JGR Atmospheres

RESEARCH ARTICLE
10.1029/2020JD034004

Special Section:
Atmospheric PM2.5 in China: physics, chemistry, measurements, and modeling

Cluster Analysis of Submicron Particle Number Size Distributions at the SORPES Station in the Yangtze River Delta of East China

Liangduo Chen[1,2], Ximeng Qi[1,2], Wei Nie[1,2], Jiaping Wang[1,2], Zheng Xu[1,2], Tianyi Wang[1,2], Yuliang Liu[1,2], Yicheng Shen[1,2,3], Zhengning Xu[1,2,4], Tom V. Kokkonen[1,2], Xuguang Chi[1,2], Pasi P. Aalto[5], Pauli Paasonen[5], Veli-Matti Kerminen[5], Tuukka Petäjä[5], Markku Kulmala[5], and Aijun Ding[1,2]

[1]Joint International Research Laboratory of Atmospheric and Earth System Sciences, School of Atmospheric Sciences, Nanjing University, Nanjing, China, [2]Jiangsu Provincial Collaborative Innovation Center for Climate Change, Nanjing University, Nanjing, China, [3]Now at State Key Joint Laboratory of Environment Simulation and Pollution Control, School of Environment, Tsinghua University, Beijing, China, [4]Now at Department of Environment and Resources, Zhejiang University, Hangzhou, China, [5]Institute for Atmospheric and Earth Systems Research/Physics, Faculty of Science, University of Helsinki, Helsinki, Finland

Abstract Submicron particles in polluted regions have received much attention because of their influences on human health and climate. A k-means clustering technique was performed on a data set of particle number size distributions (PNSD) that was obtained over more than 3 years in the Yangtze River Delta (YRD) region of East China. With simultaneous measurements of meteorological conditions, trace gases and aerosol compositions, seven clusters were categorized and interpreted. Cluster 1 and cluster 2, which accounted for 9.9% of the total PNSD data, were attributed to new particle formation (NPF) and vehicle exhaust emissions with different intensities; Cluster 3 and Cluster 4, which accounted for 10.5% of the total PNSD data, were related to the growth of nucleation mode particles; Cluster 5, which accounted for 37.9% of the total data, was attributed to the humid YRD background; and Cluster 6 and Cluster 7, which accounted for 41.6% of the total data set, were both pollution-related clusters with similar mass concentrations but completely different PNSD. Although the $PM_{2.5}$ mass concentrations were somewhat similar, the particle number concentrations of the accumulation mode particles could vary by more than one order of magnitude from the urban background cluster to the pollution-related clusters. The cluster proximity diagram and conversion flow chart of clusters clearly show the influence of NPF and growth on haze, as well as the conversion between background and polluted conditions. This study highlights the importance of PNSD for understanding urban air quality and recommends the clustering technique for analyzing complex PNSD datasets.

Plain Language Summary Submicron particles in polluted regions have significant influences on human health and climate. Based on long-term field measurements, we used the k-means clustering technique to characterize the number size distributions of submicron particles in the Yangtze River Delta (YRD) of China. Seven clusters were categorized and interpreted. New particle formation (NPF), fossil fuel combustion and biomass burning are the main sources of submicron particles in the YRD. The influences of NPF and growth on haze, as well as the conversion between background and polluted conditions, were found.

Key Points:

- New particle formation (NPF), fossil fuel combustion and biomass burning are the main sources of submicron particles in Nanjing
- The influences of NPF and growth on haze, and the conversion between background and pollution conditions were found
- The k-means cluster technique is an effective tool to categorize particle number size distribution data set

Supporting Information:

Supporting Information may be found in the online version of this article.

Correspondence to:

X. Qi,
qiximeng@nju.edu.cn

Citation:

Chen, L., Qi, X., Nie, W., Wang, J., Xu, Z., Wang, T., et al. (2021). Cluster analysis of submicron particle number size distributions at the SORPES station in the Yangtze River Delta of East China. *Journal of Geophysical Research: Atmospheres*, *126*, e2020JD034004. https://doi.org/10.1029/2020JD034004

Received 1 OCT 2020
Accepted 8 JUN 2021

1. Introduction

Atmospheric aerosols, especially submicron particles, have received much attention because of their influences on climate, air quality, and human health (Atkinson et al., 2014; Heal et al., 2012; IPCC, 2013; Knibbs et al., 2011; Kulmala et al., 2016; Rosenfeld et al., 2019; Wu et al., 2019). Both the climate and human health effects of atmospheric aerosols depend on particle sizes (Kerminen et al., 2012; Salma et al., 2015). Particles larger than ~50–100 nm not only can serve as cloud condensation nuclei but can also scatter solar radiation more effectively, which influences the Earth's radiative balance (Schmale et al., 2018; Shen et al., 2019). Recent evidence has shown that ultrafine particles smaller than 50 nm can also be activated in deep convective clouds (Fan et al., 2018). Moreover, ultrafine particles can penetrate into and be deposited

in the deeper parts of the respiratory system or can even penetrate the pulmonary epithelium and reach the cardiovascular system and thereby are more harmful to human health (Chen et al., 2016; Downward et al., 2018). Therefore, long-term measurements of particle number size distributions (PNSD), as one of the most important properties of atmospheric aerosols, are crucial for understanding the climate and human health effects of atmospheric aerosols.

In addition to being closely related to the impact of aerosols on climate and public health, PNSD also indicate the sources and atmospheric processes of aerosols. Various anthropogenic and natural emissions and multiple aerosol processes, such as new particle formation (NPF), condensation growth, evaporation, coagulation, and wet and dry deposition, can influence and cause drastic variations in PNSD (Asmi et al., 2011; Brines et al., 2015; Harrison et al., 2019; Liu et al., 2014; Wang et al., 2013). In urban environments, especially with growing urbanization, PNSD have become increasingly important for evaluating air quality and understanding haze formation. Due to the large size and complexity of PNSD data sets, the clustering technique, for example, k-means clustering, is an appropriate way to investigate the source apportionment of the number and size characteristics of submicron particles (Agudelo-Castaneda et al., 2019; Beddows et al., 2009; Brines et al., 2014, 2015; Sabaliauskas et al., 2013; Väänänen et al., 2013; Wegner et al., 2012). The k-means clustering technique categorizes PNSD data into reduced clusters with similar characteristics, such as their size peaks and temporal trends (Beddows et al., 2009). By using the k-means clustering technique, vehicle traffic was determined to be one of the main sources of submicron particles in many urban areas (Brines et al., 2015; Dall'Osto et al., 2012). Sub-10 nm particles can form behind exhaust tailpipes when the hot exhaust gases are diluted and cooled by the ambient air (Guo et al., 2020; Rönkkö et al., 2017). Although NPF events are expected to be less favored in polluted areas because of high condensation sinks, surprising events are observed with high frequency and have become another main source of submicron particles in urban areas (Chu et al., 2019; Kerminen et al., 2018; Kulmala et al., 2017; Wang, Wu et al., 2017). NPF events were even found to be one of the causes of haze formation in the megacities of China (Guo et al., 2014; Kulmala et al., 2021; Yao et al., 2018).

The Yangtze River Delta (YRD) region of East China is one of the hotspots of urban air pollution because of its high urbanization and population density (Li et al., 2011; Tie & Cao, 2009). Due to the complex anthropogenic emissions and feedbacks with the East Asian monsoons, the air pollution in the YRD is unique (Ding et al., 2019; Ding, Fu, Yang, Sun, Zheng et al., 2013). A number of studies have addressed air pollution in the YRD region, but they have mainly focused on particle mass concentrations and chemical compositions (Cheng et al., 2014; Ding et al., 2019; Sun et al., 2018; Xie et al., 2015). Only a limited number of long-term measurements of submicron PNSD have been conducted in the YRD region. Qi et al. (2015) analyzed a 2 year PNSD data set from the western part of the YRD region and found much higher submicron particle number concentrations compared with cities in Europe and North America. Seasonal variations in the PNSD with obvious signals from vehicle traffic in the cold season and NPF in the warm season were observed in the YRD, which demonstrate the complex sources of submicron particles (Ling et al., 2019; Qi et al., 2015). Although several studies have found that NPF and primary emissions from traffic and power plants have significant contributions to submicron particles (Du et al., 2012; Gao et al., 2009; Peng et al., 2014), the sources and aerosol processes governing the PNSD of submicron particles in polluted urban environments, such as the YRD region, are not well evaluated by applying clustering techniques to PNSD datasets. Meanwhile, NPF events with high formation and growth rates have frequently been observed in the YRD region even under high condensation sinks (Dai et al., 2017; Qi et al., 2015; Xiao et al., 2015). Sulfuric acid and dimethylamine were found to play an important role in nucleation in polluted areas, and high sulfuric acid concentrations can explain the initial growth of particles (Yao et al., 2018). However, the statistical characteristics and evolution of the PNSD during various types of NPF and thereby their potential impacts on air pollution in the YRD region are not yet fully understood.

Based on the continuous 6–800 nm PNSD observations from December 2011 to February 2015 and the simultaneous measurements of meteorological conditions, trace gases and aerosol compositions at the Station for Observing Regional Process of the Earth System (SORPES) in the YRD region of East China, we applied the k-means clustering technique to reduce the complexity of the data set. The purposes of this study are to (a) characterize the PNSD in polluted areas of East China, (b) identify the main sources and origins of submicron particles, (c) determine the evolution processes of submicron particles in the polluted

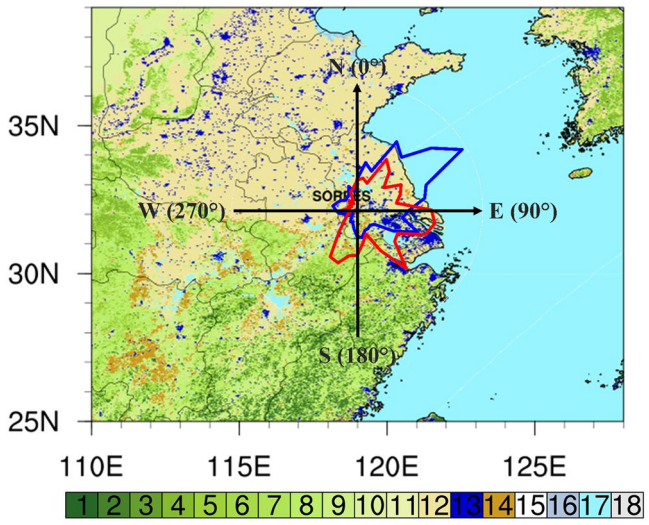

Figure 1. Location of the Station for Observing Regional Process of the Earth System with the land cover and prevailing wind information. (Red represents the prevailing wind in summer, and blue represents the prevailing wind in winter. The indexes in the color bar represent: 1 Evergreen Needleleaf, 2 Evergreen Broadleaf, 3 Deciduous Needleleaf, 4 Deciduous Broadleaf, 5 Mixed Forest, 6 Closed Shrublands, 7 Open Shrublands, 8 Woody Savannas, 9 Savannas, 10 Grasslands, 11 Permanent Wetlands, 12 Croplands, 13 Urban and Built-up, 14 Cropland and Mosaics, 15 Snow and Ice, 16 Bare Soil and Rocks, 17 Water Bodies, and 18 Unclassified).

atmosphere, and (d) investigate the PNSD in severe haze episodes and determine the relationships between NPF and air pollution.

