The Fundamental Aerosol Models Over China Region: A Cluster Analysis of the Ground-Based Remote Sensing Measurements of Total Columnar Atmosphere

Zhengqiang Li, Ying Zhang, Hua Xu, Kaitao Li, Oleg Dubovik, Philippe Goloub

AGU100 ADVANCING EARTH AND SPACE SCIENCE

Geophysical Research Letters

RESEARCH LETTER
10.1029/2019GL082056

Key Points:
- Ten fundamental aerosol models in China are derived from a cluster study based on the ground-based remote sensing measurements of SONET
- Five fine-particle aerosol models include urban polluted, secondary polluted, combined polluted, polluted fly ash, and continental background
- Five coarse-particle aerosol models include summer fly ash, winter fly ash, primary dust, transported dust, and background dust

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. Zhang and H. Xu,
zhang_ying@aircas.ac.cn;
xuhua@aircas.ac.cn

Citation:
Li, Z., Zhang, Y., Xu, H., Li, K., Dubovik, O., & Goloub, P. (2019). The fundamental aerosol models over China region: A cluster analysis of the ground-based remote sensing measurements of total columnar atmosphere. *Geophysical Research Letters*, *46*, 4924–4932. https://doi.org/10.1029/2019GL082056

Received 23 JAN 2019
Accepted 3 APR 2019
Accepted article online 10 APR 2019
Published online 1 MAY 2019

The Fundamental Aerosol Models Over China Region: A Cluster Analysis of the Ground-Based Remote Sensing Measurements of Total Columnar Atmosphere

Zhengqiang Li[1], Ying Zhang[1], Hua Xu[1], Kaitao Li[1], Oleg Dubovik[2], and Philippe Goloub[2]

[1]State Environmental Protection Key Laboratory of Satellite Remote Sensing, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing, China, [2]Laboratoire d'Optique Atmosphérique, CNRS/Université Lille, Villeneuve d'Ascq, Lille, France

Abstract Ten fundamental aerosol models in China are derived from a cluster study based on the ground-based remote sensing measurements of Sun-sky radiometer Observation NETwork. The aerosol size distribution decomposition techniques are employed to yield individual fine and coarse mode size distribution functions with independent refractive indices. The total 10,773 records containing 18 kinds of aerosol microphysical parameters are used to yield 10 typical clusters with the verification of clustering robustness. Ten clusters suggest five typical fine particle aerosol models including urban polluted, secondary polluted, combined polluted, polluted fly ash, and continental background, as well as five coarse models including summer fly ash, winter fly ash, primary dust, transported dust, and background dust over China region. The representativeness and coappearance analyses again reveal five dominative aerosol patterns on the base of fundamental models. These models can be used in the chemical model simulation, satellite remote sensing, climate, and environment analyses.

Plain Language Summary The aerosol fundamental characterization model is used to describe the typical aerosol properties in a region which is essential in the radiative transfer calculations and thus is fundamental knowledge for many applications related to Earth's atmosphere. However, the establishment of these models is not easy not only for the hard works on the continuously long-term, high-quality observation but also related to analysis methods. However, these essential models are long-term missed for China regions where is the most important aerosol regions in the world. It has turned to a bottleneck problem for accurately assessing aerosol effects in this important region. In this paper, we use the first-hand 7-year aerosol observation from the 16 sites of Chinese local observation network and employ an advanced new technique to first derive 10 fundamental aerosol characterization models over China. The obtained 10 fundamental aerosol models can provide a basis for general description of China aerosols. It is very useful for the climate, environment, and ecology studies. Meanwhile, they can be directly adapted in the development of satellite remote sensing algorithms. These results can be quickly applied to ameliorate accuracy and validity of various applications related to radiative transfer through the atmosphere in China.

1. Introduction

Atmospheric aerosols have a major influence on the radiation energy balance of the earth-atmosphere system, which makes the knowledge of the aerosol radiative forcing a critical factor to improve accuracy of climate change assessment (Boucher et al., 2013). To evaluate these important effects of aerosols, the typical or representative aerosol characterization over large areas (e.g., regional, national, or global) is needed and essentially supports atmospheric chemical model simulation and assessments of climate and environment. The establishment of typical aerosol characterization models based on continuous remote sensing observation is therefore a fundamental work for columnar atmospheric researches (Hansen, 2018). In this field, some global typical aerosol characterization models have been established from ground-based sun-sky radiometer measurements (e.g., Dubovik et al., 2002) and are widely used. However, due to the aerosol complexity and lack of observation in China (Che et al., 2018), the models are not yet well established.

Many studies (Bellouin et al., 2005; Eck et al., 2012; Gobbi et al., 2007; Lee et al., 2010; Levy et al., 2007) have performed the classification of aerosols based on multiple observations with a goal of establishing typical

aerosol models. Dubovik et al. (2002) analyzed characterization of global typical aerosol observation sites based on empirical classification, yielding aerosol models of urban industrial, biomass burning, dust, and oceanic types. It should be noted that because of the assumption of size-independent index of refraction in the Aerosol Robotic Network (AERONET) information, Dubovik et al. (2002) has restricted analysis for cases that are either fine- or coarse-mode or one-aerosol-type dominated to derive dust or smoke properties. Omar et al. (2005) also analyzed the global observation of the AERONET with a clustering approach, and the obtained six typical aerosol models are applied in the satellite remote sensing, for example, the Cloud-Aerosol Lidar with Orthogonal Polarization. For the Asia region, Wu and Zeng (2014) utilized a Gustafson-Kessel fuzzy clustering algorithm to analyze dust aerosol types. Chen et al. (2013) and Zhang, Xu, and Zheng (2017) studied aerosol types in China regions using clustering analysis, while Ma et al. (2016) focused on aerosol characterization models at Song mountain regions in central China. Compared with empirical statistics analysis, the clustering approach has advantages of nonsupervision and thus has drawn more attentions in the establishment of aerosol characterization models.

