

Letter

# Variability of Major Aerosol Types in China Classified Using AERONET Measurements

Lu Zhang  and Jing Li \* 

Department of Atmospheric and Oceanic Sciences, School of Physics, Peking University, Beijing 100871, China; luzhang72@pku.edu.cn

\* Correspondence: jing-li@pku.edu.cn

Received: 25 August 2019; Accepted: 5 October 2019; Published: 9 October 2019



**Abstract:** Aerosol type is a critical piece of information in both aerosol forcing estimation and passive satellite remote sensing. However, the major aerosol types in China and their variability is still less understood. This work uses direct sun measurements and inversion derived parameters from 47 sites within the Aerosol Robotic Network (AERONET) in China, with more than 39,000 records obtained between April 1998 and January 2017, to identify dominant aerosol types using two independent methods, namely, K means and Self Organizing Map (SOM). In total, we define four aerosol types, namely, desert dust, scattering mixed, absorbing mixed and scattering fine, based on their optical and microphysical characteristics. Seasonally, dust aerosols mainly occur in the spring and over North and Northwest China; scattering mixed are more common in the spring and summer, whereas absorbing aerosols mostly occur in the autumn and winter during heating period, and scattering fine aerosols have their highest occurrence frequency in summer over East China. Based on their spatial and temporal distribution, we also generate seasonal aerosol type maps that can be used for passive satellite retrieval. Compared with the global models used in most satellite retrieval algorithms, the unique feature of East Asian aerosols is the curved single scattering albedo spectrum, which could be related to the mixing of black carbon with dust or organic aerosols.

**Keywords:** aerosol type classification; spatial and temporal variability

## 1. Introduction

In recent decades, with its rapid economic development, China has been suffering from severe air pollution and has thus become the focus of global aerosol study. Aerosol properties in China are complex and high variable, due to the various sources including industrial pollution, seasonal biomass burning and dust [1–7]. It is thus important to understand the major types of aerosols in China and their spatial-temporal variability, in order to better estimate their climate and environmental effects.

The information of aerosol type is also critical in aerosol optical depth retrieval using passive satellite sensors, in that this information must be assumed beforehand in the retrieval algorithms [8–12]. However, many satellite aerosol products, such as those from MODIS and VIIRS, still have large uncertainties in China. Li et al. [13] evaluated the VIIRS AOD product over mainland China and found it had an overall high bias of 0.13 compared with ground measurements. Tao et al. [14] compared MODIS aerosol retrievals over China with ground observation and showed that Dark Target (DT) retrievals tended to overestimate the aerosol loading while Deep Blue (DB) retrievals exhibited obvious underestimation in northwestern and southern China. Zhang et al. [15] discussed the quality of the MODIS C6 AOD, whose result was the 3 km (10 km) AOD product performed better in suburban (urban) areas. Zhu et al. [16] found VIIRS AOD overestimated more over the North China Plain region, and AOD bias in Beijing was the largest. One of the reasons for these uncertainties is that assumption of aerosol type is not accurate.

Precise knowledge of aerosol type requires in-situ measurements of aerosol chemical composition. However, such observations, typically conducted during field campaigns, are quite limited in space and time. Ground based remote sensing techniques such as those using multi-channel sun and sky photometers, can provide information on aerosol optical and microphysical properties, which is also closely related to their type and composition. In fact, many successful attempts have been made to infer aerosol type from surface remote sensing measurements. For example, Omar et al. [17] identified six aerosol types globally from about 143,000 AERONET observations using the K means clustering technique. Logan et al. [18] examined the seasonal variation of aerosol size and composition in Asia using data from four selected AERONET sites. Chen et al. [19] analyzed AERONET data from 12 Chinese stations and found that the aerosol types in China exhibit significant spatial variability and seasonal periodicity. Xie et al. [20] presented the optical and microphysical properties of aerosol for four typical sites over China based on a newly developed remote sensing network for aerosol observation, the Sun-sky radiometer Observation NETwork (SONET). Schmeisser et al. [21] implemented different methods, including multivariate cluster analysis and back trajectory analysis, to estimate aerosol type of 24 stations within the NOAA/ESRL Federated Aerosol Monitoring Network. Che et al. [1] used sun photometer measurements from seven sites in the Yangtze River Delta (YRD) from 2011 to 2015 to characterize the aerosol microphysical and optical properties and classified the aerosols based on size and absorption. Nonetheless, these studies were either global or only used a limited number of sites in China. There is still lack of a comprehensive analysis of major aerosol types and their spatial-temporal variability for this region.

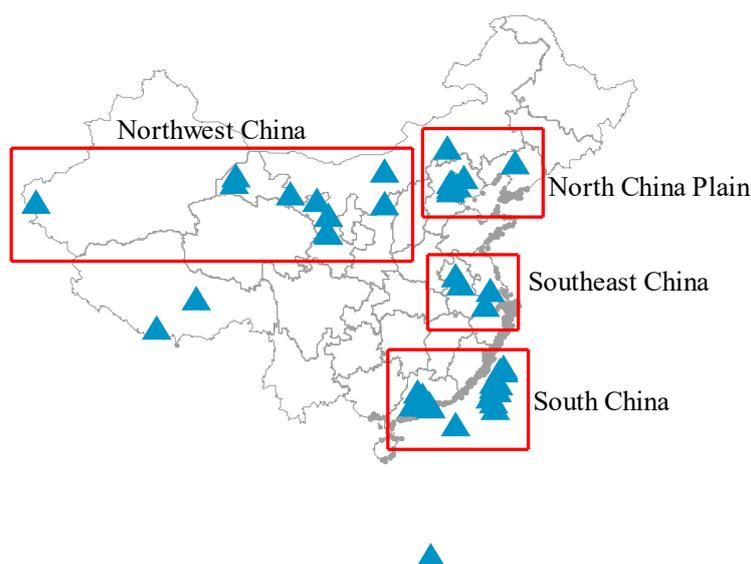
Although aerosol distribution is highly inhomogeneous, their variability exhibits certain seasonal patterns. For example, in East Asia, there are typically more dust and black carbon aerosols in spring, but in summer and autumn, the aerosol population is dominated by weakly absorbing pollution and small organic carbon aerosols, whereas in winter, it is dominated by fine particles with strong absorption [18]. Ground based sun photometers are useful tools for the continuous aerosol observation in the daytime. This also allows for the examination of diurnal variability of aerosol type, which is particularly important in the aerosol retrieval and data validation from geostationary platforms.

In this paper, we aim to generalize a representative aerosol type dataset for China based on surface observation and clustering techniques. We further analyze the seasonal variability of the major types, as well as briefly examine their diurnal variability. Compared to previous works, our study has used a more extensive set of ground observation sites that potentially contains more comprehensive information of aerosol type and variability. We further generate aerosol type maps that can be directly used in passive satellite retrieval. We hope this study can provide a useful reference for satellite remote sensing, climate change and air pollution assessments.

## 2. Data and Methods

### 2.1. AERONET Data

AERONET is a ground-based observation network established by NASA to monitor the global atmospheric aerosol characteristics [22]. There are more than 800 observation stations around the world, providing an observational basis for scientists all over the world to explore aerosol properties. This study uses Level 2.0 direct sun and inversion products download from the AERONET website (<https://aeronet.gsfc.nasa.gov/>). The data contained 150 aerosol parameters. It undergoes front and rear field of view calibration, cloud processing and manual inspection to ensure quality, which is the closest to the actual situation. Currently there are 47 sites located in China, covering major urban and desert regions (Figure 1). We selected 25 aerosol microphysical parameters from all these 47 stations for clustering, including spectral complex refraction index, effective radius of particles, single scattering albedo, asymmetric factors and so on, which are listed in Table 1.



**Figure 1.** Location of the AERONET sites in China.

**Table 1.** Aerosol parameters used in the clustering.

Parameters	Variables
Angstrom Exponent <sup>3</sup>	Angstrom Exponent-Total, Absorption Angstrom Exponent, Alpha
Single Scattering Albedo	SSA440 <sup>1</sup> -T <sup>2</sup> , SSA676 <sup>1</sup> -T <sup>2</sup> , SSA869 <sup>1</sup> -T <sup>2</sup> , SSA1020 <sup>1</sup> -T <sup>2</sup>
Imaginary Part of the Complex Refractive Index	REFI (440 <sup>1</sup> ), REFI (676 <sup>1</sup> ), REFI (869 <sup>1</sup> ), REFI (1020 <sup>1</sup> )
Real Part of the Complex Refractive Index	REFR (440 <sup>1</sup> ), REFR (676 <sup>1</sup> ), REFR (869 <sup>1</sup> ), REFR (1020 <sup>1</sup> )
Asymmetry Parameter	ASYM440 <sup>1</sup> -T <sup>2</sup> , ASYM676 <sup>1</sup> -T <sup>2</sup> , ASYM869 <sup>1</sup> -T <sup>2</sup> , ASYM1020 <sup>1</sup> -T <sup>2</sup>
Volume Concentration	VolCon-F <sup>2</sup> , VolCon-C <sup>2</sup>
Effective Radius	EffRad-F <sup>2</sup> , EffRad-C <sup>2</sup>
Standard Deviation of Effective Radius	StdDev-F <sup>2</sup> , StdDev-C <sup>2</sup>

<sup>1</sup> 440, 675, 676, 869, 870, 1020 refers to the wavelength in nm. <sup>2</sup> T refers to total. F refers to fine. C refers to coarse.