2. Materials and Methods

2.1. Measurement Site

Field measurements were conducted at the SORPES station, which is located in Nanjing in the YRD region, East China (118°57′10″E, 32°07′14″N). The SORPES station can be regarded as a suburban site that is located 20 km northeast of downtown Nanjing. As shown in Figure 1, because of the unique location of the site, the PNSD can be influenced by various air masses from different regions (Qi et al., 2015). In the cold season, the prevailing winds come from the northeast and bring pollutants that are emitted in the North China Plain to the measurement site. In the warm season, the prevailing winds come from the southeast, and the air masses from South China contain high values of biogenic volatile organic compounds (BVOCs), which cause frequent NPF events. Moreover, the site is located downwind from the cluster of cities in the YRD and is thereby able to detect pollutant plumes from the YRD. In addition to the impacts of emissions, the changes in meteorological conditions at SORPES, such as temperature, radiation and precipitation, can influence both the sources and sinks of submicron particles and thereby complicate the PNSD. More details about the SORPES station can be found in Ding et al. (2016).

2.2. The Instrumentation

The PNSD of submicron particles were measured by a differential mobility particle sizer (DMPS) that was constructed at the University of Helsinki (Herrmann et al., 2014; Qi et al., 2015). This instrument joined the intercomparison workshop under the framework of European Supersites for Atmospheric Aerosol Research and Aerosols, Clouds, and Trace gases Research InfraStructure Network, during which the performance of the instrument was well evaluated (Wiedensohler et al., 2012). This instrument is a flow-switching-type of DMPS that can cover a size range from 6 to 800 nm with 29 channels by switching the sample and sheath flow rates with a differential mobility analyzer (Salma et al., 2011). A commercial condensation particle counter (CPC, TSI model 3772) was used to detect size-segregated particles. To obtain a lower cutoff diameter, the condenser temperature of CPC was changed from 22°C (the default value) to 10°C, and therefore, the temperature difference between the saturator and the condenser (ΔT) was more than 25°C. According to the calibration described by Wiedensohler et al. (2012), the counting efficiency for 6 nm particles is higher than 75% when ΔT is greater than 25°C. The samples were dried by a dryer before entering the DMPS, and two americium 241 sources (each of ~37 kBq) were used to ensure the equilibrium charge. The total particle number concentrations were measured directly by the CPC before each 10-min scanning cycle. As shown in Figure S1, for the entire data set (e.g., from December 2011 to February 2015), the total number concentrations that were calculated from the size distributions (NC_{DMPS}) exhibited good correlations with those that were directly observed by the CPC (NC_{CPC}), which demonstrated the high quality of the PNSD data. NC_{DMPS} was slightly lower than NC_{CPC}, especially when the number concentrations of the sub-10 nm particles were high (Figure S1), which was mainly because the DMPS underestimated the number concentrations of sub-10 nm particles (Wiedensohler et al., 2012).

The meteorological conditions (e.g., incoming solar radiation, temperature, relative humidity, wind speed, and wind direction); trace gases (e.g., O_3, SO_2, NO, NO_2, and CO); and $PM_{2.5}$ and aerosol compositions (e.g., water-soluble ions) were simultaneously observed at SORPES and helped to analyze the PNSD at SORPES. The meteorological conditions were observed by sensors and radiometers (e.g., a weather station and 4 component net radiometer; Campbell). The trace gas concentrations and $PM_{2.5}$ mass concentrations were

observed using a series of online analyzers (e.g., 49i, 43i, 42i, 48i and 5030SHARP for O_3, SO_2, NO + NO_2, CO and $PM_{2.5}$, respectively; Thermo Fisher Scientific). The aerosol compositions, for example, water soluble ions, were determined by the instrument for Measuring AeRosols and Gases (MARGA; Metrohm). All of the data in this study were averaged to 1 h resolution.

2.3. The K-Means Clustering Analysis

Given the large amount of data to be analyzed and the complicated PNSD at SORPES, a clustering technique, for example, a k-means clustering analysis, was applied to the PNSD data set. K-means clustering is an iterative algorithm. The objective of the k-means clustering analysis is to group the objects in the data set into a number of k clusters in which each object belongs to the cluster with the nearest mean (Hartigan & Wong, 1979). Compared with other statistical methods, such as self-organizing maps, clustering of large applications and affinity propagation, the k-means technique performed best when analyzing PNSD data (Salimi et al., 2014). The cluster analysis was performed on the hourly averaged PNSD data from December 2011 to February 2015, with 24,312 h of available data in total.

The number of representative clusters, *k* value, needs to be determined before running k-means clustering analysis. A *k* value less than 6 is expected to be too small to distinguish the different particle sources and transformation processes, while a *k* value larger than 10 is too complex for scientific discussions (Masiol et al., 2017; Wegner et al., 2012). The Dunn index, which is defined as the ratio between the minimal intercluster distance (the distance between objects of different clusters) to maximal intracluster distance (the distance between two objects belonging to one cluster), can help to identify dense and well-separated clusters. A higher Dunn index represents the optimum number of clusters. The calculated Dunn index was highest at a *k* value equal to 7 in the range from 6 to 10 (Figure S2). Therefore, seven representative clusters were chosen for k-means cluster analysis in this study.

3. Results and Discussions

3.1. The Identification and Interpretation of Each Cluster

Figure 2 shows the PNSD (Figures 2a–2c) and proportion (Figure 2d) of each cluster. The number concentrations in the nucleation mode (6–25 nm), Aitken mode (25–100 nm) and accumulation mode (100–800 nm), geometric mean diameters (GMD) and condensation sinks of each cluster are presented in Table 1. According to their temporal occurrences and associations with meteorological parameters and pollutants, each cluster type was identified and interpreted. The diurnal and seasonal variations in the temporal occurrences of each cluster are shown in Figure 3. Figure 4 presents the meteorological parameters (e.g., temperature, relative humidity, incoming solar radiation, and wind speed) and atmospheric pollutants (e.g., $PM_{2.5}$, O_3, SO_2, CO, NO, and NO_x) for each cluster, whereas Table S1 tabulates the mean values. Figure 5 shows the air mass backward footprints of each cluster based on the method described in Ding, Wang et al. (2013). Figure 6 presents the chemical composition of $PM_{2.5}$ in each cluster, whereas Table S2 summarizes the corresponding average mass concentrations.

3.1.1. Cluster 1: High-Intensity New Particle Formation and Fresh Vehicle Exhaust Emissions

Cluster 1 (C1) accounted for 1.7% of the total measurements. C1 exhibited a bimodal size distribution with peaks at ~20 and 100 nm. The highest number concentration of nucleation mode particles (i.e., 24,200 cm^{-3}) and lowest GMD (i.e., 26 nm) were found in C1. The maximum occurrence frequency of C1 was during the daytime when the incoming solar radiation was high (Figures 3a and 4c). The $PM_{2.5}$ mass concentration, relative humidity and condensation sink were low for C1 (Figure 4 and Table 1). Strong solar radiation, low humidity, and low condensation sink are the main limiting factors for NPF in Nanjing (Herrmann et al., 2014; Qi et al., 2015). The O_3 concentrations were high, and the SO_2 concentrations were sufficient at ~11.2 ppbv on average (Table S1), which favor NPF. Therefore, the high concentration of nucleation mode particles and favorable atmospheric conditions indicate that high intensity NPF during the daytime is one of the main sources of C1. Similar to NPF (Qi et al., 2015), C1 occurred mostly in spring and autumn, which further supports the idea that NPF contributes to C1. For the aerosol chemical compositions, sulfate accounted for 48.2% of the total water soluble ions in C1, with an average mass concentration of 11.9 $\mu g/m^3$,

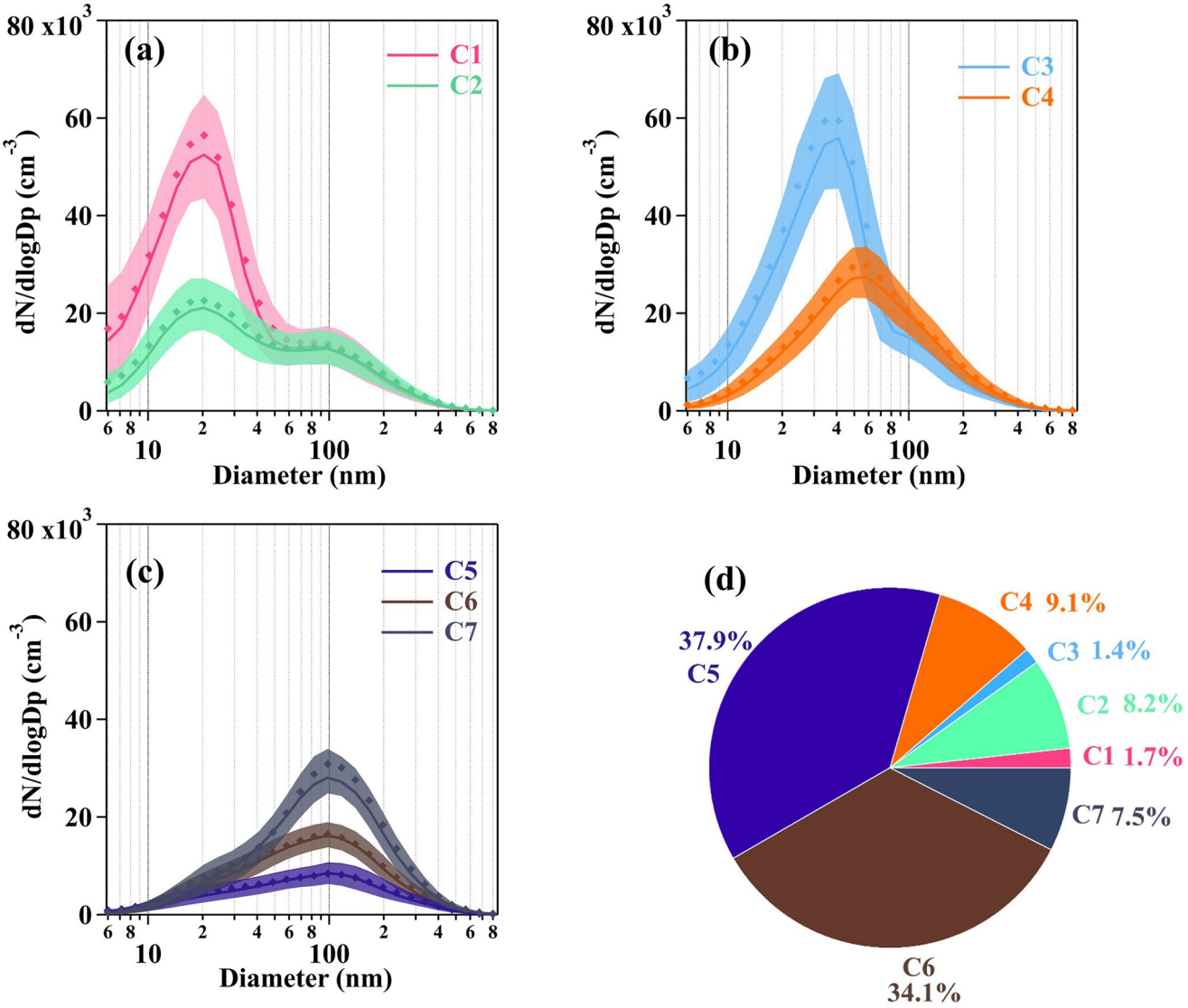

Figure 2. Particle number size distributions of (a) Clusters 1–2, (b) Clusters 3–4 and (c) Clusters 5–7 (solid lines, diamond markers and shaded areas represent the median, mean and 25th-75th percentile ranges, respectively), and (d) the frequency of occurrence of each cluster.