In this study, based on the long-term observations of local sun-sky radiometer network, together with the individual fine-/coarse-mode aerosol products (Zhang, Li, Zhang, et al., 2017), we performed a clustering analysis of 7-year aerosol observations of 16 sites over China to yield typical aerosol characterization models for this region. The second section introduces the data sources and the clustering method, while the third section presents the clustering results of 10 fundamental aerosol models. The fourth section discusses the robustness and representativeness of our classification and conclusions are summarized in the last section. It should be emphasized that this is the first time to obtain the representative aerosol models for the columnar atmosphere over entire China region, by initiatively using SONET sites and considering the long-term observation is very rare over this region and the existed AERONET sites are not enough to conduce a geologically covering analysis.

2. Data and Method

2.1. Remote Sensing Observation

The Sun-sky radiometer Observation NETwork (SONET) is a Chinese local observation network performing ground-based remote sensing measurements of aerosols (Li et al., 2018). It currently has 16 permanent observation sites distributed in China (Figure 1), covering most of typical aerosol geological (e.g., plateau, desert, mountain, hill, plain, and island) and humanistic (e.g., rural, urban, megacity, background, and polluted region) features. As seen from Figure 1, six among total 16 sites possess observation longer than 5 years, which provide a solid basis for the analysis of typical aerosol characterization over China.

2.2. Aerosol Measurement Data

The SONET is equipped with multiband polarized Sun-sky radiometer (CE318-DP) to measure direct-sun and diffused sky light with polarization. SONET data with polarization measurements have been used in the research algorithms to derive size distribution and refractive index for both fine- and coarse-mode particles simultaneously in high aerosol optical depth (AOD) conditions where both fine- and coarse-mode aerosol loadings are significant (Xu et al., 2015). However, in the operational mode, only SONET sky radiances are used in the inversion to yield aerosol optical and microphysical parameters (e.g., AOD, size distribution, and complex refractive indices) using standard AERONET retrieval algorithm (Dubovik et al., 2000, 2006), and the inversion results formulate the database for this study. The calibration and cloud-screening procedures are also consistent between SONET and AERONET, and thus, the data quality and aerosol product accuracy are comparable as suggested by data analyses (Li et al., 2018) of joint campaigns (Holben et al., 2018). The SONET data set has been employed in many applications, including validation of satellite products (e.g., Zhang, Li, Qie, et al., 2017), assimilation into chemical transport models (e.g., Chang et al., 2015), and radiative forcing studies (e.g., Li et al., 2018).

Long-term observation (e.g., Dubovik et al., 2002) shows that aerosols of the total columnar atmosphere exhibit a bimodal volume particle size distribution (VPSD) in many cases. The cutting-off size is around 1 nm (in particle radius r) which separates VPSD into fine particle ($r < \sim 1$ nm) and coarse particle ($r > \sim 1$ nm) domains. Moreover, these VPSD curves usually fit well log-normal functions and can be expressed as

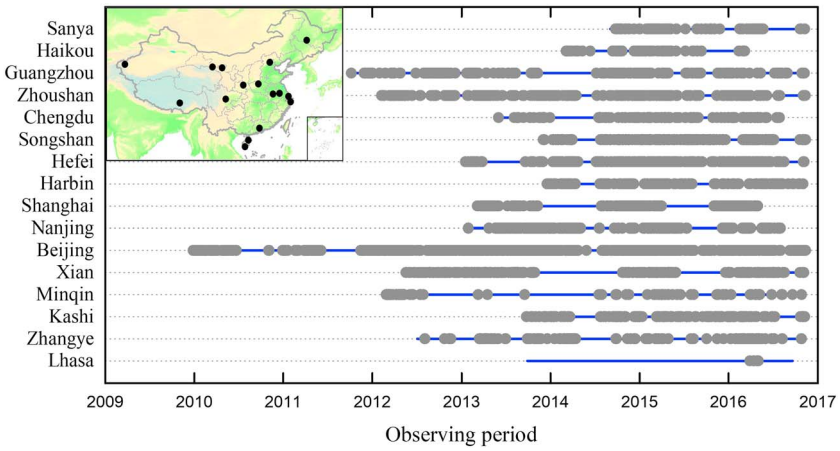

Figure 1. The geological distribution and time series of data collection of 16 Sun-sky radiometer Observation NETwork sites (blue thin lines = aerosol optical depth parameter; gray thick lines = inversion parameters, e.g., size distribution).

$$\frac{dV(r)}{dlnr} = \sum_{i=1}^{m} \frac{C_i}{\sqrt{2\pi}\left(\ln\sigma_i\right)^2} \exp\left[-\frac{1}{2}\left(\frac{lnr - \ln r_i}{\ln\sigma_i}\right)^2\right] \quad \mathrm{m} = 1, 2, \tag{1}$$

where C_i denotes the particle volume concentration, r_i is the median radius for volume, σ_i is the standard deviation, and $dV/dlnr$ is the VPSD in the unit of cubic micrometer per square micrometer. In practice, among the fine- and coarse-particle size domain, there usually exists multipeaks (>2) in VPSD (Eck et al., 2012). Scholars have suggested to mark these peaks as the standard fine mode (subscript $_{\mathrm{f}}$), the standard coarse mode (subscript $_{\mathrm{c}}$), the submicron fine mode (subscript $_{\mathrm{SMF}}$, i.e., particle radius larger than the standard fine mode but still in the fine particle size domain), and the super-micron coarse mode (subscript $_{\mathrm{SMC}}$, i.e., particle radius smaller than the standard coarse mode but still in the coarse particle size domain; Li et al., 2014; Zhang et al., 2016). Zhang, Li, Zhang, et al. (2017) method (see Text S1 in the supporting information) is used to provide the individual fine-/coarse-mode aerosol products (as listed in Table 1) from instantaneous measurements as the inputs of clustering analyses of this study. Considering that the optical parameters (e.g., AOD and single scattering albedo) can be derived from aerosol microphysical parameters, we only employ aerosol microphysical parameters in the clustering analyses.

