https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Machine Learning-Enhanced Aerosol Types Classification: Global Multi-Parameter Aerosol Modeling and an Automated Classifier Driven by Active–Passive Remote Sensing Parameters

Zhuo Yang[1, 2], Qingshan Xu[1], and Chen Cheng[1]

[1]Hefei Institutes of Physical Science, Chinese Academy of Sciences, Hefei, 230031, China.
[2]University of Science and Technology of China, Hefei, 230026, China.

Correspondence: Qingshan Xu (qshxu@aiofm.ac.cn) and Chen Cheng (chengchen@aiofm.ac.cn)

Abstract. Accurate classification of aerosol types is crucial for regional aerosol pollution source tracking and optical remote sensing studies. However, conventional classification methods are mostly relying on the single-device observation that provides limited optical parameters, and the use of fixed empirical thresholds that limits the classification accuracy. This study develops a machine learning-enhanced aerosol types classification method that integrates active and passive optical remote sensing data. First, the global multi-parameter aerosol model is built from high-quality aerosol product datasets of 19 representative AERONET sites worldwide, 7 distinct aerosol type classes were established using the K-means clustering algorithm. Then, based on the established aerosol database with multiple characteristic parameters, we propose a random forest-based aerosol types classifier, which uses only active-passive remote sensing directly retrieved parameters as inputs. This classifier is evaluated to achieve an overall accuracy of approximately 80% and successfully classifies the dominant aerosol types at 27 typical AERONET sites. This study offers a practical and reliable method for combining active and passive remote sensing data to automated aerosol types classification.

1 Introduction

Aerosols refer to solid and liquid particles in the atmosphere (Falah et al., 2022). Their composition is complex and keep changing since they can originate from volcanic eruptions, sea sprays, mineral dust, and many other natural and human sources (Carslaw et al., 2013; Filioglou et al., 2020; Rizza et al., 2019). Some aerosols can directly affect human health (Yang and Xu, 2022); Aerosol optical properties also strongly affect atmospheric radiative transfer directly and indirectly (Mallet et al., 2017; Li et al., 2023). For quantifying these influences, correctly classifying regional aerosol types via optically measured aerosol properties is crucial. An effective and accurate aerosol types classification method is especially critical for tracking aerosol pollution and determining the applicable aerosol model for remote sensing data processing (Roger et al., 2022; Shekhar et al., 2021).

The aerosol properties often used for aerosol types classification can be categorized into 2 kinds: (1) directly retrieved aerosol optical parameters. Such as Aerosol Optical Depth (AOD), Extinction Angstrom Exponent (EAE), Lidar Ratio (LR, also known as aerosol extinction-to-backscatter ratio), and Depolarization Ratio (DR), they can be obtained from detection

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

data by simple calculation or retrieval. (2) complex retrieved aerosol properties. Such as Single Scattering Albedo (SSA), Real Part of the Refractive Index (RRI), Imaginary Part of the Refractive Index (IRI), Fine Mode Fraction (FMF), they must be obtained by considering the aerosol scattering model. The traditional aerosol types classification methods are based on the empirical determined thresholds of several aerosol properties, they were usually limited in dimension (Barnaba and Gobbi, 2004; Mukhopadhyay et al., 2025; Youroi et al., 2026). Currently, some studies use machine learning algorithms to strengthen classification processing, such as K-means clustering (Fan et al., 2021), Mahalanobis distance clustering (Hamill et al., 2016), Naive Bayes classifier (Yang et al., 2026; Zhou et al., 2022), random forest classifier (Del Aguila et al., 2025; Choi et al., 2021b), etc. They take a great number of aerosol optical properties(mainly complex retrieved aerosol properties) as inputs to improve classification accuracy. Therefore, they are constrained in practice and depend on the comprehensive aerosol optical properties datasets.