Table 1
Mean Values of the Number Concentrations of Nucleation, Aitken, Accumulation Mode Particles, GMD and CS in Each Cluster

	C1	C2	C3	C4	C5	C6	C7
Nucleation mode (#/cm^3)	24,200	9,900	13,100	4,300	1,800	2,300	2,500
Aitken mode (#/cm^3)	13,700	9,100	25,400	14,900	4,100	7,700	11,300
Accumulation mode (#/cm^3)	4,600	4,800	4,800	6,100	3,500	6,200	11,300
GMD (nm)	26	38	38	56	69	74	85
CS (10^{-2} s^{-1})	3.2	3.2	3.7	4.2	3.2	4.7	7.4

Abbreviations: CS, Condensation Sinks; GMD, Geometric Mean Diameters.

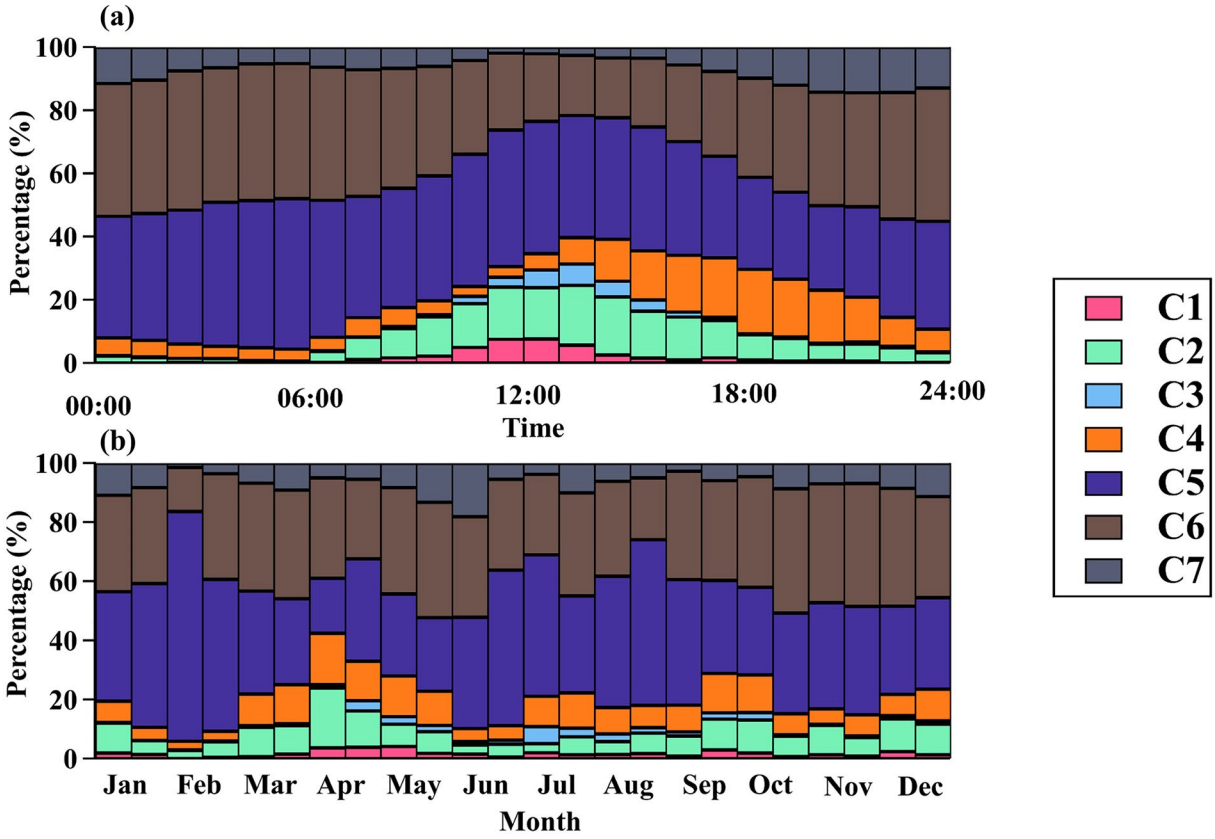

Figure 3. (a) Diurnal and (b) seasonal cycles of the occurrences of each cluster.

which was followed by ammonium (21.1%, 5.4 μg/m^3) and nitrate (20.6%, 5.1 μg/m^3). A high sulfate fraction in the water soluble ions indicates the strong photochemical oxidation of C1, which is consistent with the high solar radiation, as shown in Figure 4. Strong photochemical oxidation favors NPF as well.

Note that since C1 also occurred at night or late afternoon, fresh vehicle exhaust emissions might influence C1. As an indicator of petroleum fuel consumption, the 90th percentile of the NO_x concentrations in C1 was 59.6 ppbv, which was quite high compared with those at rural sites in the YRD, for example, ~13.4 ppbv

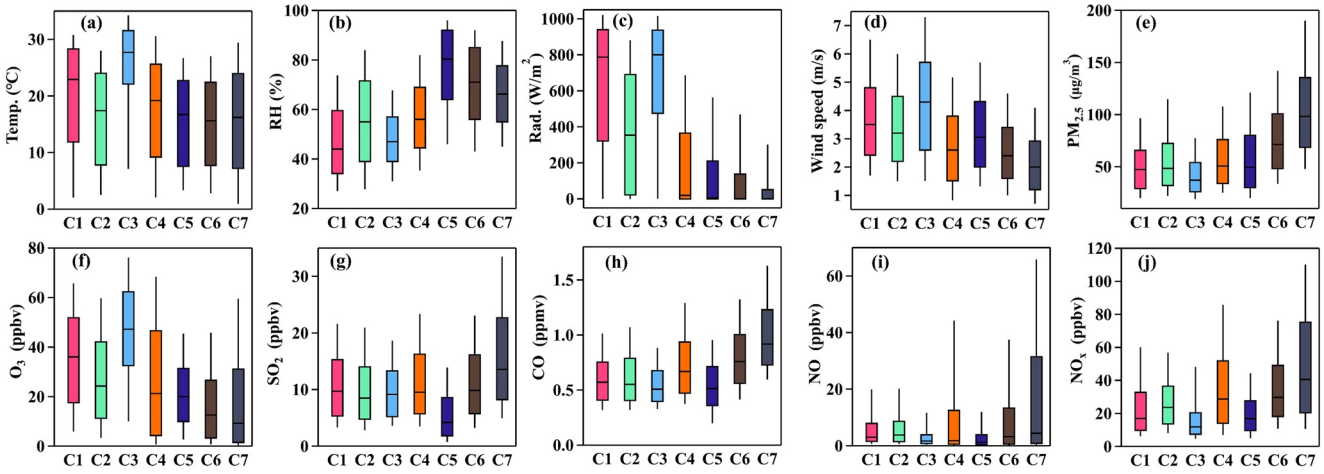

Figure 4. The (a) temperature, (b) relative humidity, (c) incoming shortwave radiation, (d) wind speed, (e) $PM_{2.5}$, (f) O_3, (g) SO_2, (h) CO, (i) NO, and (j) NO_x concentrations of each cluster. The horizontal lines, boxes and whiskers represent the median, 25–75th percentile and 10–90th percentile, respectively.

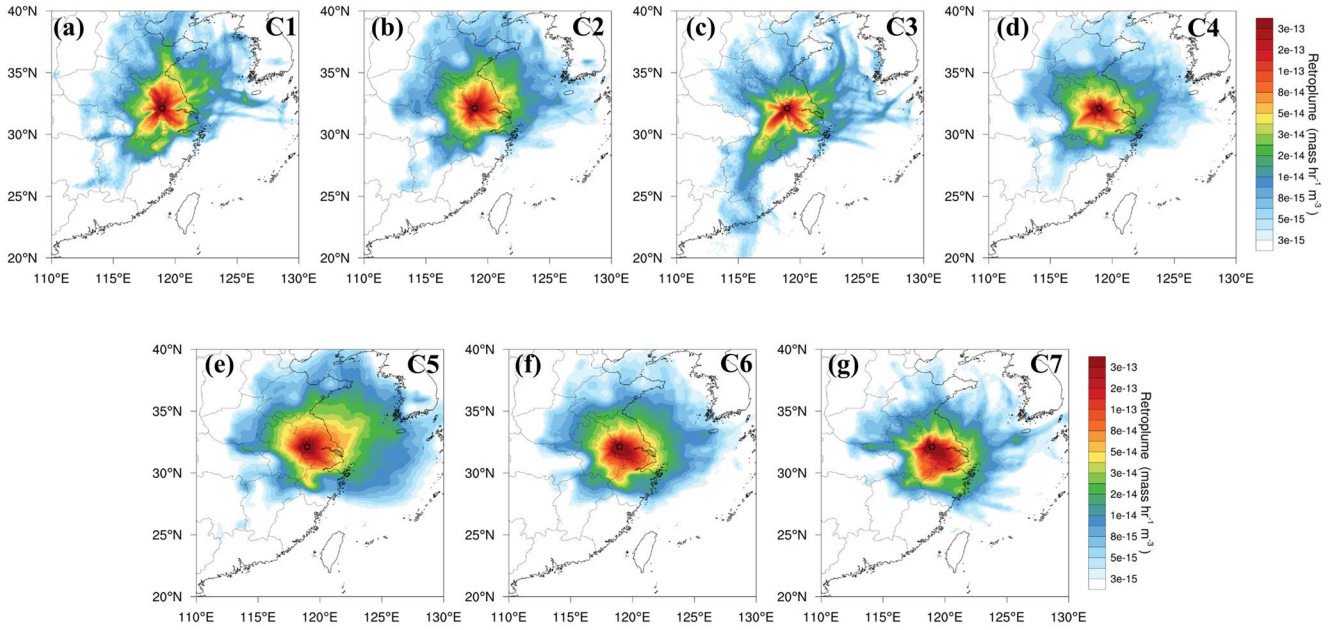

Figure 5. Backward airmass footprints of (a) Cluster 1, (b) Cluster 2, (c) Cluster 3, (d) Cluster 4, (e) Cluster 5, (f) Cluster 6, and (g) Cluster 7.

in Linan (Xu et al., 2008). A fresh traffic cluster with a major mode at ~20 nm was also found by Brines et al. (2015) and Agudelo-Castaneda et al. (2019). Rönkkö et al. (2017) found that traffic can even contribute to sub-3 nm particles in an urban environment. In this study, the NPF cluster and vehicle exhaust emissions cluster could not be separated even by increasing the representative number of clusters to 20 or by performing k-means clustering only for NPF event days (Figure S3), which was different from the studies by Agudelo-Castaneda et al. (2019) but was similar to the studies by Brines et al. (2015). The high growth rate of newly formed particles at SORPES could be one of the reasons for this, because high growth rates result in low amounts of data from the initial stage of NPF, which causes the size distributions of NPF and vehicle exhaust emissions to be similar. Moreover, because of the heavy traffic in Nanjing, the influence of vehicle emissions is frequent and can even occur simultaneously with NPF.