<sup>3</sup> Angstrom Exponent-Total and Absorption Angstrom Exponent are calculated from the AOD obtained by sky scattering light inversion. Alpha is the Angstrom index calculated from the AOD obtained by direct observation of solar radiation.

In total, the 47 sites selected in this paper have 39,249 aerosol observations (Level 2.0 AERONET data) from April 1998 to January 2017. Optical depths and alpha are retrieved by direct sun observation with higher accuracy. Therefore, these two parameters are first used as the basis of screening. For this study, the range of acceptable optical depths is  $AOD_{mean} - \sigma$  to  $AOD_{mean} + 3\sigma$ , and the acceptable alpha (Angstrom Exponent) is greater than  $alpha_{mean} - 3\sigma$  ( $\sigma$  refers to one standard deviation) [17]. Data beyond these thresholds are considered abnormal and are removed. This brings the sample size down to 36,477. Data with missing values cannot participate in clustering, so we remove it and 18,191 observations remain. There are four sites with less data removed in this process. To reduce spatial bias of the data, two sites in North China (Xianghe and Beijing\_CMA) where site density is high were eliminated because they are too close to the Beijing site while the data quality is not as high. The remaining observations of spring, summer, autumn and winter are 4432, 1387, 3207 and 4244, respectively. The summer data volume is much smaller due to increased cloudy and rainy conditions. We therefore further reduce the temporal bias by evenly reducing the spring and winter data by 2/3 (reduce 2 out of every 3), and the autumn data by 1/2 (reduce 1 out of 2). This results in the

number of observations in four seasons being 1478, 1387, 1603 and 1414, respectively, so that the data entering in the clustering algorithm are evenly distributed in time and space. After quality control and pre-processing, 5882 observations from 41 sites remain for further analysis. Figure 2 shows the data processing, and Table 2 indicates the number of stations and observations after each processing step over the four regions shown in Figure 1. All parameters are further standardized by their standard deviation before entering the clustering algorithm.

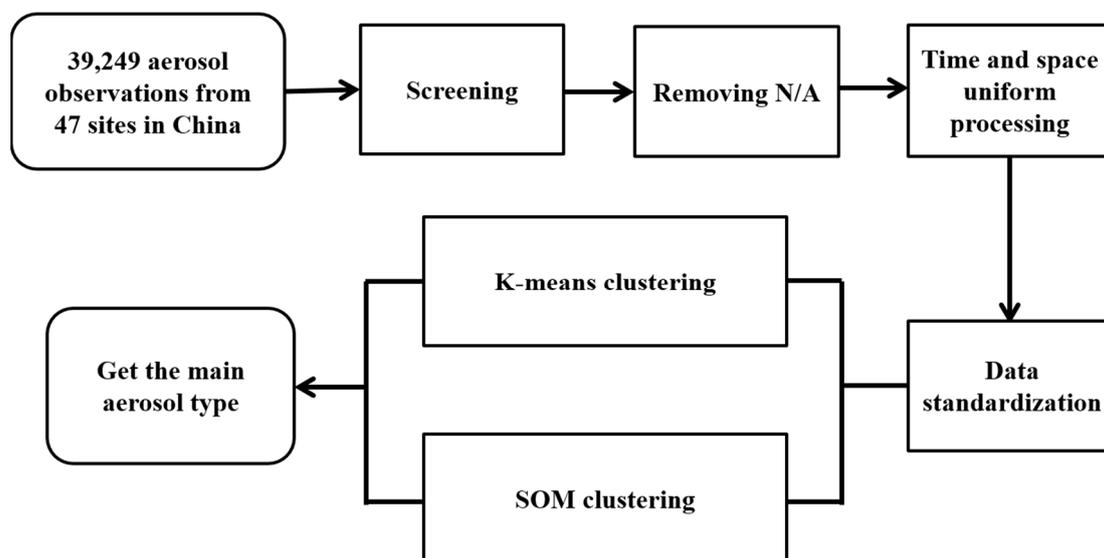


Figure 2. Data processing flow chart.

Table 2. Aerosol parameters used in the clustering.

Area	Raw Data		After Removing N/A		After Time and Space Uniform Processing	
	Number of Sites	Number of Data	Number of Sites	Number of Data	Number of Sites	Number of Data
Northwest China	12	6882	10	1937	10	896
North China Plain	8	21,237	8	9235	6	2023
Southeast China	4	2942	4	2413	4	1025
South China	23	8188	21	4606	21	1938

## 2.2. Clustering Techniques

### 2.2.1. K Means Clustering

In this paper, K means clustering method is first used to classify aerosol. K means clustering is a dynamic clustering method for clustering objects into a given number of K classes. The steps are as follows.

1. Roughly divide all objects into K initial classes and determine their central points, where K is a given natural number.
2. Adjust the existing classes and assign each object to the class nearest to its center (such as mean vector).
3. Recalculate the center point of the class that has objects called out or called in, and repeat step 2 for readjustment and so on until a reasonable classification is obtained.

This clustering method is easy to understand and calculate, and it is a classical algorithm in clustering analysis.

We further use the Davies–Bouldin index [23] to define the “best” number of clusters. The DB index is defined as the ratio of within-cluster scatter ( $S_i$ ) to between-cluster separation ( $d_{ij}$ ). The specific definitions are as follows:

$$S_i = \frac{1}{|a_i|} \sum_{x \in a_i} |x - k_i|, \quad (1)$$

$$d_{ij} = \|k_i - k_j\|, \quad (2)$$

$$DB = \frac{1}{M} \sum_{i=1}^M \max_{j, j \neq i} \left( \frac{S_i + S_j}{d_{ij}} \right), \quad (3)$$

In the equations,  $a_i$  represents the collection of class  $i$  elements,  $k_i$  represents the central value of class  $i$ ,  $M$  is the number of classes.  $S_i$  can reflect the compactness within each category and  $d_{ij}$  can reflect the dispersity between different categories. DB is thus an evaluation of in-cluster distance and between-cluster distance, and the smaller the DB index is, the better the clustering result is.

### 2.2.2. SOM Clustering

In order to increase the robustness of the clustering results, we also use the Self Organizing Map (SOM) method to evaluate whether the clustered patterns are stable.

SOM is an unsupervised learning algorithm that can obtain good visualization of the data. The structure of SOM network simulates the function of self-organizing feature mapping of human neural network. The network consists of the input layer and the output layer (competition layer), in which the number of neurons in the input layer is determined by the number of vectors in the input network, and the output layer is arranged into a two-dimensional node matrix by the neurons in a certain manner. The neurons in the input layer and the neurons in the output layer are connected together through the weights, which are constantly adjusted. The final stable network output then generates a natural feature map of the input pattern, thus achieving the goal of automatic clustering [24].

SOM automatically assigns similar data to the same node. This algorithm assumes that there are some topological structure or sequence in the input objects, and it can realize the dimensionality reduction mapping from the input space ( $n$ -dimension) to the output plane (2-dimension). While  $K$  means determines which type each data belongs to according to the closest distance between the data point to the centroid of four types as described in Section 2.2.1. The clustering effect of  $K$  means is that the points within the class are close enough and the points between different classes are far enough. In fact, SOM with a small number of nodes works similar to  $K$  means, and the criteria of clustering are both based on the distance between a data point and a given point (the centroid for  $K$  means and the node for SOM). Although  $K$  means is a highly popular algorithm, its results can be quite sensitive to initialization and may give very different results each time, whereas SOM does not have this drawback. Therefore, to increase the credibility of the clustering results, we simultaneously used these two different methods.

## 3. Results

### 3.1. Clustering Analysis

The aerosol data are classified into different clusters (2~8 clusters) by the  $K$  means method, and their DB indices are calculated respectively. The values are shown in Figure 3. It is seen that the DB index is the smallest when the cluster number is four. We therefore consider this as the reasonable number. The sample size and centroids of the four cluster results are listed in Table 3. For more robust results, we redo the clustering using an independent technique—SOM. We have actually tried different numbers of nodes. We found that using 3 nodes produces a slightly lower DB index than using 4 nodes, but the sample numbers of the three types are quite unbalanced (one type has the majority of the data points). When using 5 nodes or more, the DB index increases. We therefore consider 4 nodes as the optimum number. This result also agrees the best with  $K$  means results. The SOM results are

given in the parenthesis of Table 3 in parallel to the K means results. We notice that the results of the two techniques are almost identical, which increases the credibility of the clustering results.

