rapid formation and evolution of an extreme haze episode in Northern China during winter 2015. Sci. Rep. 6 (1), 27151. https://doi.org/10.1038/srep27151.

von Bismarck, C., Birmili, W., Ketzel, M., et al., 2013. Characterization of parameters influencing the spatio-temporal variability of urban particle number size distributions in four European cities. Atmos. Environ. 77, 415–429. https://doi.org/10.1016/j.atmosenv.2013.05.029.

Vu, T.V., Delgado, J.M., Harrison, R.M., 2015. Review: particle number size distributions from seven major sources and implications for source apportionment studies. Atmos. Environ. 122, 114–132. https://doi.org/10.1016/j.atmosenv.2015.09.027.

Wang, J., Ye, J., Zhang, Q., et al., 2021. Aqueous production of secondary organic aerosol from fossil-fuel emissions in winter Beijing haze. Proc. Natl. Acad. Sci. U. S. A. 118 (8) https://doi.org/10.1073/pnas.2022179118.

Wang, M., Kong, W., Marten, R., et al., 2020. Rapid growth of new atmospheric particles by nitric acid and ammonia condensation. Nature 581 (7807), 184–189. https://doi.org/10.1038/s41586-020-2270-4.

Wang, Y., Li, Z., Zhang, R., et al., 2019. Distinct ultrafine- and accumulation-mode particle properties in clean and polluted urban environments. Geophys. Res. Lett. 46 (19), 1. https://doi.org/10.1029/2019gl084047, 0918–10,925.

Weber, S., Kordowski, K., Kuttler, W., 2013. Variability of particle number concentration and particle size dynamics in an urban street canyon under different meteorological conditions. Sci. Total Environ. 449, 102–114. https://doi.org/10.1016/j.scitotenv.2013.01.044.

Wegner, T., Hussein, T., Hämeri, K., et al., 2012. Properties of aerosol signature size distributions in the urban environment as derived by cluster analysis. Atmos. Environ. 61, 350–360. https://doi.org/10.1016/j.atmosenv.2012.07.048.

Who, 2018. Global Urban Ambient Air Pollution Database.

Wu, H., Li, Z., Jiang, M., et al., 2021. Contributions of Traffic Emissions and New Particle Formation to the Ultrafine Particle Size Distribution in the Megacity of Beijing. Atmospheric Environment, p. 118652. https://doi.org/10.1016/j.atmosenv.2021.118652.

Zhang, F., Ren, J., Fan, T., et al., 2019. Significantly enhanced aerosol CCN activity and number concentrations by nucleation-initiated haze events: a case study in urban Beijing. J. Geophys. Res. Atmos. 124 (24), 14. https://doi.org/10.1029/2019jd031457, 102–114,113.

Zhang, Q., Jia, S., Yang, L., et al., 2021. New particle formation (NPF) events in China urban clusters given by severe composite pollution background. Chemosphere 262, 127842. https://doi.org/10.1016/j.chemosphere.2020.127842.

Zhang, R., Khalizov, A., Wang, L., et al., 2012. Nucleation and growth of nanoparticles in the atmosphere. Chem. Rev. 112 (3), 1957–2011. https://doi.org/10.1021/cr2001756.

Zhang, R., Wang, G., Guo, S., et al., 2015. Formation of urban fine particulate matter. Chem. Rev. 115 (10), 3803–3855. https://doi.org/10.1021/acs.chemrev.5b00067.

Zhang, Y., Du, W., Wang, Y., et al., 2018. Aerosol chemistry and particle growth events at an urban downwind site in North China Plain. Atmos. Chem. Phys. 18 (19), 14. https://doi.org/10.5194/acp-18-14637-2018, 637–14,651.

Zhao, J., Du, W., Zhang, Y., et al., 2017. Insights into aerosol chemistry during the 2015 China Victory Day parade: results from simultaneous measurements at ground level and 260 m in Beijing. Atmos. Chem. Phys. 17 (4), 3215–3232. https://doi.org/10.5194/acp-17-3215-2017.