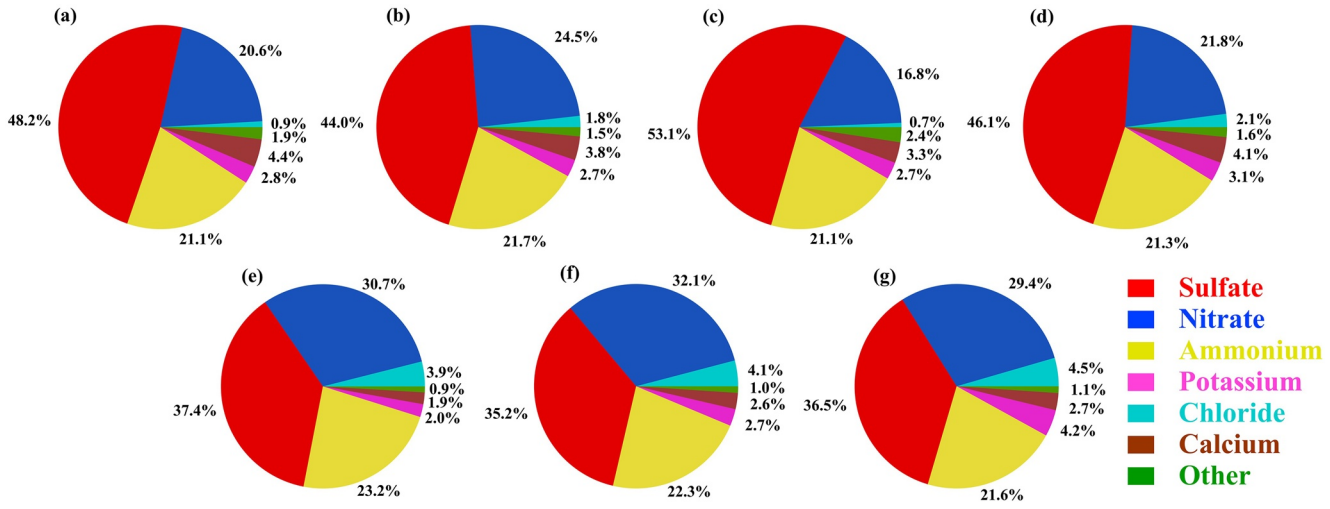

Figure 6. The average fractions of water soluble ions for (a) Cluster 1, (b) Cluster 2, (c) Cluster 3, (d) Cluster 4, (e) Cluster 5, (f) Cluster 6, and (g) Cluster 7.

3.1.2. Cluster 2: Weak New Particle Formation and Background Vehicle Exhaust Emissions

Cluster 2 (C2) accounted for 8.2% of all measurements. C2 exhibited a size distribution that was similar to that of C1 but with a weaker nucleation mode. The concentration of nucleation mode particles was 9,900 cm^{-3}, which was lower than that in C1, and the GMD was 38 nm, which was larger than that in C1. Similar to C1, the occurrence frequency of C2 exhibited a maximum during the daytime but was much higher than that of C1 during 6:00-24:00. The incoming solar radiation in C2 (386.1 W/m^2 on average, Table S1) was lower than that in C1. C2 had favorable atmospheric conditions for NPF, including relatively low values of $PM_{2.5}$ concentration (61.2 $\mu g/m^3$ on average) and CS (3.2×10^{-2} s^{-1}) and a sufficiently high SO_2 concentration (10.6 ppbv). The sulfate fraction in C2 (i.e., 44.0%) was lower than that in C1, which suggested weaker photochemical oxidation. Similar to C1, the 90th percentile of the NO_x concentrations in C2 was high and indicated the role of vehicle exhaust emissions. Since C2 exhibited more stable occurrences for diurnal and seasonal cycles, C2 represented the weak NPF and background vehicle exhaust emissions.

Both C1 and C2 represented the PNSD during NPF events, but the intensities of the NPF events were different. Figure S4 presents the J_6 (the formation rate of 6 nm particles) and $GR_{6–30\ nm}$ (the growth rate from 6 to 30 nm) of NPF events in C1 and C2. C1 had a higher formation rate and growth rate than C2 (the median J_6 and $GR_{6–30\ nm}$ values for C1 were 4.0 $cm^{-3}s^{-1}$ and 9.0 nm/h, respectively, when compared with the corresponding values of 1.5 $cm^{-3}s^{-1}$ and 8.0 nm/h for C2). The ranges of J_6 and $GR_{6–30\ nm}$ of C1 and C2 also overlapped with each other, which means that the NPF intensity at SORPES varied from low to high values in a continuous manner. As shown in Figures S5a and S5b, the formation rate mainly affected the PNSD below 10 nm, while the growth rate affected the peak at ~20 nm. When manually selecting the non-NPF period data (Figure S5c), the PNSD of C1 and C2 had lower values below 10 nm but a higher peak at ~20 nm. The intensity of vehicle emissions, which was indicated by the NO_x concentrations, mainly affected the peak at ~20 nm. Moreover, the PNSD of both C1 and C2 exhibited a second peak at ~100 nm, which represented urban background aerosols. Elevated particle number concentrations at ~100 nm were frequently observed not only at SORPES but also in other urban environments in China (Wu, et al., 2008; Xiao et al., 2015; Yue et al., 2013).

3.1.3. Cluster 3: Intensive Growth of Nucleated Particles

Cluster 3 (C3) accounted for 1.4% of the total measurements. The PNSD of C3 was unimodal with a peak in Aitken mode. The GMD of C3 was 38 nm, which was similar to C2. The Aitken mode particle number concentration of ~25,400 cm^{-3} in C3 was the highest among the seven clusters. Aitken mode particles fall in a size range that overlaps the nucleation and accumulation modes and could therefore originate from either primary emissions or from the growth of nucleation mode particles. The diurnal variations in the C3 occurrences exhibited maxima during the daytime but were slightly later than the maximum occurrence times of C1 (Figure 3). Therefore, it is likely that C3 results from the growth of nucleated C1 particles. C3 had atmospheric conditions similar to those of C1, although the temperatures and ozone concentrations were higher and the pollutant concentrations, such as $PM_{2.5}$, NO and NO_x, were lower. C3 had the lowest occurrence since it represented the initial growth stage of high-intensity NPF and occurred mostly after C1 with relatively short durations. C3 occurred mostly in the warm season, especially in spring and summer. The sulfate proportion for C3 reached 53.1% of the total water soluble ions, which was the highest percentage among the seven clusters. The combination of high solar radiation and O_3 concentrations leads to enhanced sulfate formation. As shown in Figure 5, unlike the other clusters, C3 can be influenced by air masses from South China, where BVOC emissions are high in the warm season. The oxidation of BVOCs can form highly oxidized multifunctional organic compounds with extremely low vapor pressures, which then contribute to the growth of newly formed particles (Ehn et al., 2014; Stolzenburg et al., 2018).

3.1.4. Cluster 4: Further Growth and Aged Vehicle Exhaust

Cluster 4 (C4) accounted for 9.1% of the total measurements. The PNSD of C4 exhibited a peak at 60 nm. The GMD of C4 was 56 nm, which was clearly larger than those of C1–C3. The occurrence frequencies of C4 peaked in the late afternoon, which was later than those of C1–C3. Therefore, C4 represents the further growth of nucleation mode particles. C4 had atmospheric conditions similar to those of C1–C3 but with relatively lower incoming solar radiation intensities (Figure 4). C4 was universally observed throughout the year, but with low frequencies in late January, February, and June (Figure 3). Moreover, the CO, NO,

and NO_x concentrations of C4 were higher than those of C1–C3, which indicated that aged vehicle exhaust emissions may also influence the PNSD of C4. For the aerosol compositions of C4 (Figure 6 and Table S2), sulfate (46.1%, 14.0 μg/m^3) dominated the water-soluble ions in $PM_{2.5,}$ which was similar to C1–C3. The water soluble ion mass concentrations of C4 were higher than those of C1–C3, which suggested the formation of secondary inorganic aerosols during aging.

Taken together, C1–C4 occurred mostly in the daytime and thus had higher temperatures and radiation intensities compared with the other clusters. Trace gases such as O_3 also followed the diurnal variations, with higher concentrations during the daytime. To exclude the effect of diurnal variations, an interval during the daytime period of 9:00-15:00 L.T. was selected to compare the atmospheric conditions of each cluster (Figure S6). The results were generally similar and indicated that diurnal variations were not the only reason for the different atmospheric conditions of each cluster.

3.1.5. Cluster 5: Humid YRD Urban Background Cluster

Cluster 5 (C5) was the most frequent cluster at SORPES and accounted for 37.9% of the total measurements. C5 exhibited a unimodal size distribution with a mode centered at 100 nm. The particle number concentration of C5 was the lowest among all of these clusters (e.g., 1,800 cm^{-3}, 4,100 cm^{-3}, and 3,500 cm^{-3} in the nucleation, Aitken and accumulation modes, respectively), which indicated that C5 represented a background urban aerosol size distribution in the YRD region. C5 was associated with high relative humidity (75.6%), which means that C5 occurred during cloudy or rainy days, when the wet removal efficiency was remarkable. According to Figure 5, the air masses mostly came from marine areas, which carried clean and wet air to the SORPES station. The concentrations of trace gas pollutants such as SO_2, CO, NO, and NO_x were lower in C5 and reflected the background conditions in the YRD region. However, as shown in Figure 4d and Table S2, the median and average $PM_{2.5}$ concentrations of C5 were not the lowest among the seven clusters. Although C5 had a $PM_{2.5}$ mass concentration similar to that of C4, the water soluble ion concentration of C5 was much higher, which suggested different sources and transformation processes of the particles in C5. As the highest relative humidity was in C5, heterogeneous reactions might play an important role in secondary inorganic aerosol formation (Ravishankara, 1997).