In this study, we establish 7 distinct aerosol type classes with multiple characteristic parameters from the high-quality aerosol product datasets of global typical AERONET sites using the K-means clustering algorithm. The aerosol properties selected for clustering are not only contain various key aerosol characteristics but also have been frequently used in other classification studies. Then, taking established aerosol database as training sample, we propose an aerosol types classifier based on random forest algorithm, which takes only frequently used directly retrieved optical parameters(AOD, EAE, LR and DR) as inputs. Compared with other studies, this proposed classifier is more universally applicable and will not be affected by the uncertainty introduced by complex model calculation. Besides, this classifier is constructed from a comprehensive multi-dimensional database established by clustering, rather than from an empirical database, which ensures objectivity. Evaluated with validation data, the overall accuracy of the constructed classifier is nearly 80%. By testing, this classifier can correctly classify the dominant aerosol types in 27 typical AERONET sites.

2 Global multi-parameter aerosol model

2.1 AERONET data

AERONET is a globally ground-based aerosol monitoring network developed by NASA. The sun-sky radiometers in AERONET sites measure direct solar radiance at 8 wavelengths (340, 380, 440, 500, 675, 870, 940, and 1020 nm) and sky radiance at 4 wavelengths (440, 670, 865, and 1020 nm), to retrieve and provide various aerosol properties with high accuracy and time resolution (Sinyuk et al., 2020). Many studies have used the aerosol properties given by AERONET in clustering for aerosol types classification (Wang et al., 2024; Zhang and Li, 2019). The AERONET version 3 data have 3 various data quality levels: Level 1.0 (unscreened data), Level 1.5 (cloud-screened and quality-controlled data), and Level 2.0 (cloud-screened and quality-ensured data) (Nie et al., 2025). In this study, the quality-ensured AERONET version 3 Level 2.0 aerosol inversion product data is chosen to build the multi-parameter aerosol model.

Since this study aims to develop the multi-parameter model of global typical aerosol types, AERONET sites were selected according to the following criteria:

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Table 1. The basic information of AERONET sites used in clustering.

Sites	Location	The dominant aerosol type
Anmyon, Beijing, Chen-Kung_Univ, Gosan_SNU	East Asia	Mixed
Abracos_Hill, Alta_Floresta, Cuiaba	South America	Biomass Burning
Mongu, Senanga, Zambezi	South Africa	Biomass Burning
Banizoumbou, Capo_Verde	North Africa	Dust
Mussafa, Solar_Village	Arabia	Dust
GSFC, MD_Science_Center	North America	Urban/Industrial
Ispra, Leipzig, Lille	Europe	Urban/Industrial

(1) The selected sites must be recognized as typical sites that have a single dominant aerosol type. Previous studies have defined many typical AERONET sites dominated by a single aerosol type. However, some studies may defined different dominant aerosol types for the same site, such as the site "Shirahama" (Giles et al., 2012; Hamill et al., 2016). Therefore, this study only selected the sites that most studies consistently attribute to the same aerosol type.

In addition, there are inherent limitations in the application of AERONET products. The quality control criteria of AERONET Level 2.0 inversion products require an aerosol optical depth (AOD) larger than 0.4, while the actual measurement results of marine aerosol can hardly meet this requirement. Consequently, many AERONET-based aerosol types classification studies did not consider marine aerosol as an independent aerosol type in their classification schemes (Giles et al., 2012; Lee et al., 2010; Wei et al., 2024). Marine aerosol site data were also excluded in this study.

(2) The AERONET sites dominated by the same aerosol type but with different geographical locations were also selected. Some previous studies have defined many geographically separated sites dominated by the same aerosol types (Russell et al., 2010). By taking these sites into account, we may find the subtle aerosol property differences or even subtypes caused by regional differences.

(3) All observation data were collected up to the end of 2025. Since aerosol properties may change within observation time, in this study, all available observation data at each station through the end of 2025 were adopted, so that these data can characterize the long-term real regional aerosol properties.