2.3. Clustering Analysis Method

We employ the K-means clustering approach (Anderberg, 1973; Duda et al., 2001; Hartigan, 1975; Jain & Dubes, 1988; Sokal & Sneath, 1963) to classify total 18 kinds of aerosol microphysical property parameters (as shown in Table 1, r_{f}, r_{SMF}, σ_{f}, σ_{SMF}, C_{f}, C_{SMF}, n_{f}, $k_{\mathrm{f,\ 440}}$, k_{f} for fine particles; r_{c}, r_{SMC}, σ_{c}, σ_{SMC}, C_{c}, C_{SMC}, n_{c}, $k_{\mathrm{c,\ 440}}$, k_{c} for coarse particles, and the number subscript denoting the wavelength of 440 nm). The K-means method deals with the data set consisting of M n-dimensional vector. When element number M is greater than the designated cluster number k (i.e., M > k), the data set can be classified into k groups, that is, $S = \{S_1, S_2, \ldots, S_k\}$. Mathematically, it converts into a problem of finding the group center values μ when the total distance between all data set elements and all group's center is minimized:

$$J = \min_{S} \sum_{i=1}^{k} \sum_{x_j \in S_i} \left\| x_j - \mu_i \right\|^2, \tag{2}$$

where J is the dissimilarity, x is the element value, and μ is the value of a group center. Given the group number k, each group center is randomly selected. The dissimilarity between all element and all group centers are calculated. Then (a) the element is classified into a cluster when the dissimilarity is the lowest; (b) the group center value is recalculated by average of the new cluster elements; (c) above steps are iterated until the J value reaches the minimum. Because the physical unit spaces of n-dimensional element are different, the normalization by standard deviation is necessary. We use the standardized Euclidean distance to cluster the data (see Text S2), assuming the parameters of fine mode and coarse mode are irrelevant.

Table 1
Parameters and Characteristics of the Representative Clusters of the Sun-Sky Radiometer Observation Network Data Set

Fine modes												
Type	r_f	r_{SMF}	σ_f	σ_{SMF}	C_f	C_{SMF}	n_f	$k_{f,440}$	k_f	Ratio (%)	Size	Ref. ind.	Absorptivity
f-ULW	0.200	/	1.669	/	0.136	/	1.410	0.007	0.009	27.6	Unimodal	Low	Weak
f-UHS	0.146	/	1.710	/	0.063	/	1.515	0.014	0.017	32.6	Unimodal	High	Strong
f-BLW	0.142	0.320	1.456	1.637	0.087	0.069	1.392	0.007	0.010	18.9	Bimodal	Low	Weak
f-BNS	0.107	0.236	1.339	1.710	0.046	0.081	1.459	0.016	0.020	15.3	Bimodal	Normal	Strong
f-BNM	0.135	0.644	1.516	2.742	0.073	0.058	1.477	0.011	0.014	5.7	Bimodal	Normal	Moderate
Coarse modes													
Type	r_{SMC}	r_c	σ_{SMC}	σ_c	C_{SMC}	C_c	n_c	$k_{c,440}$	k_c	Ratio (%)	Size	Ref. ind.	Absorptivity
c-ULW	/	2.751	/	1.941	/	0.089	1.437	0.006	0.009	46.1	Unimodal	Low	Weak
c-UHS	/	3.133	/	1.890	/	0.090	1.522	0.015	0.028	17.1	Unimodal	High	Strong
c-UNW	/	2.250	/	1.675	/	0.482	1.495	0.003	0.003	4.9	Unimodal	Normal	Weak
c-BNM	1.532	3.448	1.986	1.724	0.059	0.105	1.492	0.009	0.019	10.9	Bimodal	Normal	Moderate
c-BHM	2.026	4.788	1.941	1.439	0.121	0.076	1.518	0.008	0.012	20.9	Bimodal	High	Moderate

Note. r is median radius (μm), σ is the standard deviation (μm), C is the peak volume concentration ($\mu m^3/\mu m^2$), n is the real part of refractive index, $k_{f,440}$ ($k_{c,440}$) and k_f (k_c) are imaginary parts of refractive index of fine (coarse) model at 440 nm and from 675 to 1,020 nm. The subscript f is for the standard fine mode, SMF for submicron fine mode, SMC for super-micron coarse mode, and c for the standard coarse mode. Ratio is the appearance probability within either fine or coarse clusters.

3. Results

3.1. Representative Clusters

From the representative clusters, we totally obtain 10 fundamental aerosol models that are uniformly denoted by four characters. The features of volume size distribution have two types (Unimodal and Bimodal) in fine and coarse mode, while the refractivity and absorption capability all have three levels (Low, Normal, and High and Weak, Moderate, and Strong, respectively).