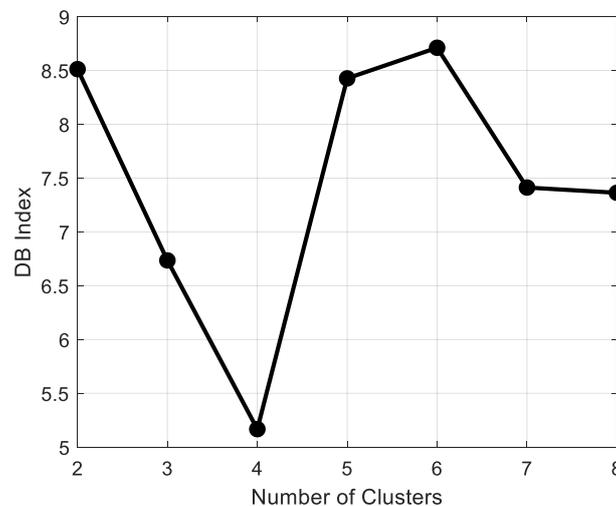


Figure 3. The DB indices of different classifications numbers.

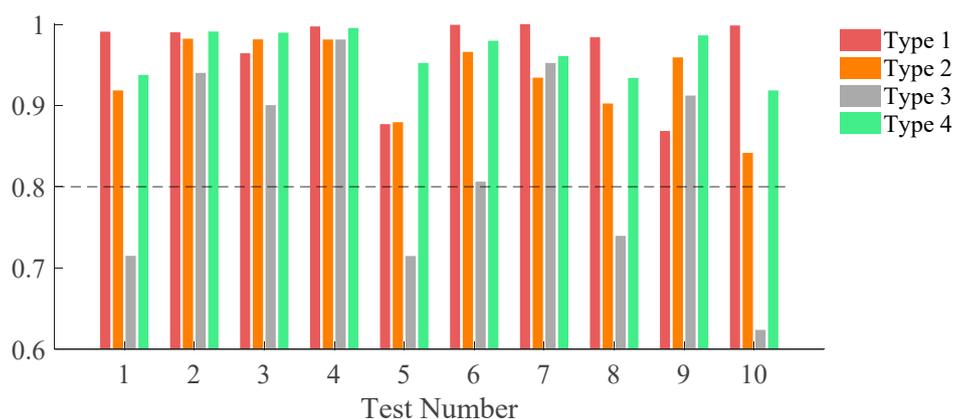
Table 3. K means Self Organizing Map (SOM) clustering results. Outside the brackets is the result of K means, and inside the brackets is the result of SOM.

	Type 1 (Desert Dust)	Type 2 (Scattering Mixed Type)	Type 3 (Absorbing Mixed Type)	Type 4 (Scattering Fine Type)
Number of record	802 (792)	2071 (2090)	976 (969)	2033 (2031)
Angstrom Exponent – Total	0.679 (0.674)	1.375 (1.373)	1.265 (1.264)	1.269 (1.269)
SSA (440 nm)	0.890 (0.890)	0.912 (0.912)	0.851 (0.850)	0.949 (0.949)
SSA (676 nm)	0.930 (0.930)	0.919 (0.918)	0.856 (0.856)	0.950 (0.950)
SSA (869 nm)	0.936 (0.936)	0.911 (0.911)	0.839 (0.838)	0.944 (0.944)
SSA (1020 nm)	0.938 (0.938)	0.905 (0.905)	0.827 (0.827)	0.939 (0.939)
Absorption Angstrom Exponent	1.577 (1.580)	1.362 (1.361)	1.184 (1.183)	1.177 (1.178)
REFR (440 nm)	1.500 (1.500)	1.443 (1.443)	1.489 (1.489)	1.420 (1.420)
REFR (676 nm)	1.527 (1.527)	1.461 (1.462)	1.508 (1.508)	1.423 (1.423)
REFR (869 nm)	1.530 (1.530)	1.472 (1.472)	1.520 (1.520)	1.424 (1.424)
REFR (1020 nm)	1.527 (1.527)	1.471 (1.471)	1.523 (1.523)	1.419 (1.419)
REFI (440 nm)	0.008 (0.008)	0.011 (0.011)	0.022 (0.022)	0.007 (0.006)
REFI (676 nm)	0.005 (0.005)	0.008 (0.008)	0.017 (0.018)	0.006 (0.006)
REFI (869 nm)	0.004 (0.004)	0.008 (0.008)	0.018 (0.018)	0.006 (0.006)
ASYM (440 nm)-T	0.723 (0.723)	0.713 (0.712)	0.694 (0.694)	0.749 (0.749)
ASYM (676 nm)-T	0.685 (0.686)	0.643 (0.643)	0.635 (0.635)	0.700 (0.700)
ASYM (869 nm)-T	0.683 (0.683)	0.612 (0.612)	0.617 (0.618)	0.666 (0.666)
ASYM (1020 nm)-T	0.688 (0.688)	0.604 (0.605)	0.616 (0.616)	0.649 (0.649)
VolCon-F ( $\mu\text{m}^3/\mu\text{m}^2$ )	0.059 (0.059)	0.105 (0.104)	0.084 (0.084)	0.118 (0.118)
EffRad-F ( $\mu\text{m}$ )	0.137 (0.137)	0.163 (0.163)	0.150 (0.150)	0.214 (0.214)
StdDev-F	0.527 (0.527)	0.491 (0.491)	0.504 (0.505)	0.543 (0.543)
VolCon-C ( $\mu\text{m}^3/\mu\text{m}^2$ )	0.277 (0.278)	0.094 (0.094)	0.111 (0.111)	0.067 (0.067)
EffRad-C ( $\mu\text{m}$ )	1.982 (1.981)	2.199 (2.198)	2.249 (2.252)	2.481 (2.481)
StdDev-C	0.621 (0.620)	0.627 (0.628)	0.639 (0.638)	0.583 (0.583)
Alpha	0.668 (0.663)	1.378 (1.376)	1.262 (1.261)	1.279 (1.280)
Fine Mode Fraction	0.177 (0.175)	0.528 (0.526)	0.431 (0.432)	0.637 (0.637)

When analyzing the spatial distribution and seasonal variability of aerosol types in the next section, we hope to know the aerosol type information from all sites. So we use all data in this step, but the data not used for clustering analysis must be manually classified according to its distance from

the clustered centers. For the observations used in the clustering step, reclassification will not affect the results. This is because our reclassification criterion is the same as that used in K means. K means determines to which type each data belongs to according to the closest distance between this data point and the centroid of four types. When we decide which type this data should belong to, we also assign it the type whose centroid is the closest to this data point. Therefore, there will not be cases when a data being classified into one type by K means is later assigned to a different type by reclassification. Then, all 39,249 data are used and classified into the defined four types according to the Euclidean distance between the data and the centroid of the four types, i.e., the selecting the centroid with the smallest distance and classifying the data into that type. In addition, we define a distance threshold as the longest distance from the centroid of each type, i.e., if the shortest distance between a data point and a centroid is still greater than the distance threshold of this type, this data is not classified and is thus discarded. In this way, we classified 39,249 data records by discarding only 47 records.