3.1.6. Cluster 6: Pollution From Fossil Fuel Combustion

Cluster 6 (C6) was the second most frequent cluster, which accounted for 34.1% of the total measurements. Similar to C5, the size distribution of C6 was unimodal with a peak at 100 nm. However, the number concentrations of all modes (e.g., 2,300 cm^{-3}, 7,700 cm^{-3} and 6,200 cm^{-3} in the nucleation, Aitken and accumulation modes, respectively) in C6 were higher than those in C5, which indicated that C6 was a more polluted cluster than C5. C6 occurred mostly at night during the presence of a shallower nocturnal boundary layer, and therefore, the pollutant levels were enriched. As a result, C6 was associated with high $PM_{2.5}$ levels and high concentrations of the gas pollutants such as NO, NO_x, and CO. The average $PM_{2.5}$ mass concentration was 82.4 μg/m^3, which was higher than the National Ambient Air Quality Standards in China (e.g., 75 μg/m^3 in annual average). According to the footprint shown in Figure 5, the pollutants mostly came from the YRD area, and fossil fuel combustion could be the main source. During the spring festival in early February, the lowest C6 occurrence was observed because of the closing of factories and decreased traffic, which further supported the view that anthropogenic activities were causing pollution. The nitrate proportion (e.g., 32.1%) was the highest among this cluster, with an average mass concentration of 19.4 μg/m^3. The daytime gas-phase oxidation of NO_2 by OH radicals under relatively low temperatures and nighttime N_2O_5 hydrolysis (Sun et al., 2018) could be reasons for the high nitrate proportion in this cluster.

3.1.7. Cluster 7: Heavy Pollution From Biomass Burning and Fossil Fuel Combustion

Cluster 7 (C7) accounted for 7.5% of the total measurements. C7 exhibited a unimodal size distribution and the highest number concentration of particles in the accumulation mode, which reached 11,300 cm^{-3}. The GMD of C7 was the largest among the seven clusters (e.g., 85 nm). This cluster was characterized by the lowest incoming solar radiation intensity and O_3 concentration, as well as the highest $PM_{2.5}$ concentrations (111.6 μg/m^3 on average), SO_2 (16.7 ppbv on average), CO (1.0 ppmv on average), NO (22.3 ppbv on average) and NO_x (53.5 ppbv on average). The high SO_2, CO and NO_x concentrations suggest that C7 was influenced by fossil fuel combustion, such as coal and petroleum fuel combustion. The average nitrate, sulfate and ammonium mass concentrations were 20.4 μg/m^3, 20.6 μg/m^3, and 13.6 μg/m^3, respectively, which

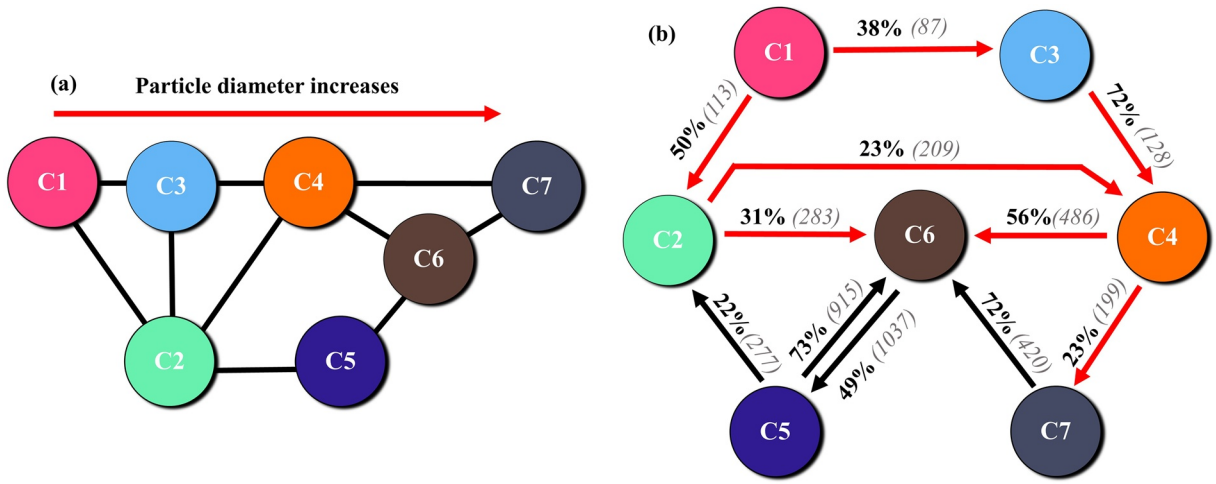

Figure 7. (a) Cluster proximity diagram and (b) "converted-to" flow chart. Note: the red arrows represent the increase in particle diameter. The frequencies and numbers of conversion are marked with arrows. Only frequencies larger than 20% are presented in Figure 7b.

were the highest among the seven clusters. The mass concentration of water soluble ions was comparable with those measured in other cities of China during haze pollution periods (An et al., 2019; Wang, Huang et al., 2017; Zhang et al., 2019). Similar to C6, C7 occurred mostly during nighttime. The wind speeds for C7 were the lowest among the seven clusters, which suppressed the diffusion of pollutants. Therefore, the heavy pollution of C7 may be related to fossil fuel combustion and accumulation under relatively stable atmospheric conditions. Moreover, this pollution could also be attributed to pollutant transport, such as from biomass burning. As shown in Figure 6 and Table S2, the mass concentration of potassium ions (K^+) in C7 was 2.7 $\mu g/m^{3,}$ and its fraction reached 4.2%, which was much higher than those in the other clusters and indicated the influence of biomass burning (Ding, Fu, Yang, Sun, Petäjä et al., 2013; Nie et al., 2015). As shown in Figure 3b, C7 occurred most frequently during the harvest period from late May to early June, which further supported the idea that biomass burning was one of the main reasons for the severe pollution in C7. The biomass burning plumes were transported to Nanjing from Anhui Province (to the west of Nanjing) as well as from the YRD area (Figure 5). The heavily polluted biomass burning plumes increased atmospheric stability and thereby enhanced local pollutant accumulation in Nanjing (Ding, Fu, Yang, Sun, Petäjä et al., 2013; Huang et al., 2018).

3.2. The Associations of Each Cluster

The cluster proximity diagram, which was obtained using the silhouette width (Beddows et al., 2009), is shown in Figure 7a. This diagram positions each cluster according to its similarity with the other clusters. Closer clusters represent those with the most similar characteristics, whereas more distant clusters represent those with more dissimilar characteristics. The average particle diameter of each cluster increases from left to right. The nucleation-related clusters, for example, C1 and C2, are positioned on the left side of the diagram. C3 and C4 are located next to the C1 and C2 clusters and connect the nucleation-related clusters (e.g., C1 and C2) with the urban pollution-related clusters (e.g., C6 and C7). Although the urban pollution-related clusters (e.g., C6 and C7) are positioned far from the nucleation-related clusters (e.g., C1 and C2), the contribution of nucleation to urban pollution is clear through the growth clusters (e.g., C3 and C4) as intermediate steps. C5, which represents the YRD urban background cluster, connects the nucleation-related cluster (C2) with the urban pollution-related cluster (C6).

To further study how each cluster evolves with time, the frequency of each cluster occurring after a certain cluster is presented in Figure 7b. The event occurrence frequency of each cluster is tabulated in Table S3, and the frequencies of certain clusters that are converted from another cluster are shown in Figure S7. Note that due to the very different event occurrence frequencies of each cluster (Table S3), the main pathways that describe the fates of the clusters (Figure 7) are somewhat different from the main pathways that describe their origins (Figure S7). C1 had 228 events in total (Table S3) and was mostly from C2, the weak NPF

cluster (Figure S7). As shown in Figure 7b, C1 could convert to C2 with a frequency of 50%, which indicated weakening of the nucleation. Moreover, C1 could convert to C3 with a frequency of 38%, and C3 could then convert continuously to C4 with a frequency of 72%. This conversion chain shows the typical process of NPF and growth with a "banana" shape (Figure S8a). C2 had 910 events (Table S3), and most of C2 came from C5 and C6 (Figure S7). Similar to the conversion of C1 to C4, 23% of C2 could convert to C4, which indicated the growth of nucleation mode particles (Figure S8b). During the measurements, there were 45 C1-C3-C4 NPF events and 70 C2-C4 NPF events in total. The C1-C3-C4 NPF events had higher formation and growth rates than the C2-C4 NPF events. The averaged J_6 and $GR_{6–30\ nm}$ values of the C1-C3-C4 NPF events were 3.8 s^{-1} cm^{-3} and 11.0 nm/h, respectively, compared with 2.8 s^{-1} cm^{-3} and 8.6 nm/h, respectively, for the C2-C4 NPF events.

After the nucleation-related clusters (C1, C2) convert to C4, 79% of C4 will convert to C6 (i.e., 56%) and C7 (i.e., 23%), which were the two clusters associated with haze in urban areas. As pollution-related clusters, C6 and C7 had 2,114 and 581 events, respectively. 23% of C6 and 34% of C7 came from C4 (Figure S7). The strong relationships among NPF events and pollution episodes indicated that NPF and growth could somehow contribute to pollution in urban environments. Figure S9 presents a typical case for the conversions from nucleation-related clusters to pollution-related clusters. Through a case study, Guo et al. (2014) also found that nucleation precedes pollution episodes in Beijing. Kulmala et al. (2021) reported that nearly all of the observed haze episodes originated from NPF in Beijing and found that the reduced growth rates could delay the buildup of haze episodes. The newly formed particles, which had different chemical compositions from those of accumulation mode particles, might facilitate heterogeneous reactions by providing distinct aerosol surfaces and volumes and thereby produce secondary particle masses (Kulmala et al., 2021). Despite the clear evidence, the physical and chemical mechanisms of for nucleation contributing to pollution in urban environments are still unclear.

The humid urban background cluster and two pollution clusters, for example, C5, C6 and C7, are closely related to each other. C5–C7 represent three levels of urban pollution, and therefore, the conversions between them are related to the changes in the PNSD of submicron particles. C5 had 1,246 events (Table S3), and most of C5 were from C6 (Figure S7). The conversion from C5 to C6 occurred at a frequency of 73% and represented the accumulation of pollutants, while the conversion of C6 to C5 occurred at a frequency of 49% and represented the gradual dissipation of pollution. C7, the heaviest pollution cluster, can convert to C6 with a frequency of 72% and involves the processes of atmospheric cleansing. Wet/dry deposition, evolution of the planetary boundary layer, and passing through a cold front can be causes of atmospheric cleansing.