The 10 fundamental aerosol models are presented in Table 1. The detailed aerosol microphysical parameters, the appearance probability, and corresponding characteristics for 10 clusters are presented. In average, as to the fine-particle clusters, the standard fine mode median radius (r_f) are about 0.1–0.2 μm, while those of submicron fine modes (r_{SMF}) can vary from 0.2 to 0.6 μm. As to the coarse-particle clusters, the standard coarse-mode median radius (r_c) varies in a relatively large range 2–5 μm, while those of super-micron coarse modes (r_{SMC}) are about 1.5–2 μm. Meanwhile, as to the refractive index, the fine- and coarse-particle clusters also show different characteristics, for example, real part of fine clusters has relatively low values but large variation changing from 1.39 to 1.51, while those of coarse clusters have high values but small variation range from 1.43 to 1.52. All these characteristics indicate that the fundamental 10 clusters may have clear representative features which will be analyzed in the next section in detail.

3.2. Fine-Mode Aerosol Models

In the fine-particle size range (see Figure 2), we find two single-peak aerosol models (f-ULW and f-UHS) and three double-peak (f-BLW, f-BNS, and f-BNM) aerosol models. It should be noted that the double-peak feature may not be explicit and invisible to the naked eyes because one small peak may be covered by another big peak. Among five fine models, the unimodal types (f-ULW and f-UHS) are dominative with a total ratio of 60.2%, that is, these two models represent nearly two thirds of the cases with fine particle characteristics in China region. The f-ULW model has a high volume concentration ($C_f = 0.136$) but weak light absorption (i.e., smaller k_f). Compared to f-ULW, f-UHS model has a lower volume concentration ($C_f = 0.063$, only about half of that of f-ULW) but stronger light absorbing capability ($k_{f,440} = 0.014$, $k_f = 0.017$). As to the three bimodal fine models, they all have two single-peak models (see Table 2). The f-BLW is the most frequent model (18.9%) among these three models, with a higher volume concentration ($C_f + C_{SMF} = 0.156$) compared to that of dominative unimodal model f-ULW (see Figure 2). Moreover, the lower refractivity ($n_f = 1.392$) agrees well with that of hydrophilic polluted aerosols during serious haze pollution in China (Müller et al., 2006). In contrast, both f-BNS (15.3%) and f-BNM (5.7%) have relatively high refractivity (n_f) and strong absorption ($k_{f,440}$, k_f) similar to characteristics of polluted aerosols in the megacities of

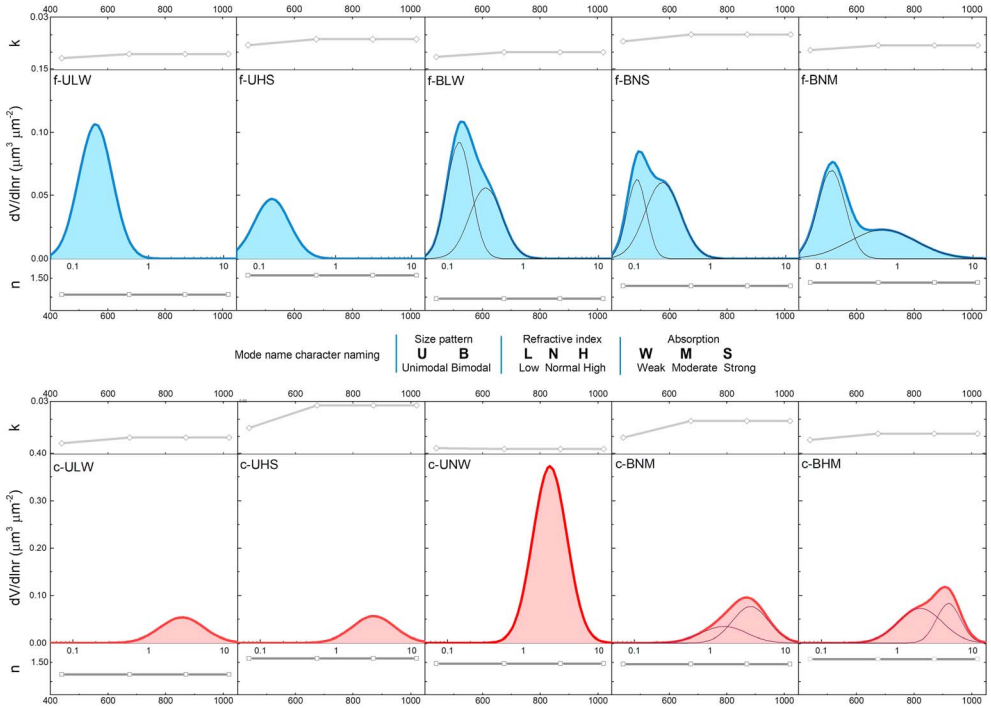

Figure 2. The fine- and coarse-particle aerosol models (volume size distribution and refractive indices) in China region based on clustering of Sun-sky radiometer Observation NETwork measurements. The thin line in the subgraph represents a single peak in the bimodal types. The *x* axis for volume size distribution is radius (unit: μm) and that for refractive indices are wavelength (unit: nm).

north China. Especially, we notice that the average median radius of the submicron mode (r_{SMF}) is 0.4 μm, which is close to the radius of typical SMF mode peak (0.44 μm) observed in Beijing winter heavy haze (Li et al., 2014). All these characteristics indicate that these three bimodal models have strong potentials pointing to haze polluted aerosols of North China cities.