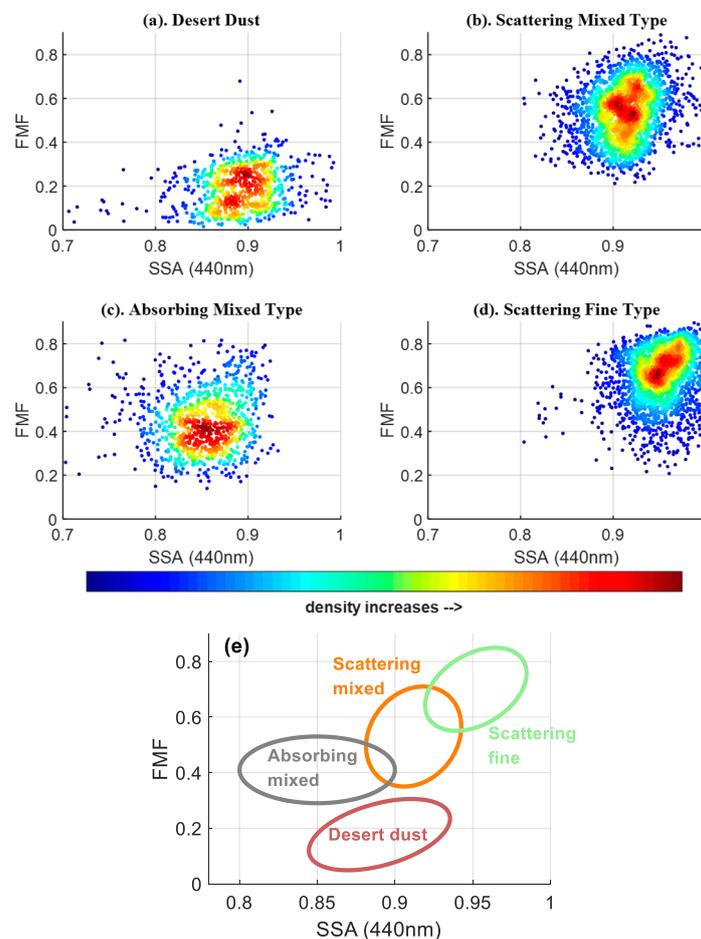
Moreover, to rule out the effect of sampling on the clustering results, we further perform a Monte Carlo type validation by randomly selecting half of the data from the original data as a subset to conduct data processing and clustering analysis and repeat the process ten times. Each time we compare the clustering results of the subset with those of all data. If one observation data is classified as the same aerosol type in both cases, it is marked as “correct”, thus the accuracy rate of the four types can be obtained by counting the number of correct classifications. Figure 4 shows the accuracy rate for these ten experiments using the K means method. It is seen that most accuracy rates are above 80%, with more than half of them exceeding 90%. Therefore, we consider that the four types identified are stable.



**Figure 4.** Accuracy rate of the result compared to clustering the entire data set.

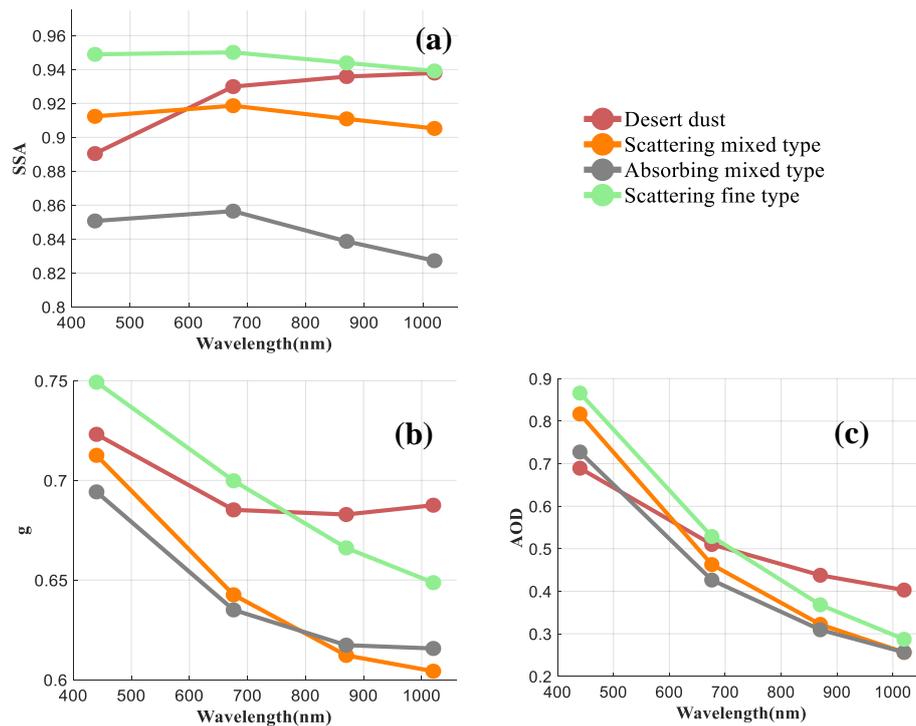
The four types identified by the analysis are tentatively named as dust, scattering mixed type, absorbing mixed type and scattering fine type respectively, primarily based on their scattering/absorption parameters (single scattering albedo, refractive indices, absorption Angstrom Exponent) and size parameters (effective radius, fine mode fraction, extinction Angstrom Exponent). Because of the difficulty in high dimensional visualization, Figure 5a–d shows the scatter plot between two representative parameters—single scattering albedo (SSA) and fine mode fraction (FMF). It is observed that type 3 has 440 nm SSA about 0.85 which indicates the strongest absorption. Type 2 and type 4 have SSA above 0.9 in four wavelengths, which implies relatively strong scattering. On the other hand, Type 1 has the lowest FMF, while Type 4 has the highest FMF. As indicated by Chen et al. [19],  $FMF < 0.4$  indicates that coarse particles dominate the aerosol model,  $FMF > 0.6$  indicates that fine particles dominate the aerosol model, and between them, the aerosol model is mixed. In Figure 5e, we further define a range for each type based on their scatter plots. Although there are slight overlaps between scattering mixed with absorbing mixed and scattering fine types, overall the ranges of four types are well separated. Dust has the lowest FMF and moderate SSA; scattering fine has the highest FMF and highest SSA; absorbing mixed has the lowest SSA and moderate FMF, while the properties of

scattering mixed lie between those of scattering fine and absorbing mixed. These results support the validity of our clustering results.



**Figure 5.** (a)–(d) is the density graph of SSA and FMF. Red indicates high density value, and blue indicates small density value. (e) gives a range for different aerosol types.

In addition to single wavelength properties, the spectral dependencies of many parameters are also associated with aerosol composition. For example, as indicated by several previous studies [25–28], dust SSA increases with wavelength due to its UV absorption, whereas anthropogenic aerosols show a reversed behavior. We therefore continue to examine the spectra for three parameters of the four types in Figure 6: (a) single scattering albedo (SSA), (b) asymmetry parameter ( $g$ ), (c) aerosol optical depth (AOD). In Figure 6a, it is clearly seen that SSA of desert dust increases with wavelength while that of other types decreases with wavelength, which is due to the absorption of dust in the UV wavelengths and is consistent with the previously documented results. Moreover, the SSA spectra of absorbing and scattering mixed aerosols both exhibit non-monotonic behavior, i.e., SSA first increases with wavelength from 440 to 675 nm, then decreases. This is an indication of mixing of black carbon with dust or organic carbon and is a typical feature of East Asian aerosols [27]. According to the more quantitative method proposed by Li et al. [27], absorbing mixed (with greater SSA spectral curvature) is likely black carbon mixed more with dust, whereas scattering mixed likely contain more organic carbon. With respect to asymmetry parameter, desert dust has the  $g$  factor above 0.7 and varies little with wavelength, whereas that of other types decreases rapidly with wavelength (Figure 6b). Figure 6c also shows that dust AOD has a much more flat spectrum than others. These clear differences between dust and the other aerosol types confirm that the former indeed primarily consists of larger particles.



**Figure 6.** Optical parameters of the aerosol types: (a) single scattering albedo, (b) asymmetry parameter, (c) aerosol optical depth.

### 3.2. Spatial Distribution and Seasonal Variability of the Aerosol Types

With the clustering information in hand, it becomes possible to examine the spatial and temporal variability of major aerosol types in China. Figure 7 shows the distribution of frequency of occurrence of each type for each season. Note that although the clustering only uses a selected portion of AERONET data described in Section 2, Figure 7 is produced using all data. Each data point is classified into one of the four types according to the threshold mentioned in Section 3.1 and the Euclidian distance between this data point and the center of each type. Some sites do not appear on the map during certain seasons because they have available data only for some seasons but not for others. In addition, five sites representative of major pollution source regions in China (SACOL, Beijing, Taihu, Taiwan Cheng Kung University and Hong Kong) are further selected to examine their seasonal variability of aerosol types in detail (Figure 8).

Figure 7 shows that desert dust mainly occurs in northwest China in the spring (MAM), close to the major dust sources of the Taklimakan and Gobi deserts. The North China Plain also has significant amounts of dust in the spring due to the transport by the prevailing west winds in this season. These results are consistent with previous studies [29–31]. Figure 8 also reveals that dust has high frequency of occurrence in the spring at SACOL (Northwest China) and Beijing (North China Plain). Moreover, Taihu is also dominated by dust aerosol in spring. Scattering mixed type occurs mostly in the spring and summer (JJA) over East China and South China, and it also appears relatively frequently in the South China all year round (Figures 7 and 8). A large number of such aerosols in south China are due to the emission from biomass burning, industrial activities, automobile exhaust and so on. Another reason may be that organic carbon aerosols released by extensive biomass burning activities in Southeast Asia are transported to South China by the monsoon [32]. The total amount of absorbing mixed type is relatively small and is mainly found over North China Plain and East China and occurs most frequently in the fall (SON) and winter (DJF) (Figures 7 and 8). The seasonal feature is consistent with the heating time when many carbonaceous aerosols are released from coal combustion and vehicle exhausts [33,34]. For the scattering fine type (Figures 7 and 8), the occurrence mainly concentrates over North China Plain and East China during the summer and fall. These two seasons are featured with

high humidity and precipitation, and the suspended water droplets can facilitate the conversion of gases to particles, leading to more fine mode aerosols with strong scattering (such as the conversion of sulfur dioxide to sulfate particles, [33,35]). Moreover, hygroscopic growth of aerosols also increases their scattering efficiency. Both factors favor the higher scattering properties retrieved by AERONET. Nonetheless, due to the limitation of remote sensing data, we cannot isolate the relative humidity effect.