(4) Balancing sample sizes was considered when selecting sites. The amount of valid data(Level 2) varies significantly among different AERONET sites. Since these site data were collected for clustering, it is desirable to maintain a basic balance in the amount of data for different aerosol types.

Finally, 19 typical AERONET sites across various continents were selected for clustering. The geographic locations of these sites are displayed in Fig. 1. The data from these 19 sites can capture various aerosol types' characteristics(except for marine aerosol) and regional characteristics. Some sites are geographically close to each other. Their basic information (locations and dominant aerosol types) are listed in Table 1.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Figure 1. Geographic locations of 19 typical AERONET sites used in clustering. The basemap is sourced from Tianditu © MNR, China (http://www.tianditu.gov.cn) with Map Approval No. GS(2025)1508.

The aerosol properties selected for clustering are listed below. We chose these properties not only because they can capture various aerosol characteristics without repetition, but also because they have been widely used as aerosol types classification criteria in various studies.

(1)Extinction Angstrom Exponent (EAE). The AOD at different wavelengths(λ) are considered to have the following relationship (Angström, 1964; Russell et al., 2010; Yuan et al., 2023):

$$AOD\left(\lambda\right) = B \times \lambda^{-EAE} \tag{1}$$

Here B is the turbidity coefficient. In AERONET aerosol products, EAE is calculated from AOD at 440 and 870 nm:

$$EAE = -\frac{\ln\left[AOD\left(870\right)/AOD\left(440\right)\right]}{\ln\left(870/440\right)} \tag{2}$$

A small value of EAE indicates that in the aerosol the coarse particles are dominant; a large value of EAE indicates that the fine particles are dominant (Dubovik et al., 2002). EAE is related to aerosol particle size characteristics and is widely used in aerosol types classification (Chen et al., 2018; Evgenieva et al., 2022; Han et al., 2025; Sreekanth, 2014).

(2)Single Scattering Albedo (SSA). SSA is the ratio of the aerosol scattering coefficient(σ_s) to the aerosol extinction coefficient(σ_e) of the entire atmospheric column.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

$$SSA = \sigma_s/\sigma_e \tag{3}$$

SSA can describe the relative magnitude of aerosol scattering and extinction capacity. Since aerosol extinction is composed of aerosol scattering and absorption, SSA is frequently used to distinguish absorbing from non-absorbing aerosols in aerosol types classification (Che et al., 2018; Lee et al., 2010; Sabetghadam et al., 2020).

(3)Sphericity Factor. AERONET inversions provide a percentage value of aerosol particle sphericity, which is related to the aerosol shape characteristic. It can be used to distinguish low spherical aerosols(especially dust) from high spherical aerosols (Hamill et al., 2016; Taylor et al., 2015). Since this aerosol property is not available in AERONET version 3 Level 2.0 datasets, here we can use the sphericity factor from Level 1.5 for supplementation (Roger et al., 2022).

(4)Refractive Index. Refractive Index is composed of 2 parts, namely the Real Part of the Refractive Index (RRI) and the Imaginary Part of the Refractive Index (IRI). The RRI and IRI are indicate aerosol scattering and absorption capacity, respectively (Awais and Wang, 2025). They have been used in many clustering classification cases (Fan et al., 2021; Hamill et al., 2016).

(5)Effective Radius. AERONET inversions can provide the effective radius of aerosols in fine mode and coarse mode. They indicate fine and coarse aerosol particle size characteristics, and were used in many AERONET-based classification cases (Choi et al., 2021a; Wei et al., 2024).

This study only adopts these five discriminative and non-redundant key features for clustering, compared with clustering using numerous aerosol characteristic parameters, which avoids disturbances introduced by non-informative and highly correlated features, and achieves better physical rationality.