3.3. Coarse-Mode Aerosol Models

For the coarse particles (see Figure 2), we find three single-peak aerosol models (c-ULW, c-UHS, and c-UNW) and two double-peak (c-BNM and c-BHM) aerosol models. The dominative types are also unimodal

Table 2
Types and Characteristics of the Fundamental Models in China Region Derived From Sun-Sky Radiometer Observation Network Measurements

Mode	Cluster	Aerosol type	Description of characteristics
Fine	f-ULW	Urban polluted	Urban-polluted fine-mode aerosols, with high particle concentration
	f-UHS	Continental background	Background fine-mode aerosols, with low particle concentration
	f-BLW	Secondary polluted	Secondary aerosols containing mainly polluted particles, with low refractivity and higher scattering capability
	f-BNS	Combined polluted	Mixture of direct anthropogenic emissions and secondary aerosols, with high light absorption
	f-BNM	Polluted fly ash	Mixture of anthropogenic-polluted particles and fine fly ash particles, with large submicron fine particles
Coarse	c-ULW	Summer fly ash	Background of fly ash coarse particles with low light absorption, for example, natural emission of primary organic aerosols
	c-UHS	Winter fly ash	Background of fly ash coarse particles with high light absorption, for example, fly ash polluted by anthropogenic components
	c-UNW	Primary dust	Coarse particles of natural dust, with very high volume concentration
	c-BNM	Transported dust	Dust particles after long-term transportation and sedimentation
	c-BHM	Background dust	Background dust suspended over continental regions, with large coarse standard mode

clusters (68.2%) similar to the case of fine-particle models. Among three unimodal coarse models, the most frequent model is the c-ULW (46.1%) with the lowest refractivity (n_c = 1.437) and weakest absorption (k_c = 0.009). The c-UHS model has the similar size distribution shape to the c-ULW model but with much stronger light absorbing capability (k_c = 0.028). Considering that most of high absorption coarse-particle components result from soot or flying ash related to coal burning, highly emitted during winter season in China, we think that the high-absorbing c-UHS is related to the coarse particles in north winter. In contrast, it is also reasonable to infer that the c-ULW model represents the summer conditions, especially considering its high water content (i.e., small value n_c) as suggested in Zhang, Li, Zhang, et al. (2017). The c-UNW model shows an extraordinarily high volume concentration (C_c = 0.482) together with weak absorption and moderate refractivity, which implies the natural dust. The weak absorbing characteristics of coarse particles are also observed by Levin et al. (1980) and Otterman et al. (1982), while the unique spectral feature on the increasing absorption at short wavelength (i.e., $k_{c,440} > k_c$) of dust aerosols (Dubovik et al., 2002) is also clearly observed in c-UNW model.

As to the two bimodal coarse models (see Figure 2), we notice that the characteristics of significantly larger median radius and moderately higher absorption agree with that of dusts in North China. However, the major difference between c-BNM and c-BHM models is their median radii. The radius of c-BHM (4.7 μm) is much larger than that of c-BNM (3.4 μm), while their r_{SMC} are close to each other (in an average of 1.78 μm). Considering the standard coarse mode (e.g., r_c) represents the major components of these models, it is reasonable to infer that c-BNM and c-BHM suffer different staying time in the air, which affects the sedimentation of larger particles much greater than smaller particles. Moreover, it can also be noted that the median radius of our unimodal coarse models (about 2.2–3.1 μm) is close to that of global-observed coarse particles (Dubovik et al., 2002), but the standard mode median radius of our bimodal coarse models is significantly larger (~4.1 μm in average) than that of dust aerosols in Dubovik et al. (2002). We think this might be a feature of Asia dust nearing or transported from desert sources (Xu et al., 2014).

4. Discussion

4.1. The Robustness Experiment

For evaluating the robustness of our clustering results, we perform a random-grouping test to examine the validity of classification and the stability of the cluster center values. First, we randomly divide the total 10,773 data (called T data set) into A and B subdata set (equally half of T data set). Next, we perform clustering on A and B exactly following the method mentioned above. Then, we conduct two kinds of evaluations: (a) comparison of cluster numbers of A and B and (b) comparison of the cluster center values of A and B. If the cluster numbers of T, A, and B are stable and the group center values are close (average bias less than 10%), we can confirm that the clustering result is robust (Omar et al., 2005). The result of our robustness experiment meets well the two criteria as mentioned above (Figure S3). The Cluster numbers are exactly the same as that of sub data sets, and the group center values are also stable. The average biases of group center values for the two experiments are less than ±2%. Most of characteristic parameters are stable with bias less than 5%, and only the volume concentration of the standard coarse mode (C_c) is relatively large (~12%). The bias on C_c is explained by the fact that the standard coarse mode is mainly composed of mineral dust which is the biggest source of global atmospheric aerosols with a great variety of concentration in space and time. We conclude that the clustering results obtained from SONET measurements are robust and the obtained aerosol models show good representativeness in China region.

4.2. Representativeness of Typical Aerosol Models

Some scholars preliminarily studied aerosol characterization modeling in China (e.g., Chen et al., 2013; Lee et al., 2010; Wu & Zeng, 2014: Zhang, Xu, & Zheng, 2017). However, all of the aerosol models in these studies are established based on the entire volume size distribution and supposing the fine and coarse modes have the same refractive indices. In our paper, 10 basic modes are obtained for either fine or coarse modes, and thus, they can compose theoretically 36 kinds of bimodal total size distribution patterns, besides the relative amount of fine and coarse modes can adjusted again which products much more combination possibilities. Following the characterization of the clustering results (Figure 2 and Table 1) and the analysis of coappearance, we can summarize 10 fundamental aerosol models to estimate the representativeness of all ground-based measurements and the corresponding typical types. In Table 2, we listed our appointment between