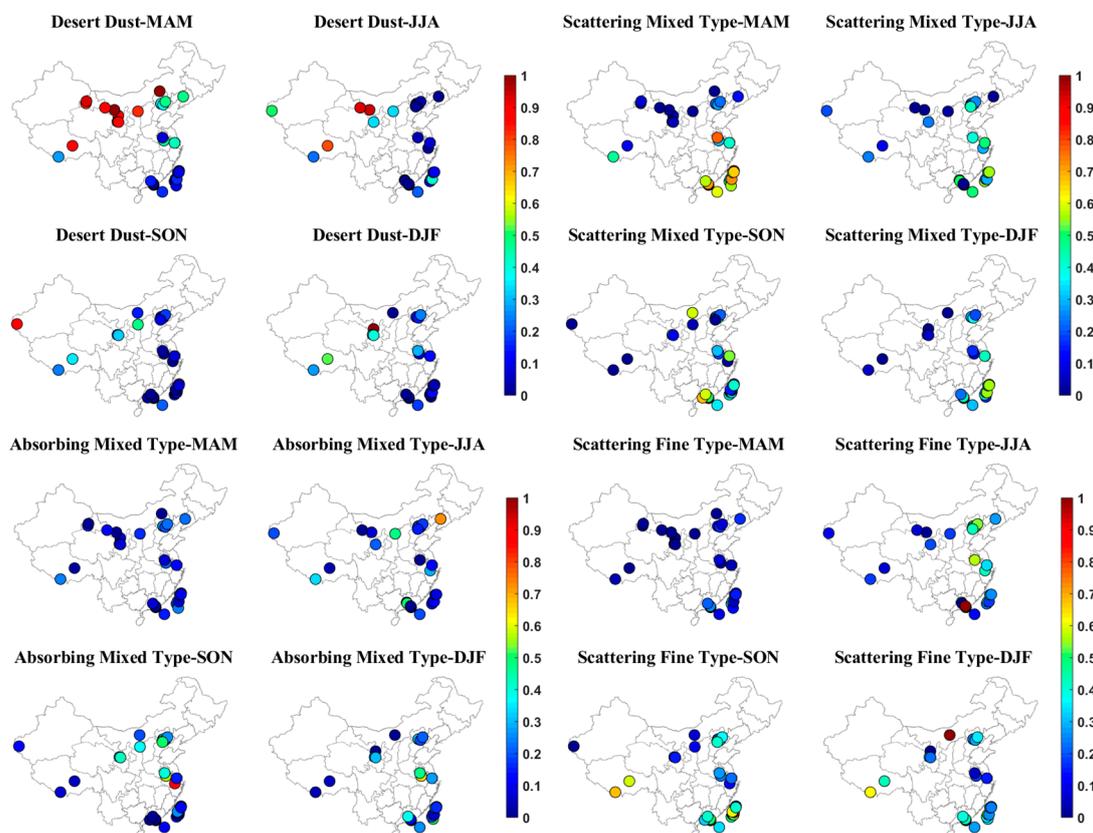
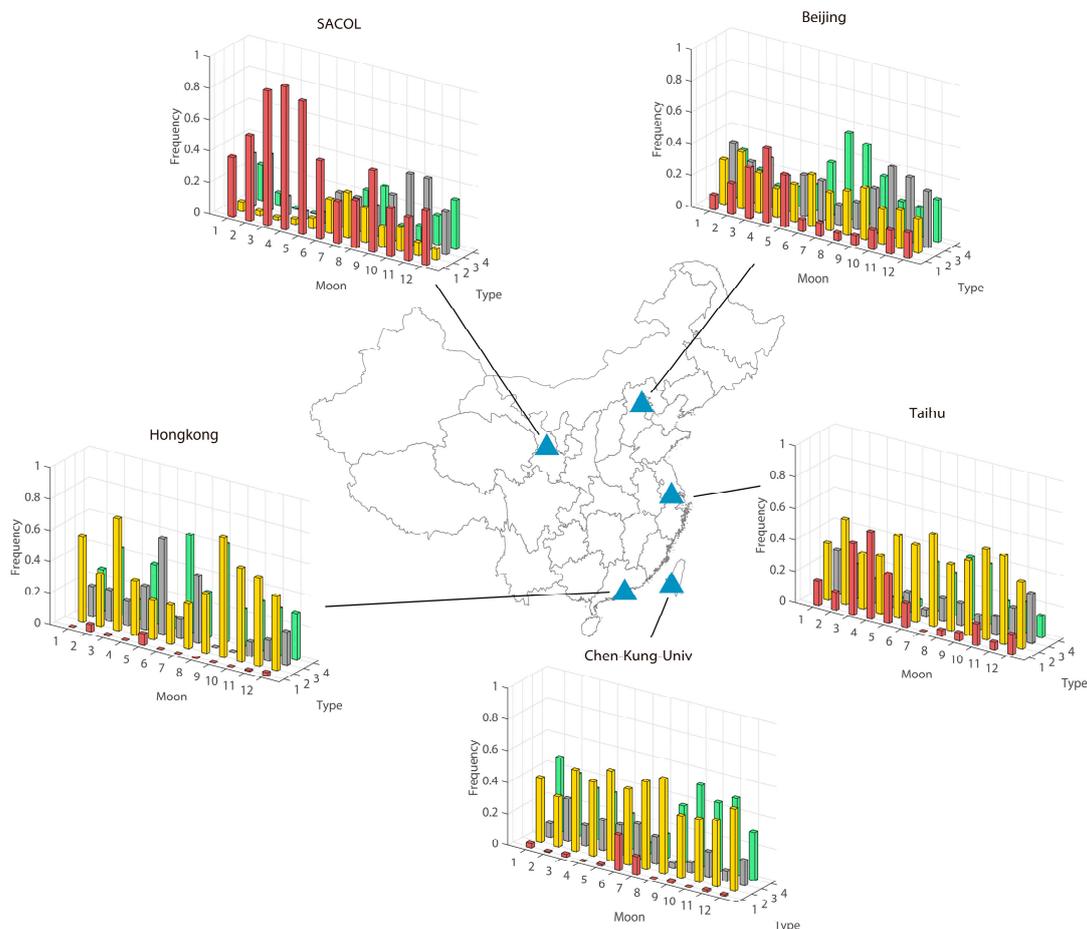


Figure 7. The occurrence frequency of four aerosol types.

Note that here we focus on identifying the optical and microphysical properties of different aerosol types, which are critical in aerosol forcing estimation and satellite aerosol retrievals. The remote sensing measurements that we used cannot give the detailed chemical composition of aerosols. By comparing Figure 7 with existing studies that report aerosol chemical composition in China ([5,7,17,36,37]), we can make some inference of the composition of four aerosol types. The dust type should mainly contain sand dust, urban fugitive dust and coal ash ([7]), which is the main component of aerosol in northwest China. Zhang et al. [7] also found black carbon with strong absorption has its high level of mass concentrations in North China especially during autumn and winter, consistent with spatial and seasonal distribution of our absorbing mixed type. Many previous studies [5,17,36,37] showed the small particles with strong scattering property such as organic carbon, sulfate, nitrate and ammonium all have high concentration in China especially in urban due to various emission sources. They also account for a large proportion of aerosol composition in summer which maybe driven by the aqueous processing and gas-phase photochemical production. Therefore, it is likely that the two scattering types mainly consist of organic carbon, ammonium sulfate and nitrate. Moreover, the scattering mixed type should contain more organic carbon due to its SSA spectra. Besides, Figure 8 also shows the typical aerosol types in representative regions of China. For example, SACOL (Northwest China) is dominated by dust aerosol with big particles almost throughout the year. Beijing (North China) and Taihu (Southeast China) are mixed regions of all aerosol types where the four types all have obvious frequency of occurrence, so there is a significant mixture of coarse particles and fine particles in these

regions. The frequency of dust aerosols is close to 0 in Taiwan Cheng Kung University and Hong Kong (South China), therefore this region is mainly dominated by fine particles.



**Figure 8.** Multi-year mean monthly frequency of aerosols in five sites. Type 1 is dust desert. Type 2 is scattering mixed type. Type 3 is absorbing mixed type. Type 4 is scattering fine type.