3.3. Severe Haze Episodes in Cluster 6 and Cluster 7

C6 and C7 are two urban pollution-related clusters with high $PM_{2.5}$ concentrations but with different PNSD. This section focuses on the extremely severe haze episodes in C6 and C7 to investigate how the PNSD varies during air pollution in the urban YRD region. As shown in Figure 8, two typical episodes in C6 and C7, for example, Episode 1 from December 3 to 9, 2013, and Episode 2 from June 1 to 5, 2012, were selected for detailed case studies. Episode 1 was observed in the cold season, during which severe smog occurred frequently in East China (Ding, Fu, Yang, Sun, Zheng et al., 2013). The average $PM_{2.5}$ concentration during Episode 1 was 258 $\mu g/m^3$, and the maximum was 488 $\mu g/m^3$. As shown in Figure S10, the average nitrate, sulfate and ammonium concentrations during Episode 1 were 66.3 $\mu g/m^3$, 44.1 $\mu g/m^3$, and 37.5 $\mu g/m^3$, respectively, compared with the corresponding annual average values of ~15.8 $\mu g/m^3$, 16.8 $\mu g/m^3$, and 11.0 $\mu g/m^3$, respectively. The average and maximum CO concentrations reached 1.8 ppmv and 3.0 ppmv, respectively, which were much higher than the annual average value of ~0.7 ppmv. Therefore, fossil fuel combustion coupled with steady atmospheric stability and low boundary layer heights were presumably the main reasons for this regional-scale smog. Episode 2 occurred in early June during the harvest season in East China. Extremely high $PM_{2.5}$ concentrations were observed at night when the planetary boundary layer height was low (Figure S11). The average $PM_{2.5}$ concentration during Episode 2 was 117 $\mu g/m^3$, and the maximum $PM_{2.5}$ concentration was 364 $\mu g/m^3$. The average and maximum potassium ion (K^+) concentrations reached 5.3 $\mu g/m^3$ and 21.7 $\mu g/m^3$, respectively, which indicated that the severe pollution was mainly caused by straw burning.

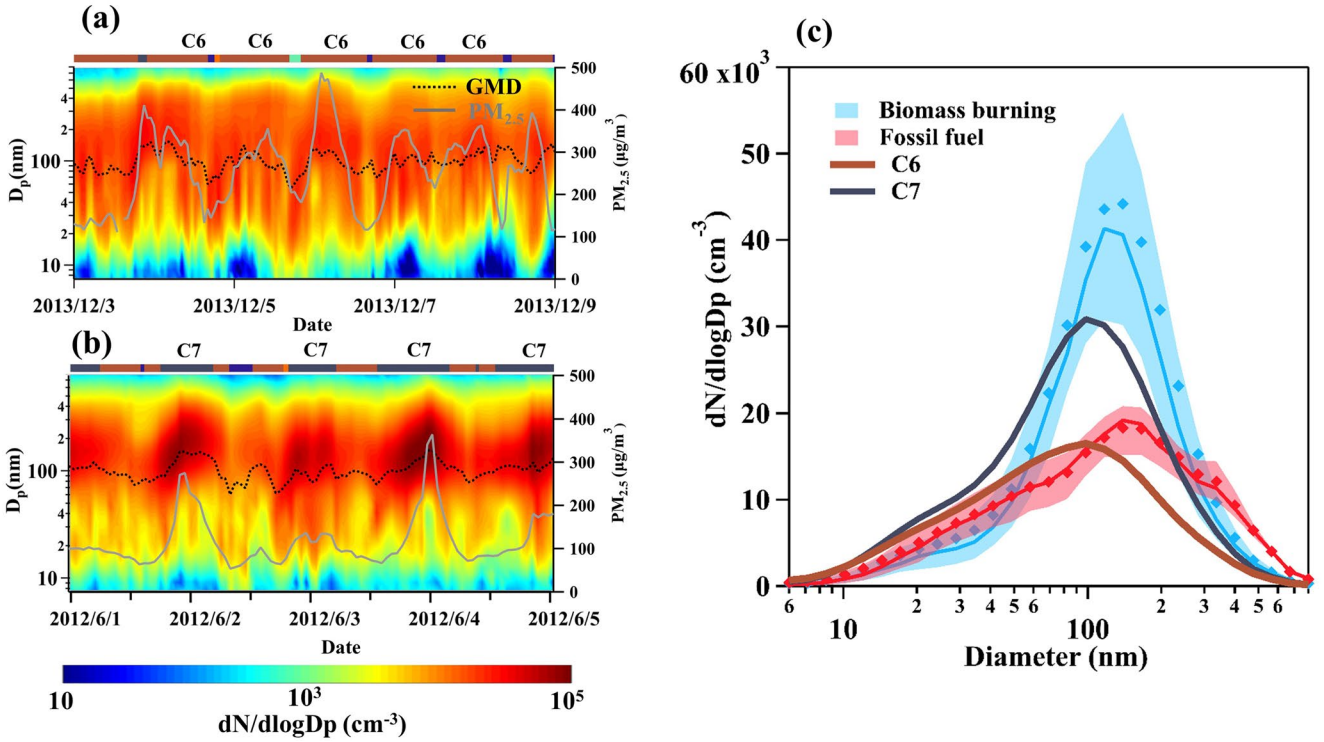

Figure 8. Variations in particle number size distributions, geometric mean diameters and $PM_{2.5}$ concentrations during severe haze pollution episodes in (a) Cluster 6 and (b) Cluster 7 and (c) the size distributions of fossil fuel episodes, biomass burning episodes, Cluster 6 and Cluster 7 (solid lines, diamond markers and shaded areas represent the median, mean and 25th-75th percentile ranges, respectively). Additional parameters, such as meteorological conditions, trace gases, and water soluble ions, during the episodes are presented in Figures S10 and S11.

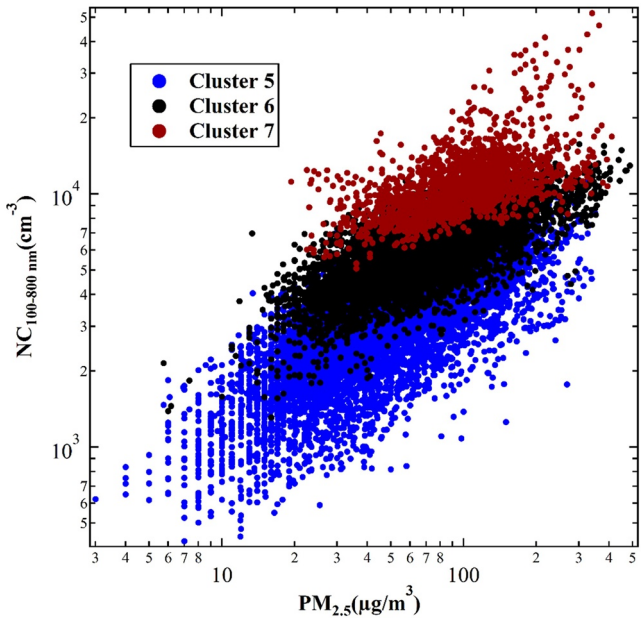

Figure 9. Scatter plot of $PM_{2.5}$ mass concentrations versus number concentrations of accumulation mode particles in Clusters 5–7.

Although both Episode 1 and Episode 2 represented extremely high $PM_{2.5}$ episodes, the PNSD was significantly different between these two episodes. As shown in Figure 8, most of the PNSD during Episode 1 belonged to C6, while the PNSD during Episode 2 belonged to C7 when the $PM_{2.5}$ concentrations were high. Different pollutant sources can cause completely different PNSD. Although the average $PM_{2.5}$ concentration during biomass burning pollution, such as for Episode 2, was lower than that during fossil fuel pollution, such as for Episode 1 (117 μg/m^3 vs. 258 μg/m^3), the number concentrations in the biomass burning plumes were much higher than those in the fossil fuel combustion plumes (26,000 cm^{-3} vs. 18,300 cm^{-3}). Figure 9 shows the relationships among the $PM_{2.5}$ mass concentrations and accumulation mode particle number concentrations for the urban background cluster (C5) and two pollution-related clusters (C6, C7). The accumulation mode particle number concentrations had a positive relationship with the $PM_{2.5}$ mass concentrations. C5-C7 exhibited relatively similar ranges of $PM_{2.5}$ mass concentrations, although the median $PM_{2.5}$ concentrations were different (Figure 4e). C7 clearly had a higher number concentration of accumulation mode particles than C6, whereas C5 had the lowest accumulation mode particle number concentration. Although the $PM_{2.5}$ mass concentrations were somewhat similar in C5, C6 and C7, the particle number concentrations of the accumulation mode particles could vary by more than one order of magnitude among these clusters. This was mainly because the $PM_{2.5}$ mass concentrations were dominated by supermicron particles (i.e., particles with sizes larger than 1 μm) whose contribution

to the particle number concentrations was negligible. However, since the number concentrations and particle sizes are directly related to the human health and climate effects of aerosols, in addition to the PM mass concentration, the PNSD is important for understanding the air quality in urban environments.

4. Conclusions

More than 3 years of continuous measurements of submicron PNSD (6–800 nm) were conducted at the SORPES station in the YRD region of East China. To understand the sources and transformations of submicron particles, a k-means clustering technique was applied, and seven clusters were categorized from the PNSD data set. The temporal occurrences of the PNSD, meteorological parameters and concentrations of pollutant gases, $PM_{2.5}$ and particle chemical composition were used to define and interpret each cluster.

C1 and C2, which accounting for 9.9% of the total PNSD data, were related to NPF and vehicle exhaust emissions, but they represented different intensities of these two sources. The submicron particles from NPF and vehicle exhaust emissions could not be separated because of the high particle growth rates and heavy traffic in the YRD. C3 and C4, which accounted for 10.5% of the total PNSD data, represented the growth of newly formed particles. C5, which accounted for 37.9% of the total PNSD data, was attributed to humid urban background conditions. C6 and C7, which accounted for 41.6% of the total PNSD data, were attributed to urban haze pollution. The cluster proximity diagram and conversion flow chart show the close relationships among the different clusters. The nucleation clusters (C1, C2) can influence the haze pollution clusters (C6, C7) through the growth clusters (C3, C4). The urban background cluster (C5) and haze pollution clusters (C6, C7) frequently convert to each other. According to the case studies of the selected severe haze episodes in C6 and C7, although the $PM_{2.5}$ mass concentrations during biomass burning pollution were lower than those during periods with haze from fossil fuel combustion, the particle number concentrations, especially the accumulation mode particle number concentrations, were much higher. A further analysis showed that although the $PM_{2.5}$ mass concentrations were somewhat similar, the particle number concentrations of the accumulation mode particles could vary by more than one order of magnitude from the urban background cluster (C5) to the pollution-related clusters (C6, C7).

In general, this study applied the cluster analysis technique to analyze a long-term PNSD data set from the YRD for the first time. The k-means cluster technique has been proven to be an effective tool to categorize PNSD datasets and can thereby elucidate the sources and transformations of atmospheric aerosols. Considering the increasing public concern regarding air quality, we highlight the importance of the PNSD to understand the air quality in urban environments and recommend clustering techniques for analyzing complex PNSD datasets.

Data Availability Statement

The land cover data in Figure 1 are from the Moderate Resolution Imaging Spectroradiometer (MODIS) Land Cover Type (MCD12Q1) Version 6 data product, which is available at https://doi.org/10.5067/MODIS/MCD12Q1.006. The SORPES measurement data used in this study are available at https://doi.org/10.5281/zenodo.4060571.