clusters and their type and characteristics. Among the five fine-particle models, the f-ULW model is recognized as urban-polluted type, considering its low refractivity and light absorption together with relatively high peak volume concentration. The f-UHS has the characteristics in contrast with f-ULW and thus is identified as the continental background type. The f-BLW is recognized as the secondary-polluted type, of which refractivity, light absorption, and peak concentration are usually comparable with urban-polluted type f-ULW. The discrepancy of SMF model between f-BLW and f-ULW suggested that the bimodal of fine particles has a close relationship with the seriously polluted haze dominated by secondary aerosols. In contrast, the f-BNS and f-BNM have strong light absorption together with large-size fine particles, which is the typical feature of the complex mixture of polluted aerosols in China. The large-size fine particles of f-BNS are mainly affected by anthropogenic pollutions and have high light absorption, while those of f-BNM are usually affected by fine natural dust or fly ash with an increased SMF peak radius. Among those coarse models, the c-ULW and c-UHS are recognized as representative summer and winter background types, based on their seasonal appearance patterns (not presented in the main text). The c-UNW can be recognized as primary dust because it shows a high frequency at typical desert or acrid regions. In contrast, the c-BNM is recognized as transported dust considering it usually appears far from dust source regions. At the end, the c-BHM is identified as continental background dust considering it appears at all sites and has a very large standard coarse mode peak-volume radius indicating local production by winds. The temporal and spatial variations of the aerosol models are listed in Figures S3 and S5 for more information.

4.3. The Coappearance of the Fine and Coarse Models

In section 4.2, we have analyzed in details the characteristics of the five fine models and five coarse models. They are fundamental elements which can combine into various atmospheric aerosols in nature. In Figure 3a, we present the coappearance matrix between five fine-particle models and five coarse-particle models based on a statistics of clustering results. In general, we find that the coappearance probability of those aerosols which couple the fine and coarse models with close refractivity and absorption properties is very high. For example, f-ULW and f-BLW accompanying c-ULW models appear very frequently. However, the overall coappearance probability of f-BLW model together with all coarse models is obviously lower than that of f-ULW model. Meanwhile, the f-ULW model has a low refractivity ($n_f = 1.392$), suggesting that a close relationship with haze polluted aerosols which usually have high water content ($n = 1.33$) accompanying low concentration of coarse particles. The f-UHS model and c-BHM model usually appear together, which agrees with their common originating regions, mostly Northwest China. When fine model contains bimodal structures (i.e., f-BLW, f-BNS, and f-BNM), the coappearance probability of c-UNW model is obviously low. For the case of f-BNM model, the coappearance chance of bimodal coarse models is nearly zero. This suggests that in the case of larger median radius of SMF mode, the coarse model has rare chance to have bimodal structures (c-BHM and c-BNM). Similarly, the coappearance of SMF (f-B types) and SMC (c-B types) mainly occurs in the condition that f-BNS combines with coarse models. In addition, we also find that the cases of solo model (i.e., only fine or coarse model) are rare (only four cases f-BNM, c-BHM, c-BNM, and c-UNW). For these cases, most of standard fine/coarse mode is always accompanied by the SMF or SMC modes (f-B or c-B types), while the pure solo mode appears only in the case of high concentration c-UNW model.

Following the coappearance analyses in Figure 3a, we plot the combined fine-coarse model with coappearance probability greater than 5% in Figures 3b and 3c (left: fine-particle dominative cases; right: coarse-particle dominative cases). It can be seen from Figure 3b that the three fine-dominative cases are all related to pollutions (Table 2), together with two low concentration coarse models which are all unimodal type. With the pollution condition changes, the standard fine mode median radius increases from 0.1 (combined polluted, f-BNS) to 0.2 (urban polluted, f-ULW), with a decrease of light absorption (k from 0.02 to 0.009), which may reveal a process of haze pollution from formation to spread (Li et al., 2014). In Figure 3c, the both coarse-dominative cases contain the same fine model (continental background, f-UHS) but combined with different coarse particles (background dust, c-BHM and winter fly ash, c-UHS). First, it suggests that for the regions with prevailing natural dust, the anthropogenic pollution is usually low. Second, the background dust model (c-BHM) has a large standard coarse mode peak-volume radius (4.7 μm), much bigger than that (3.1 μm) of the winter fly ash model (c-UHS), which suggests again that China continental background dust has relatively large coarse-particle size. As a summary, the frequently appeared aerosol types in SONET measurements are mainly three polluted and two background types which are prevailing in China region.

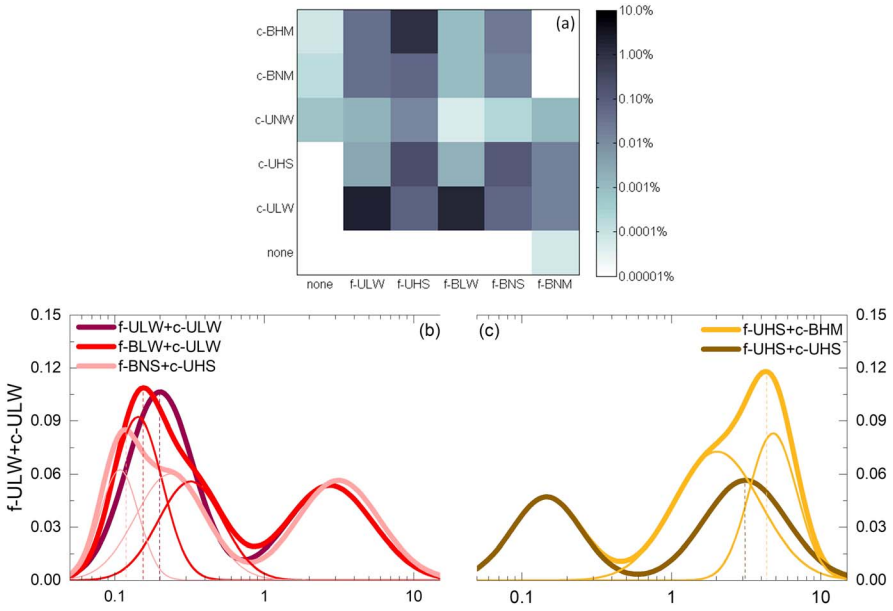

Figure 3. (a) The coappearance probability with respect to the clustered fine and coarse models. The *x* axis is for fine-particle models and *y* axis for coarse-particle models. (b and c) The combined fine-coarse model couples with high coappearance probability (>5%).