Briefly, dust aerosols are mainly found over Northwest China and the North China Plain in the spring, scattering mixed aerosols over Southeast China in the spring and summer, absorbing mixed type over East China in the fall and winter, and scattering fine type over East China in the summer and fall.

An advantage of sun photometers is their continuous measurements during the daytime, allowing the examination of diurnal variability of aerosol type. We conduct a statistical analysis on the diurnal variation of aerosol types by calculating the averaged frequency of diurnal aerosol type change. Specifically, for each station, the number of days with at least four daily observations is counted, and then the number of days on which more than one type is found is calculated; the ratio between the latter and the former thus represents the frequency of diurnal change of aerosol type. The results are shown in Figure 9. We can see that except for Northwest China, diurnal type change is frequently observed (frequency > 0.5) for most East and South China sites. This is reasonable as in the Northwest aerosol composition is relatively simple (mainly dust as discussed in the previous section), whereas in East and South China, complicated mixtures are often found due to both local emission and remote transport. We also analyzed the frequency of occurrence of the four aerosol types in the morning and afternoon but did not find outstanding patterns. Therefore, diurnal variability of aerosol types is a common phenomenon and cannot be ignored. This has important implications for both climate modeling and satellite remote sensing. Current climate models generally use averaged optical parameters while different aerosol type will result in different radiative forcing. In addition, aerosol retrieval

from geostationary platforms has become increasingly popular. Such practice also needs to assume the aerosol model and without accounting for the diurnal variability of aerosol type will inevitably introduce uncertainties in the results. We will present detailed analysis of the effect of aerosol diurnal change on radiative forcing and aerosol retrieval in a following study.

### The Frequency of Diurnal Variability in Aerosol Type

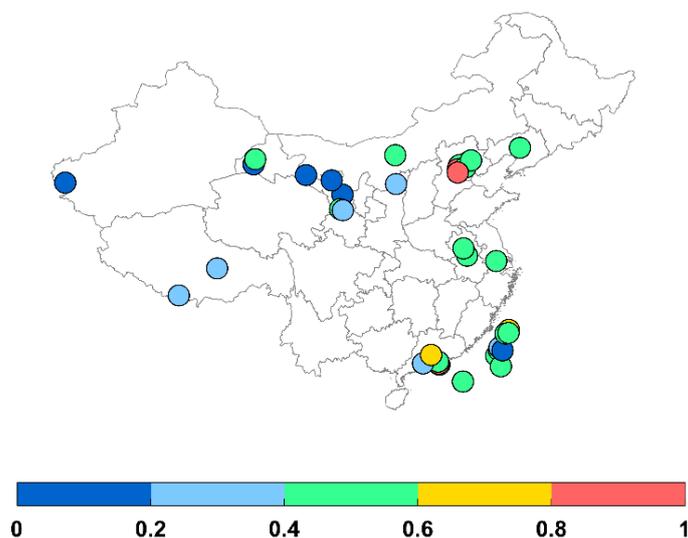


Figure 9. Frequency distribution of diurnal aerosol type change.

#### 3.3. Aerosol Type Map for Satellite Remote Sensing

Another important practical use of aerosol classification is to facilitate the assumption of aerosol model in passive satellite retrievals, such as those from Moderate Resolution Spectroradiometer (MODIS) and the Visible Infrared Imaging Radiometer Suite (VIIRS). For example, the aerosol model used in the MODIS retrieval algorithm is based on the classification study using AERONET measurements by Omar et al. [17]. However, many of these global models are not quite representative of East Asian aerosols. In fact, only one station in China, Beijing, is used in the Omar et al. [17] work. On the one hand, East Asia is a global pollution hotspot with complicated emission sources and aerosol composition and deserves special attention in aerosol observation. On the other hand, current satellite aerosol products still have large uncertainties over this region, and aerosol model assumption has been suggested as a major source of uncertainties (e.g., [13]). As a result, it is thus necessary to refine the representation of aerosol optical and microphysical properties in East Asia in the retrieval algorithms in order to improve the retrieval accuracy over this region.

The SSA and  $g$  spectra used by MODIS and VIIRS are shown in Figure 10 whose aerosol model information comes from reference [38,39]. By comparing Figures 6 and 10, obviously desert dust relates to dust in Figure 10, and we can infer that absorbing mixed type should contain more black carbon. Scattering fine type maybe dominated by sulfate and nitrate aerosols. However, because in reality the aerosols are almost always mixed and the sun photometers retrieved the integrated properties, it is not possible to accurately define their chemical composition. We find that the major difference is the lack of obviously curved SSA spectral in the MODIS and VIIRS fine aerosol models, which is a characteristic frequently occurring over East Asia. We further attempt to generate a seasonal map of aerosol type distribution in China based on the spatial and seasonal variability of the four aerosol types revealed by Figure 7, as shown in Figure 11. Although the resolution is coarse and there are many places not covered due to the limited number of AERONET sites, this map provides more detailed information about aerosol type variability in China than that used for many satellite aerosol retrieval algorithms such as MODIS and VIIRS. With the ongoing ground observation efforts in China, it is promising that this map could become more accurate in the future.

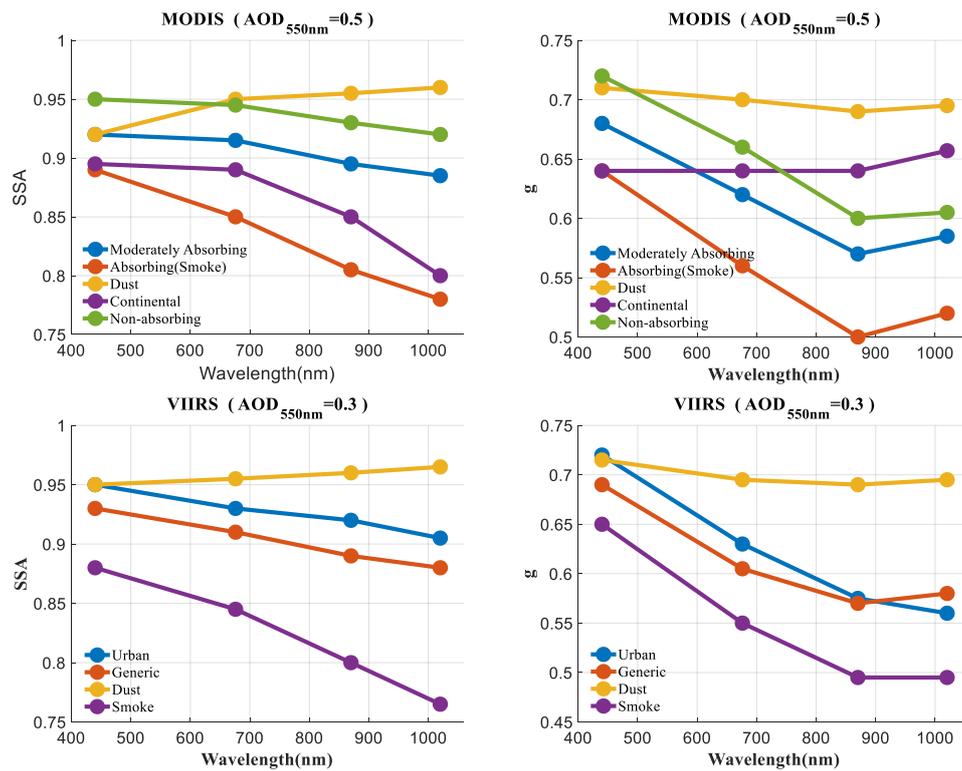


Figure 10. Aerosol models used by MODIS [38] and VIIRS [39].