Acknowledgments
This study was supported by the National Key R&D Program of China (2016YFC0202000, 2018YFC0213800), National Natural Science Foundation of China (41805101), Natural Science Foundation of Jiangsu Province (BK20180351), International cooperation project of Jiangsu Provincial Science and Technology Agency (BZ2017066) and the Academy of Finland (311932).

References

Agudelo-Castaneda, D. M., Teixeira, E. C., Braga, M., Rolim, S. B. A., Silva, L. F. O., Beddows, D. C. S., et al. (2019). Cluster analysis of urban ultrafine particles size distributions. *Atmospheric Pollution Research*, *10*(1), 45–52. https://doi.org/10.1016/j.apr.2018.06.006

An, Z. S., Huang, R.-J., Zhang, R., Tie, X., Li, G., Cao, J., et al. (2019). Severe haze in northern China: A synergy of anthropogenic emissions and atmospheric processes. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(18), 8657–8666. https://doi.org/10.1073/pnas.1900125116

Asmi, A., Wiedensohler, A., Laj, P., Fjaeraa, A.-M., Sellegri, K., Birmili, W., et al. (2011). Number size distributions and seasonality of submicron particles in Europe 2008-2009. *Atmospheric Chemistry and Physics*, *11*(11), 5505–5538. https://doi.org/10.5194/acp-11-5505-2011

Atkinson, R. W., Kang, S., Anderson, H. R., Mills, I. C., & Walton, H. A. (2014). Epidemiological time series studies of $PM_{2.5}$ and daily mortality and hospital admissions: A systematic review and meta-analysis. *Thorax*, *69*(7), 660–665. https://doi.org/10.1136/thoraxjnl-2013-204492

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

Brines, M., Dall'Osto, M., Beddows, D. C. S., Harrison, R. M., Gómez-Moreno, F., Núñez, L., et al. (2015). Traffic and nucleation events as main sources of ultrafine particles in high-insolation developed world cities. *Atmospheric Chemistry and Physics*, *15*(10), 5929–5945. https://doi.org/10.5194/acp-15-5929-2015

Brines, M., Dall'Osto, M., Beddows, D. C. S., Harrison, R. M., & Querol, X. (2014). Simplifying aerosol size distributions modes simultaneously detected at four monitoring sites during SAPUSS. *Atmospheric Chemistry and Physics*, *14*(6), 2973–2986. https://doi.org/10.5194/acp-14-2973-2014

Chen, R., Hu, B., Liu, Y., Xu, J., Yang, G., Xu, D., & Chen, C. (2016). Beyond $PM_{2.5}$: The role of ultrafine particles on adverse health effects of air pollution. *Biochimica et Biophysica Acta (BBA) - General Subjects*, *1860*(12), 2844–2855. https://doi.org/10.1016/j.bbagen.2016.03.019

Cheng, Z., Wang, S., Fu, X., Watson, J. G., Jiang, J., Fu, Q., et al. (2014). Impact of biomass burning on haze pollution in the Yangtze River delta, China: A case study in summer 2011. *Atmospheric Chemistry and Physics*, *14*(9), 4573–4585. https://doi.org/10.5194/acp-14-4573-2014

Chu, B., Kerminen, V. M., Bianchi, F., Yan, C., Petäjä, T., & Kulmala, M. (2019). Atmospheric new particle formation in China. *Atmospheric Chemistry and Physics*, *19*(1), 115–138. https://doi.org/10.5194/acp-19-115-2019

Dai, L., Wang, H., Zhou, L., An, J., Tang, L., Lu, C., et al. (2017). Regional and local new particle formation events observed in the Yangtze River Delta region, China. *Journal of Geophysical Research: Atmospheres*, *122*(4), 2389–2402. https://doi.org/10.1002/2016jd026030

Dall'Osto, M., Beddows, D. C. S., Pey, J., Rodriguez, S., Alastuey, A., Harrison, R. M., & Querol, X. (2012). Urban aerosol size distributions over the Mediterranean city of Barcelona, NE Spain. *Atmospheric Chemistry and Physics*, *12*(22), 10693–10707. https://doi.org/10.5194/acp-12-10693-2012

Ding, A., Huang, X., Nie, W., Chi, X., Xu, Z., Zheng, L., et al. (2019). Significant reduction of $PM_{2.5}$ in eastern China due to regional-scale emission control: Evidence from SORPES in 2011-2018. *Atmospheric Chemistry and Physics*, *19*(18), 11791–11801. https://doi.org/10.5194/acp-19-11791-2019

Ding, A. J., Fu, C. B., Yang, X. Q., Sun, J. N., Petäjä, T., Kerminen, V.-M., et al. (2013). Intense atmospheric pollution modifies weather: A case of mixed biomass burning with fossil fuel combustion pollution in eastern China. *Atmospheric Chemistry and Physics*, *13*(20), 10545–10554. https://doi.org/10.5194/acp-13-10545-2013

Ding, A. J., Fu, C. B., Yang, X. Q., Sun, J. N., Zheng, L. F., Xie, Y. N., et al. (2013). Ozone and fine particle in the western Yangtze River Delta: An overview of 1 yr data at the SORPES station. *Atmospheric Chemistry and Physics*, *13*(11), 5813–5830. https://doi.org/10.5194/acp-13-5813-2013

Ding, A. J., Nie, W., Huang, X., Chi, X., Sun, J., Kerminen, V.-M., et al. (2016). Long-term observation of air pollution-weather/climate interactions at the SORPES station: A review and outlook. *Frontiers of Environmental Science & Engineering*, *10*(5), 15. https://doi.org/10.1007/s11783-016-0877-3

Ding, A. J., Wang, T., & Fu, C. B. (2013). Transport characteristics and origins of carbon monoxide and ozone in Hong Kong, South China. *Journal of Geophysical Research: Atmospheres*, *118*(16), 9475–9488. https://doi.org/10.1002/jgrd.50714

Downward, G. S., van Nunen, E. J. H. M., Kerckhoffs, J., Vineis, P., Brunekreef, B., Boer, J. M. A., et al. (2018). Long-term exposure to ultrafine particles and incidence of cardiovascular and cerebrovascular disease in a prospective study of a Dutch cohort. *Environmental Health Perspectives*, *126*(12), 8. https://doi.org/10.1289/ehp3047

Du, J. F., Cheng, T., Zhang, M., Chen, J., He, Q., Wang, X., et al. (2012). Aerosol size spectra and particle formation events at urban Shanghai in Eastern China. *Aerosol and Air Quality Research*, *12*(6), 1362–1372. https://doi.org/10.4209/aaqr.2011.12.0230

Ehn, M., Thornton, J. A., Kleist, E., Sipilä, M., Junninen, H., Pullinen, I., et al. (2014). A large source of low-volatility secondary organic aerosol. *Nature*, *506*(7489), 476–479. https://doi.org/10.1038/nature13032

Fan, J. W., Rosenfeld, D., Zhang, Y., Giangrande, S. E., Li, Z., Machado, L. A. T., et al. (2018). Substantial convection and precipitation enhancements by ultrafine aerosol particles. *Science*, *359*(6374), 411–418. https://doi.org/10.1126/science.aan8461

Gao, J., Wang, T., Zhou, X. H., Wu, W. S., & Wang, W. X. (2009). Measurement of aerosol number size distributions in the Yangtze River delta in China: Formation and growth of particles under polluted conditions. *Atmospheric Environment*, *43*(4), 829–836. https://doi.org/10.1016/j.atmosenv.2008.10.046

Guo, S., Hu, M., Peng, J., Wu, Z., Zamora, M. L., Shang, D., et al. (2020). Remarkable nucleation and growth of ultrafine particles from vehicular exhaust. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(7), 3427–3432. https://doi.org/10.1073/pnas.1916366117

Guo, S., Hu, M., Zamora, M. L., Peng, J., Shang, D., Zheng, J., et al. (2014). Elucidating severe urban haze formation in China. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(49), 17373–17378. https://doi.org/10.1073/pnas.1419604111

Harrison, R. M., Beddows, D. C. S., Alam, M. S., Singh, A., Brean, J., Xu, R., et al. (2019). Interpretation of particle number size distributions measured across an urban area during the FASTER campaign. *Atmospheric Chemistry and Physics*, *19*(1), 39–55. https://doi.org/10.5194/acp-19-39-2019

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm, *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *28*(1), 100–108. https://doi.org/10.2307/2346830

Heal, M. R., Kumar, P., & Harrison, R. M. (2012). Particles, air quality, policy and health. *Chemical Society Reviews*, *41*(19), 6606–6630. https://doi.org/10.1039/c2cs35076a

Herrmann, E., Ding, A. J., Kerminen, V.-M., Petäjä, T., Yang, X. Q., Sun, J. N., et al. (2014). Aerosols and nucleation in eastern China: First insights from the new SORPES-NJU station. *Atmospheric Chemistry and Physics*, *14*(4), 2169–2183. https://doi.org/10.5194/acp-14-2169-2014

Huang, X., Wang, Z., & Ding, A. (2018). Impact of aerosol-PBL interaction on haze pollution: Multiyear observational evidences in North China. *Geophysical Research Letters*, *45*(16), 8596–8603. https://doi.org/10.1029/2018gl079239

IPCC (2013). *Climate change 2013: The Physical Science basis. Contribution of working group I to the fifth assessment report of the Intergovernmental Panel on Climate Change* (p. 1535). Cambridge, United Kingdom and New York, NY, USA: Cambridge University Press. https://doi.org/10.1017/CBO9781107415324

Kerminen, V. M., Chen, X. M., Vakkari, V., Petaja, T., Kulmala, M., & Bianchi, F. (2018). Atmospheric new particle formation and growth: Review of field observations. *Environmental Research Letters*, *13*(10), 38. https://doi.org/10.1088/1748-9326/aadf3c

Kerminen, V. M., Paramonov, M., Anttila, T., Riipinen, I., Fountoukis, C., Korhonen, H., et al. (2012). Cloud condensation nuclei production associated with atmospheric nucleation: A synthesis based on existing literature and new results. *Atmospheric Chemistry and Physics*, *12*(24), 12037–12059. https://doi.org/10.5194/acp-12-12037-2012

Knibbs, L. D., Cole-Hunter, T., & Morawska, L. (2011). A review of commuter exposure to ultrafine particles and its health effects. *Atmospheric Environment*, *45*(16), 2611–2622. https://doi.org/10.1016/j.atmosenv.2011.02.065

Kulmala, M., Dada, L., Daellenbach, K. R., Yan, C., Stolzenburg, D., Kontkanen, J., et al. (2021). Is reducing new particle formation a plausible solution to mitigate particulate air pollution in Beijing and other Chinese megacities? *Faraday Discussions*, *226*, 334–347. https://doi.org/10.1039/D0FD00078G

Kulmala, M., Kerminen, V. M., Petaja, T., Ding, A. J., & Wang, L. (2017). Atmospheric gas-to-particle conversion: Why NPF events are observed in megacities? *Faraday Discussions*, *200*, 271–288. https://doi.org/10.1039/c6fd00257a

Kulmala, M., Luoma, K., Virkkula, A., Petäjä, T., Paasonen, P., Kerminen, V.-M., et al. (2016). On the mode-segregated aerosol particle number concentration load: Contributions of primary and secondary particles in Hyytiala and Nanjing. *Boreal Environment Research*, *21*(3–4), 319–331.