5. Conclusions

The multiyear SONET data set at 16 sites are used to conduct the aerosol clustering analyses in China region. We find five fundamental fine models (urban polluted, continental background, secondary polluted, combined polluted, and polluted fly ash) and five fundamental coarse models (summer fly ash, winter fly ash, primary dust, transported dust, and background dust). The random-grouping experiment proves that the clustering results are stable and robust and weakly dependent on data set. The coappearance relationship analyses also show that the aerosol pollution cases can have different optical patterns (e.g., light scattering or light absorption dominated) and usually accompanied by a low concentration background coarse mode. In contrast, the ordinary dust or fly ash aerosols usually have high light absorption and appear in the light-polluted conditions. The obtained 10 fundamental aerosol models can provide a basis for general description of China aerosols. It is very useful for the climate, environment, and ecology studies. Meanwhile, these fine-/coarse-separated models can be directly adapted in the development of satellite remote sensing algorithms to improve the calculations on the atmospheric scattering and absorption by aerosols, ameliorating accuracy and validity of various applications related to radiative transfer through the atmosphere in China.

Acknowledgments
The data used for this study are available at http://www.sonet.ac.cn/en/index.php (SONET website). We would like to thank all the SONET site staffs for the data used for this study. The authors are greatly appreciating the constructive comments and suggestions of anonymous reviewers for improving the expression and structure of this paper. This work was supported by the National Key R&D Program of China (grant 2016YFE0201400) and the National Natural Science Fund of China (41601386, 41671367, and 41671364).

References

Anderberg, M. R. (1973). *Cluster analysis for applications*. Academic Press.

Bellouin, N., Boucher, O., Haywood, J., & Reddy, M. S. (2005). Global estimate of aerosol direct radiative forcing from satellite measurements. *Nature*, *438*(7071), 1138–1141. https://doi.org/10.1038/nature04348

Boucher, O., Randall, D., Artaxo, P., Bretherton, C., Feingold, G., Forster, P., et al. (2013). Clouds and aerosols. In T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, et al. (Eds.), *Climate change 2013: The physical science basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change* (Chap. 8). Cambridge, UK and New York: Cambridge University Press.

Chang, W., Liao, H., Xin, J., Li, Z., Li, D., & Zhang, X. (2015). Uncertainties in anthropogenic aerosol concentrations and direct radiative forcing induced by emission inventories in eastern China. *Atmospheric Research*, *166*, 129–140. https://doi.org/10.1016/j.atmosres.2015.06.021

Che, Y., Xue, Y., Guang, J., She, L., & Guo, J. (2018). Evaluation of the AVHRR DeepBlue aerosol optical depth dataset over mainland China. *ISPRS Journal of Photogrammetry and Remote Sensing*, *146*, 74–90. https://doi.org/10.1016/j.isprsjprs.2018.09.004

Chen, H., Gu, X., Cheng, T., Li, Z., & Yu, T. (2013). The spatial–temporal variations in optical properties of atmosphere aerosols derived from AERONET dataset over China. *Meteorology and Atmospheric Physics*, *122*(1-2), 65–73. https://doi.org/10.1007/s00703-013-0268-2

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., et al. (2002). Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *Journal of the Atmospheric Sciences*, *59*(3), 590–608. https://doi.org/10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2

Dubovik, O., Sinyuk, A., Lapyonok, T., Holben, B. N., Mishchenko, M., Yang, P., et al. (2006). Application of light scattering by spheroids for accounting for particle nonsphericity in remote sensing of desert dust. *Journal of Geophysical Research*, *111*, D11208. https://doi.org/10.1029/2005JD006619

Dubovik, O., Smirnov, A., Holben, B. N., King, M. D., Kaufman, Y. J., Eck, T. F., & Slutsker, I. (2000). Accuracy assessments of aerosol optical properties retrieved from aerosol robotic network (AERONET) Sun and sky radiance measurements. *Journal of Geophysical Research*, *105*(D8), 9791–9806. https://doi.org/10.1029/2000JD900040

Duda, R., Hart, P., & Stork, D. (2001). *Pattern classification* (2nd ed.). New York: John Wiley.

Eck, T. F., Holben, B. N., Reid, J. S., Giles, D. M., Rivas, M. A., Singh, R. P., et al. (2012). Fog- and cloud-induced aerosol modification observed by the Aerosol Robotic Network (AERONET). *Journal of Geophysical Research*, *117*, D07206. https://doi.org/10.1029/2011JD016839

Gobbi, G. P., Kaufman, Y. J., Koren, I., & Eck, T. F. (2007). Classification of aerosol properties derived from AERONET direct sun data. *Atmospheric Chemistry and Physics*, *7*(2), 453–458. https://doi.org/10.5194/acp-7-453-2007

Hansen, J. (2018). Air quality and climate change: Science and politics. AGU Jing Meeting: Atmospheric PM2.5 in China: Change, Impact, Mitigation and Global Perspective. Xi'an.

Hartigan, J. A. (1975). *Clustering algorithms*. John Wiley.