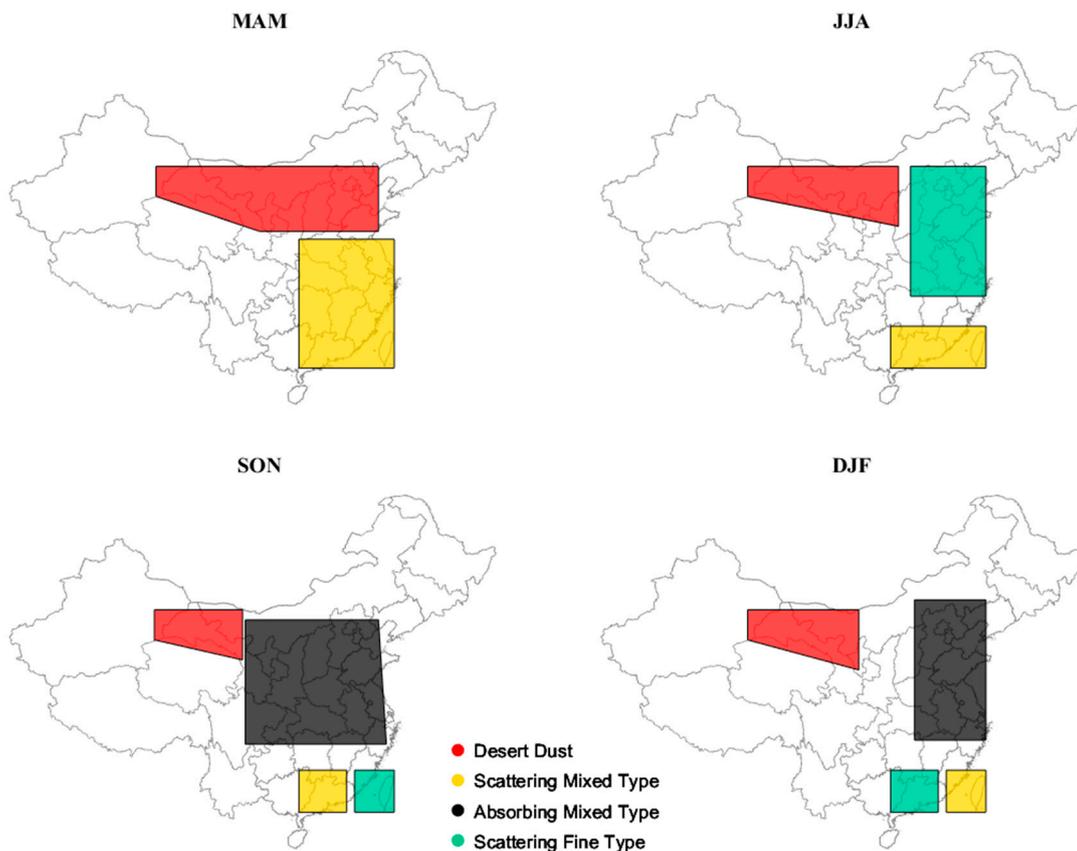


Figure 11. Seasonal distribution of aerosol types used recommended for satellite aerosol retrieval.

#### 4. Conclusions and Discussion

In this study, we identify major aerosol types in China using cluster analysis of AERONET measurements. Two independent clustering techniques, K means and SOM, are used, and they produce highly consistent results. The main conclusions of this study are summarized below:

(1) We identify four types, namely desert dust, scattering mixed type, absorbing mixed type and scattering fine type, respectively, according to their optical and microphysical properties.

(2) Briefly, dust aerosols have the lowest FMF, slowest variation of AOD and asymmetry factor with wavelength and increased SSA with wavelength. Absorbing mixed type has the lowest overall SSA, while scattering fine has the highest SSA values. The latter also has the highest FMF. Absorbing mixed and scattering mixed types both have non-monotonic SSA spectra, i.e., their SSA first increases with wavelength from 440 to 675 nm and then decreases, which is an indication of the mixing of black carbon with dust or organic carbon aerosols.

(3) In terms of space–time variability, dust aerosols are typically found in the spring season over Northwest China and the North China Plain. Scattering mixed type mostly occurs in the spring and summer over Southeast China. The absorbing mixed aerosols appear frequently in autumn and winter over most of the East China region, corresponding to heating period. The scattering fine aerosol has the highest frequency of occurrence in the summer and fall over East China.

(4) Based on these spatial and temporal variability, we further produce a seasonal aerosol type map to facilitate aerosol retrieval from passive satellite sensors such as MODIS and VIIRS.

The aerosol type classification presented in this paper can provide an useful reference for the aerosol parameterization in climate models, as well as aerosol model assumption in satellite retrieval algorithms. To our knowledge, our study is the most comprehensive aerosol type classification work using ground based remote sensing data. The results have added to our understanding of aerosol composition and their optical and microphysical properties in China. The newly established aerosol model can also help improve satellite retrieval accuracy. In fact, in another paper [40], we have used this updated aerosol model to retrieve aerosol optical depth over the Beijing area and obtained very promising results.

Nonetheless, the use of remote sensing data can only define the types based on their optical properties, while a detailed study of aerosol composition requires in-situ chemical measurements. In the future, we plan to incorporate such data to validate and improve our classification results. We will also try to incorporate observations from domestic sun photometer networks. Moreover, the change of aerosol type on seasonal and diurnal scale has significant impact on both aerosol forcing estimation and satellite retrieval. In the future we will investigate these effects in more detail.

**Author Contributions:** J.L. conceptualized and supervised the study. L.Z. conducted the analysis and prepared the figures and tables. J.L. and L.Z. wrote the manuscript together.

**Funding:** This study is funded by the National Natural Science Foundation of China (No. 41575018), the National Key Research and Development Program of China No. 2017YFC0212803, and the Young Elite Scientists Sponsorship (YESS) Program by CAST (No. 2016QNRC001).

**Acknowledgments:** We gratefully acknowledge the AERONET site principal investigators for the data used for this study. The AERONET data is downloaded from <https://aeronet.gsfc.nasa.gov/>.

**Conflicts of Interest:** The authors declare no conflict of interest.

#### References

1. Che, H.; Qi, B.; Zhao, H.; Xia, X.; Eck, T.F.; Goloub, P.; Dubovik, O.; Estelles, V.; Cuevasagullo, E.; Blarel, L.; et al. Aerosol optical properties and direct radiative forcing based on measurements from the China Aerosol Remote Sensing Network (CARSNET) in eastern China. *Atmos. Chem. Phys.* **2018**, *18*, 405–425. [[CrossRef](#)]
2. Miao, Y.; Liu, S. Linkages between aerosol pollution and planetary boundary layer structure in China. *Sci. Total Environ.* **2019**, *650*, 288–296. [[CrossRef](#)]

3. Mu, Q.; Liao, H. Simulation of the interannual variations of aerosols in China: Role of variations in meteorological parameters. *Atmos. Chem. Phys.* **2014**, *14*, 9597–9612. [[CrossRef](#)]
4. Reddy, K.R.O.; Zhang, X.; Bi, L. Seasonal aerosol variations over a coastal city, Zhoushan, China from CALIPSO observations. *Atmos. Res.* **2019**, *218*, 117–128. [[CrossRef](#)]
5. Sun, Y.; Wang, Z.F.; Fu, P.; Yang, T.; Jiang, Q.; Dong, H.; Li, J.; Jia, J.J. Aerosol composition, sources and processes during wintertime in Beijing, China. *Atmos. Chem. Phys.* **2013**, *13*, 4577–4592. [[CrossRef](#)]
6. Tian, P.; Cao, X.; Zhang, L.; Sun, N.; Sun, L.; Logan, T.; Shi, J.; Wang, Y.; Ji, Y.; Lin, Y.; et al. Aerosol vertical distribution and optical properties over China from long-term satellite and ground-based remote sensing. *Atmos. Chem. Phys.* **2017**, *17*, 2509–2523. [[CrossRef](#)]
7. Zhang, X.Y.; Wang, Y.; Niu, T.; Zhang, X.; Gong, S.L.; Zhang, Y.; Sun, J. Atmospheric aerosol compositions in China: Spatial/temporal variability, chemical signature, regional haze distribution and comparisons with global aerosols. *Atmos. Chem. Phys.* **2012**, *12*, 779–799. [[CrossRef](#)]
8. Levy, R.C.; Remer, L.A.; Kleidman, R.; Mattoo, S.; Ichoku, C.; Kahn, R.A.; Eck, T.F. Global evaluation of the Collection 5 MODIS dark-target aerosol products over land. *Atmos. Chem. Phys.* **2010**, *10*, 10399–10420. [[CrossRef](#)]
9. Levy, R.C.; Munchak, L.; Mattoo, S.; Patadia, F.; Remer, L.A.; Holz, R.E. Towards a long-term global aerosol optical depth record: Applying a consistent aerosol retrieval algorithm to modis and viirs-observed reflectance. *Atmos. Meas. Tech.* **2015**, *8*, 4083–4110. [[CrossRef](#)]
10. Mielonen, T.; Levy, R.C.; Aaltonen, V.; Komppula, M.; Leeuw, G.D.; Huttunen, J.; Lihavainen, H.; Kolmonen, P.; Lehtinen, K.E.J.; Arola, A. Evaluating the assumptions of surface reflectance and aerosol type selection within the MODIS aerosol retrieval over land: The problem of dust type selection. *Atmos. Meas. Tech.* **2011**, *4*, 201–214. [[CrossRef](#)]
11. Oo, M.M.; Jerg, M.; Hernandez, E.; Picon, A.; Gross, B.; Moshary, F.; Ahmed, S. Improved MODIS Aerosol Retrieval Using Modified VIS/SWIR Surface Albedo Ratio Over Urban Scenes. *IEEE Trans. Geosci. Remote Sens.* **2010**, *48*, 983–1000. [[CrossRef](#)]
12. Zhang, H.; Kondragunta, S.; Laszlo, I.; Liu, H.; Remer, L.A.; Huang, J.; Superczynski, S.; Ciren, P. An enhanced VIIRS aerosol optical thickness (AOT) retrieval algorithm over land using a global surface reflectance ratio database. *J. Geophys. Res.* **2016**, *121*, 10738. [[CrossRef](#)]
13. Li, C.; Li, J.; Xu, H.; Li, Z.; Xia, X.; Che, H. Evaluating VIIRS EPS Aerosol Optical Depth in China: An intercomparison against ground-based measurements and MODIS. *J. Quant. Spectrosc. Radiat.* **2019**, *224*, 368–377. [[CrossRef](#)]
14. Tao, M.; Chen, L.; Wang, Z.; Tao, J.; Che, H.; Wang, X.; Wang, Y. Comparison and evaluation of the MODIS Collection 6 aerosol data in China. *J. Geophys. Res.* **2015**, *120*, 6992–7005. [[CrossRef](#)]
15. Zhang, J.; Xin, J.; Zhang, W.; Wang, S.; Wang, L.; Xie, W.; Xiao, G.; Pan, H.; Kong, L. Validation of MODIS C6 AOD Products Retrieved by the Dark Target Method in the Beijing-Tianjin-Hebei Urban Agglomeration, China. *Adv. Atmos. Sci.* **2017**, *34*, 993–1002. [[CrossRef](#)]
16. Zhu, J.; Xia, X.; Wang, J.; Che, H.; Chen, H.; Zhang, J.; Xu, X.; Levy, R.C.; Oo, M.; Holz, R.; et al. Evaluation of Aerosol Optical Depth and Aerosol Models from VIIRS Retrieval Algorithms over North China Plain. *Remote Sens.* **2017**, *9*, 432. [[CrossRef](#)]
17. Omar, A.H.; Won, J.; Winker, D.M.; Yoon, S.; Dubovik, O.; McCormick, M.P. Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements. *J. Geophys. Res.* **2005**, *110*, D10S14. [[CrossRef](#)]
18. Logan, T.; Xi, B.; Dong, X.; Li, Z.; Cribb, M. Classification and investigation of Asian aerosol absorptive properties. *Atmos. Chem. Phys.* **2013**, *13*, 2253–2265. [[CrossRef](#)]
19. Chen, H.; Gu, X.; Cheng, T.; Yu, T.; Li, Z. Characteristics of aerosol types over China. *J. Remote Sens.* **2013**, *17*, 1559–1571. [[CrossRef](#)]
20. Xie, Y.; Li, Z.; Li, D.; Xu, H.; Li, K. Aerosol Optical and Microphysical Properties of Four Typical Sites of SONET in China Based on Remote Sensing Measurements. *Remote Sens.* **2015**, *7*, 9928–9953. [[CrossRef](#)]
21. Schmeisser, L.; Andrews, E.; Ogren, J.A.; Sheridan, P.J.; Jefferson, A.; Sharma, S.; Kim, J.; Sherman, J.; Sorribas, M.; Kalapov, I.; et al. Classifying aerosol type using in situ surface spectral aerosol optical properties. *Atmos. Chem. Phys.* **2017**, *17*, 12097–12120. [[CrossRef](#)]