Li, L., Chen, C. H., Fu, J. S., Huang, C., Streets, D. G., Huang, H. Y., et al. (2011). Air quality and emissions in the Yangtze River Delta, China. *Atmospheric Chemistry and Physics*, *11*(4), 1621–1639. https://doi.org/10.5194/acp-11-1621-2011

Ling, Y., Wang, Y., Duan, J., Xie, X., Liu, Y., Peng, Y., et al. (2019). Long-term aerosol size distributions and the potential role of volatile organic compounds (VOCs) in new particle formation events in Shanghai. *Atmospheric Environment*, *202*, 345–356. https://doi.org/10.1016/j.atmosenv.2019.01.018

Liu, Z. R., Hu, B., Liu, Q., Sun, Y., & Wang, Y. S. (2014). Source apportionment of urban fine particle number concentration during summertime in Beijing. *Atmospheric Environment*, *96*, 359–369. https://doi.org/10.1016/j.atmosenv.2014.06.055

Masiol, M., Harrison, R. M., Vu, T. V., & Beddows, D. C. S. (2017). Sources of sub-micrometre particles near a major international airport. *Atmospheric Chemistry and Physics*, *17*(20), 12379–12403. https://doi.org/10.5194/acp-17-12379-2017

Nie, W., Ding, A. J., Xie, Y. N., Xu, Z., Mao, H., Kerminen, V.-M., et al. (2015). Influence of biomass burning plumes on HONO chemistry in eastern China. *Atmospheric Chemistry and Physics*, *15*(3), 1147–1159. https://doi.org/10.5194/acp-15-1147-2015

Peng, J. F., Hu, M., Wang, Z. B., Huang, X. F., Kumar, P., Wu, Z. J., et al. (2014). Submicron aerosols at thirteen diversified sites in China: Size distribution, new particle formation and corresponding contribution to cloud condensation nuclei production. *Atmospheric Chemistry and Physics*, *14*(18), 10249–10265. https://doi.org/10.5194/acp-14-10249-2014

Qi, X. M., Ding, A. J., Nie, W., Petäjä, T., Kerminen, V.-M., Herrmann, E., et al. (2015). Aerosol size distribution and new particle formation in the western Yangtze River Delta of China: 2 years of measurements at the SORPES station. *Atmospheric Chemistry and Physics*, *15*(21), 12445–12464. https://doi.org/10.5194/acp-15-12445-2015

Ravishankara, A. R. (1997). Heterogeneous and multiphase chemistry in the troposphere. *Science*, *276*(5315), 1058–1065. https://doi.org/10.1126/science.276.5315.1058

Rönkkö, T., Kuuluvainen, H., Karjalainen, P., Keskinen, J., Hillamo, R., Niemi, J. V., et al. (2017). Traffic is a major source of atmospheric nanocluster aerosol. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(29), 7549–7554. https://doi.org/10.1073/pnas.1700830114

Rosenfeld, D., Zhu, Y., Wang, M., Zheng, Y., Goren, T., & Yu, S. (2019). Aerosol-driven droplet concentrations dominate coverage and water of oceanic low-level clouds. *Science*, *363*, eaav0566. https://doi.org/10.1126/science.aav0566

Sabaliauskas, K., Jeong, C.-H., Yao, X., Jun, Y.-S., & Evans, G. (2013). Cluster analysis of roadside ultrafine particle size distributions. *Atmospheric Environment*, *70*, 64–74. https://doi.org/10.1016/j.atmosenv.2012.12.025

Salimi, F., Ristovski, Z., Mazaheri, M., Laiman, R., Crilley, L. R., He, C., et al. (2014). Assessment and application of clustering techniques to atmospheric particle number size distribution for the purpose of source apportionment. *Atmospheric Chemistry and Physics*, *14*(21), 11883–11892. https://doi.org/10.5194/acp-14-11883-2014

Salma, I., Borsos, T., Weidinger, T., Aalto, P., Hussein, T., Dal Maso, M., & Kulmala, M. (2011). Production, growth and properties of ultrafine atmospheric aerosol particles in an urban environment. *Atmospheric Chemistry and Physics*, *11*(3), 1339–1353. https://doi.org/10.5194/acp-11-1339-2011

Salma, I., Fueri, P., Nemeth, Z., Balashazy, I., Hofmann, W., & Farkas, A. (2015). Lung burden and deposition distribution of inhaled atmospheric urban ultrafine particles as the first step in their health risk assessment. *Atmospheric Environment*, *104*, 39–49. https://doi.org/10.1016/j.atmosenv.2014.12.060

Schmale, J., Henning, S., Decesari, S., Henzing, B., Keskinen, H., Sellegri, K., et al. (2018). Long-term cloud condensation nuclei number concentration, particle number size distribution and chemical composition measurements at regionally representative observatories. *Atmospheric Chemistry and Physics*, *18*(4), 2853–2881. https://doi.org/10.5194/acp-18-2853-2018

Shen, Y., Virkkula, A., Ding, A., Luoma, K., Keskinen, H., Aalto, P. P., et al. (2019). Estimating cloud condensation nuclei number concentrations using aerosol optical properties: Role of particle number size distribution and parameterization. *Atmospheric Chemistry and Physics*, *19*(24), 15483–15502. https://doi.org/10.5194/acp-19-15483-2019

Stolzenburg, D., Fischer, L., Vogel, A. L., Heinritzi, M., Schervish, M., Simon, M., et al. (2018). Rapid growth of organic aerosol nanoparticles over a wide tropospheric temperature range. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(37), 9122–9127. https://doi.org/10.1073/pnas.1807604115

Sun, P., Nie, W., Chi, X., Xie, Y., Huang, X., Xu, Z., et al. (2018). Two years of online measurement of fine particulate nitrate in the western Yangtze River Delta: Influences of thermodynamics and N_2O_5 hydrolysis. *Atmospheric Chemistry and Physics*, *18*(23), 17177–17190. https://doi.org/10.5194/acp-18-17177-2018

Tie, X. X., & Cao, J. J. (2009). Aerosol pollution in China: Present and future impact on environment. *Particuology*, *7*(6), 426–431. https://doi.org/10.1016/j.partic.2009.09.003

Väänänen, R., Kyrö, E.-M., Nieminen, T., Kivekäs, N., Junninen, H., Virkkula, A., et al. (2013). Analysis of particle size distribution changes between three measurement sites in northern Scandinavia. *Atmospheric Chemistry and Physics*, *13*(23), 11887–11903. https://doi.org/10.5194/acp-13-11887-2013

Wang, Y. C., Huang, R.-J., Ni, H. Y., Chen, Y., Wang, Q. Y., Li, G. H., et al. (2017). Chemical composition, sources and secondary processes of aerosols in Baoji city of northwest China. *Atmospheric Environment*, *158*, 128–137. https://doi.org/10.1016/j.atmosenv.2017.03.026

Wang, Z. B., Hu, M., Sun, J. Y., Wu, Z. J., Yue, D. L., Shen, X. J., et al. (2013). Characteristics of regional new particle formation in urban and regional background environments in the North China Plain. *Atmospheric Chemistry and Physics*, *13*(24), 12495–12506. https://doi.org/10.5194/acp-13-12495-2013

Wang, Z. B., Wu, Z., Yue, D., Shang, D., Guo, S., Sun, J., et al. (2017). New particle formation in China: Current knowledge and further directions. *The Science of the Total Environment*, *577*, 258–266. https://doi.org/10.1016/j.scitotenv.2016.10.177

Wegner, T., Hussein, T., Hameri, K., Vesala, T., Kulmala, M., & Weber, S. (2012). Properties of aerosol signature size distributions in the urban environment as derived by cluster analysis. *Atmospheric Environment*, *61*, 350–360. https://doi.org/10.1016/j.atmosenv.2012.07.048

Wiedensohler, A., Birmili, W., Nowak, A., Sonntag, A., Weinhold, K., Merkel, M., et al. (2012). Mobility particle size spectrometers: Harmonization of technical standards and data structure to facilitate high quality long-term observations of atmospheric particle number size distributions. *Atmospheric Measurement Techniques*, *5*(3), 657–685. https://doi.org/10.5194/amt-5-657-2012

Wu, G. Y., Brown, J., Zamora, M. L., Miller, A., Satterfield, M. C., Meininger, C. J., et al. (2019). Adverse organogenesis and predisposed long-term metabolic syndrome from prenatal exposure to fine particulate matter. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(24), 11590–11595. https://doi.org/10.1073/pnas.1902925116

Wu, Z. J., Hu, M., Lin, P., Liu, S., Wehner, B., & Wiedensohler, A. (2008). Particle number size distribution in the urban atmosphere of Beijing, China. *Atmospheric Environment*, *42*(34), 7967–7980. https://doi.org/10.1016/j.atmosenv.2008.06.022

Xiao, S., Wang, M. Y., Yao, L., Kulmala, M., Zhou, B., Yang, X., et al. (2015). Strong atmospheric new particle formation in winter in urban Shanghai, China, *Atmospheric Chemistry and Physics*, *15*(4), 1769–1781. https://doi.org/10.5194/acp-15-1769-2015

Xie, Y. N., Ding, A., Nie, W., Mao, H., Qi, X., Huang, X., et al. (2015). Enhanced sulfate formation by nitrogen dioxide: Implications from in situ observations at the SORPES station. *Journal of Geophysical Research: Atmospheres*, *120*(24), 12679–12694. https://doi.org/10.1002/2015jd023607

Xu, X., Lin, W., Wang, T., Yan, P., Tang, J., Meng, Z., & Wang, Y. (2008). Long-term trend of surface ozone at a regional background station in eastern China 1991-2006: Enhanced variability. *Atmospheric Chemistry and Physics*, *8*(10), 2595–2607. https://doi.org/10.5194/acp-8-2595-2008

Yao, L., Garmash, O., Bianchi, F., Zheng, J., Yan, C., Kontkanen, J., et al. (2018). Atmospheric new particle formation from sulfuric acid and amines in a Chinese megacity. *Science*, *361*(6399), 278–281. https://doi.org/10.1126/science.aao4839

Yue, D. L., Hu, M., Wang, Z. B., Wen, M. T., Guo, S., Zhong, L. J., et al. (2013). Comparison of particle number size distributions and new particle formation between the urban and rural sites in the PRD region, China. *Atmospheric Environment*, *76*, 181–188. https://doi.org/10.1016/j.atmosenv.2012.11.018

Zhang, X. Y., Zhao, X., Ji, G. X., Ying, R. R., Shan, Y. H., & Lin, Y. S. (2019). Seasonal variations and source apportionment of water-soluble inorganic ions in $PM_{2.5}$ in Nanjing, a megacity in southeastern China. *Journal of Atmospheric Chemistry*, *76*(1), 73–88. https://doi.org/10.1007/s10874-019-09388-z