Holben, B. N., Kim, J., Sano, I., Mukai, S., Eck, T. F., Giles, D. M., et al. (2018). An overview of mesoscale aerosol processes, comparisons, and validation studies from DRAGON networks. *Atmospheric Chemistry and Physics*, *18*, 655–671. https://doi.org/10.5194/acp-18-655-2018

Jain, A. K., & Dubes, R. C. (1988). *Algorithms for clustering data*. Prentice Hall.

Lee, J., Kim, J., Song, C. H., Kim, S. B., Chun, Y., Sohn, B. J., & Holben, B. N. (2010). Characteristics of aerosol types from AERONET sunphotometer measurements. *Atmospheric Environment*, *44*(26), 3110–3117. https://doi.org/10.1016/j.atmosenv.2010.05.035

Levin, Z., Joseph, J. H., & Mekler, Y. (1980). Properties of Sharav (Khamsin) dust—Comparison of optical and direct sampling data. *Journal of the Atmospheric Sciences*, *37*(4), 882–891. https://doi.org/10.1175/1520-0469(1980)037<0882:POSDOO>2.0.CO;2

Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., & Kaufman, Y. J. (2007). Second generation operational algorithm: Retrieval of aerosol properties over land from inversion of moderate resolution imaging spectroradiometer spectral reflectance. *Journal of Geophysical Research*, *112*, D13211. https://doi.org/10.1029/2006JD007811

Li, Z., Eck, T., Zhang, Y., Zhang, Y., Li, D., Li, L., et al. (2014). Observations of residual submicron fine aerosol particles related to cloud and fog processing during a major pollution event in Beijing. *Atmospheric Environment*, *86*, 187–192. https://doi.org/10.1016/j.atmosenv.2013.12.044

Li, Z. Q., Xu, H., Li, K. T., Li, D. H., Xie, Y. S., Li, L., et al. (2018). Comprehensive study of optical, physical, chemical, and radiative properties of total columnar atmospheric aerosols over China: An overview of Sun–Sky Radiometer Observation Network (SONET) measurements. *Bulletin of the American Meteorological Society*, *99*(4), 739–755. https://doi.org/10.1175/BAMS-D-17-0133.1

Ma, Y., Li, Z., Li, Z., Xie, Y., Fu, Q., Li, D., et al. (2016). Validation of MODIS aerosol optical depth retrieval over mountains in Central China based on a sun-sky radiometer site of SONET. *Remote Sensing*, *8*(2), 111. https://doi.org/10.3390/rs8020111

Müller, D., Tesche, M., Eichler, H., Engelmann, R., Althausen, D., Ansmann, A., et al. (2006). Strong particle light absorption over the Pearl River Delta (south China) and Beijing (north China) determined from combined Raman lidar and Sun photometer observations. *Geophysical Research Letters*, *33*, L20811. https://doi.org/10.1029/2006GL027196

Omar, A. H., Won, J.-G., Winker, D. M., Yoon, S.-C., Dubovik, O., & McCormick, M. P. (2005). Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements. *Journal of Geophysical Research*, *110*, D10S14. https://doi.org/10.1029/2004JD004874

Otterman, J., Fraser, R. S., & Bahethi, O. P. (1982). Characterization of tropospheric desert aerosols at solar wavelengths by multispectral radiometry from Landsat. *Journal of Geophysical Research*, *87*(C2), 1270–1278. https://doi.org/10.1029/JC087iC02p01270

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: W.H. Freeman.

Wu, L., & Zeng, Q. (2014). Classifying Asian dust aerosols and their columnar optical properties using fuzzy clustering. *Journal of Geophysical Research: Atmospheres*, *119*, 2529–2542. https://doi.org/10.1002/2013JD020751

Xu, H., Li, Z., Li, D., Li, L., Chen, X., Xie, Y., et al. (2014). Ground-based polarimetric remote sensing of dust aerosol properties in Chinese deserts near Hexi corridor. *Advances in Meteorology*, *2014*, 1–10. https://doi.org/10.1155/2014/240452

Xu, X., Wang, J., Zeng, J., Spurr, R., Liu, X., Dubovik, O., et al. (2015). Retrieval of aerosol microphysical properties from AERONET photo-polarimetric measurements: 2. A new research algorithm and case demonstration. *Journal of Geophysical Research: Atmopheres*, *120*, 7079–7098. https://doi.org/10.1002/2015JD023113

Zhang, W., Xu, H., & Zheng, F. (2017). Classifying aerosols based on fuzzy clustering and their optical and microphysical properties study in Beijing, China. *Advances in Meteorology*, *2017*, 1–18. https://doi.org/10.1155/2017/4197652

Zhang, Y., Li, Z., Qie, L., Hou, W., Liu, Z., Zhang, Y., et al. (2017). Retrieval of aerosol optical depth uing the empirical orthogonal function (EOFs) based on PARASOL multi-angle intensity data. *Remote Sensing*, *9*(578), 1–12.

Zhang, Y., Li, Z., Zhang, Y., Li, D., Qie, L., Che, H., & Xu, H. (2017). Estimation of aerosol complex refractive indices for both fine and coarse modes simultaneously based on AERONET remote sensing products. *Atmospheric Measurement Techniques*, *10*(9), 3203–3213. https://doi.org/10.5194/amt-10-3203-2017

Zhang, Y., Li, Z., Zhang, Y.-H., Chen, Y., Cuesta, J., & Ma, Y. (2016). Multi-peak accumulation and coarse modes observed from AERONET retrieved aerosol volume size distribution in Beijing. *Meteorology and Atmospheric Physics*, *128*(4), 537–544. https://doi.org/10.1007/s00703-016-0435-3