22. Holben, B.N.; Eck, T.F.; Slutsker, I.; Tanre, D.; Buis, J.P.; Setzer, A.W.; Vermote, E.F.; Reagan, J.A.; Kaufman, Y.J.; Nakajima, T.; et al. AERONET—A Federated Instrument Network and Data Archive for Aerosol Characterization. *Remote Sens. Environ.* **1998**, *66*, 1–16. [[CrossRef](#)]
23. Davies, D.L.; Bouldin, D.W. A cluster separation measure. *IEEE Trans. Pattern Anal.* **1979**, *1*, 224–227. [[CrossRef](#)]
24. Vesanto, J. SOM-Based Data Visualization Methods. *Intell. Data Anal.* **1999**, *3*, 111–126. [[CrossRef](#)]
25. Bergstrom, R.W.; Pilewskie, P.; Russell, P.B.; Redemann, J.; Bond, T.C.; Quinn, P.K.; Sierau, B. Spectral absorption properties of atmospheric aerosols. *Atmos. Chem. Phys.* **2007**, *7*, 5937–5943. [[CrossRef](#)]
26. Dubovik, O.; Holben, B.; Eck, T.F.; Smirnov, A.; Kaufman, Y.J.; King, M.D.; Tanre, D.; Slutsker, I. Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations. *J. Atmos. Sci.* **2002**, *59*, 590–608. [[CrossRef](#)]
27. Li, J.; Carlson, B.E.; Laci, A.A. Using single-scattering albedo spectral curvature to characterize East Asian aerosol mixtures. *J. Geophys. Res. Atmos.* **2015**, *120*, 2037–2052. [[CrossRef](#)]
28. Russell, P.B.; Bergstrom, R.W.; Shinzuka, Y.; Clarke, A.D.; Decarlo, P.F.; Jimenez, J.L.; Livingston, J.M.; Redemann, J.; Holben, B.; Dubovik, O.; et al. Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition. *Atmos. Chem. Phys.* **2010**, *9*, 21785–21817. [[CrossRef](#)]
29. Cao, C.; Zheng, S.; Singh, R.P. Characteristics of aerosol optical properties and meteorological parameters during three major dust events (2005–2010) over Beijing, China. *Atmos. Res.* **2014**, *150*, 129–142. [[CrossRef](#)]
30. Wehner, B.; Wiedensohler, A.; Tuch, T.M.; Wu, Z.J.; Hu, M.; Slanina, J.; Kiang, C.S. Variability of the aerosol number size distribution in Beijing, China: New particle formation, dust storms, and high continental background. *Geophys. Res. Lett.* **2004**, *31*, L22108. [[CrossRef](#)]
31. Yu, X.; Lu, R.; Kumar, K.R.; Ma, J.; Zhang, Q.; Jiang, Y.; Kang, N.; Yang, S.; Wang, J.; Li, M. Dust aerosol properties and radiative forcing observed in spring during 2001–2014 over urban Beijing, China. *Environ. Sci. Pollut. Res.* **2016**, *23*, 15432–15442. [[CrossRef](#)] [[PubMed](#)]
32. Sahu, L.K.; Sheel, V. Spatio-temporal variation of biomass burning sources over South and Southeast Asia. *J. Atmos. Chem.* **2014**, *71*, 1–19. [[CrossRef](#)]
33. Che, H.; Xia, X.; Zhu, J.; Li, Z.; Dubovik, O.; Holben, B.N.; Goloub, P.; Chen, H.; Estelles, V.; Cuevasagullo, E.; et al. Column aerosol optical properties and aerosol radiative forcing during a serious haze-fog month over North China Plain in 2013 based on ground-based sunphotometer measurements. *Atmos. Chem. Phys.* **2014**, *14*, 2125–2138. [[CrossRef](#)]
34. Zhang, X.Y.; Wang, Y.Q.; Zhang, X.; Guo, W.; Gong, S.L. Carbonaceous aerosol composition over various regions of China during 2006. *J. Geophys. Res.* **2008**, *113*, D14111. [[CrossRef](#)]
35. Hennigan, C.J.; Bergin, M.H.; Dibb, J.E.; Weber, R.J. Enhanced Secondary Organic Aerosol Formation due to Water Uptake by Fine Particles. *Geophys. Res. Lett.* **2008**, *35*, L18801. [[CrossRef](#)]
36. Tao, J.; Zhang, L.; Cao, J.; Zhang, R. A review of current knowledge concerning PM<sub>2.5</sub> chemical composition, aerosol optical properties, and their relationships across China. *Atmos. Chem. Phys.* **2017**, *17*, 9485–9518. [[CrossRef](#)]
37. Sun, Y.; Wang, Z.; Du, W.; Zhang, Q.; Wang, Q.Q.; Fu, P.; Pan, X.; Li, J.; Jayne, J.T.; Worsnop, D.R. Long-term real-time measurements of aerosol particle composition in Beijing, China: Seasonal variations, meteorological effects, and source analysis. *Atmos. Chem. Phys.* **2015**, *15*, 10149–10165. [[CrossRef](#)]
38. Levy, R.C.; Remer, L.A.; Dubovik, O. Global aerosol optical properties and application to Moderate Resolution Imaging Spectroradiometer aerosol retrieval over land. *J. Geophys. Res.* **2007**, *112*. [[CrossRef](#)]
39. Laszlo, I.; Liu, H. *EPS Aerosol Optical Depth (AOD) Algorithm Theoretical Basis Document*; NOAA NESDIS Center for Satellite Applications and Research: College Park, MD, USA, 2016.
40. Li, C.; Li, J. Impact of aerosol vertical distribution of aerosol optical depth retrieval from passive satellite sensors. *Remote Sens. Environ.* (paper submitted).