AERONET inversion products can provide SSA, RRI, and IRI at 4 different wavelengths; however, aerosol properties at different wavelength are basically describe the same characteristics. Since strongly correlated features would affect the performance of K-means clustering, referring to other studies (Annapurna et al., 2024; Wei et al., 2024), we can use the properties at 440 nm. Finally, there are 7 specific features used in clustering: EAE, SSA_{440nm}, Sphericity Factor, RRI_{440nm}, IRI_{440nm}, Effective Radius(fine) and Effective Radius(coarse).

In addition, for improving the data reliability, we did not use the data points with RRI of 1.33 or 1.6, IRI of 0.0005 or 0.5, and sphericity factor of 0.1 or 99. Because they are the minimum and maximum values in AERONET inversion data. By the end of 2025, there are 18,439 valid data points from these 19 AERONET sites. The volume of data is sufficient for clustering modeling.

2.2 K-means clustering

To create a comprehensive multi-parameter aerosol model with high-quality, K-means clustering algorithm is employed. As an unsupervised machine learning technique, it does not rely on empirical assumptions. K-means clustering algorithm has high computational efficiency, strong robustness, and clear physical meaning; it has been widely adopted in many aerosol types

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

classification studies (Assiri, 2024; Li et al., 2022; Omar et al., 2005). This study uses it to classify AERONET data points into K reference clusters based on the 7 features mentioned in Sect. 2.1.

K-means clustering algorithm was implemented through following steps:

(1) Calculating the mean(μ_i) and standard deviation(σ_i) of various features, then for each feature selecting K points randomly within the interval of (μ_i-σ_i) to (μ_i+σ_i), to compose initial K cluster centroids.

(2) Calculating the normalized Euclidean distance between each data point and K cluster centroids, and then marking each data point with the cluster tag corresponding to the minimum value of K distances.

Here the normalized Euclidean distance (D) was calculated from Equation 4:

$$D = \sqrt{\sum_{i=1}^{7} \left(\frac{x_i - c_i}{\sigma_i} \right)^2} \tag{4}$$

Where x_i is the data point value of feature i, c_i is the cluster centroid of feature i, and σ_i is the standard deviation of the data points of feature i. K normalized Euclidean distances were obtained by calculating with K cluster centroids.

(3) Updating cluster centroids by calculating the average value of each feature within the cluster.

(4) The steps (2) and (3) were iteratively performed until the position change between the old cluster centroids and the new cluster centroids was less than the set threshold value.

In addition, K-means clustering requires the determination of the number of clusters (K). The choice of K directly influences the final clustering results. In this study, the Elbow method is adopted to select the optimal value of K. The Elbow method selects K by analyzing the decline in the within-cluster sum of squared distances (WCSS) as K increases. When K reaches a turning point, the rate of decline in WCSS obviously slows down. This turning point is defined as the "elbow point", which is regarded as the optimal K (Elmourssi et al., 2026). WCSS can be calculated from Equation 5:

$$WCSS^{(K)} = \sum_{k=1}^{K} \sum_{\boldsymbol{x} \in C^{(k)}} \|x_i - c_i^{(k)}\|^2 \tag{5}$$

Where $WCSS^{(K)}$ denotes the WCSS of K, $C^{(k)}$ is the cluster k, x_i is the value of data point $\boldsymbol{x}$ of feature i, and $c_i^{(k)}$ is the cluster k centroid of feature i. The WCSS of different K values were calculated and demonstrated in Fig. 2. As is shown in Fig. 2, K = 7 can be recognized as the "elbow point". This value is finally selected for K-means clustering in this study.

2.3 Aerosol model analysis

The K-means clustering results are shown in Table 2. There are 7 clusters classified from the AERONET data.

From clustering results, we found that the basic aerosol types "Mixed", "Biomass Burning", and "Dust" were each classified into two clusters. Some previous studies have reported similar further classifications of these basic aerosol types (Cattrall et al., 2005; Hamill et al., 2016; Wei et al., 2024), therefore these clustering results meet our primary expectations. Each aerosol type cluster and its multi-parameter characteristics are analyzed below:

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

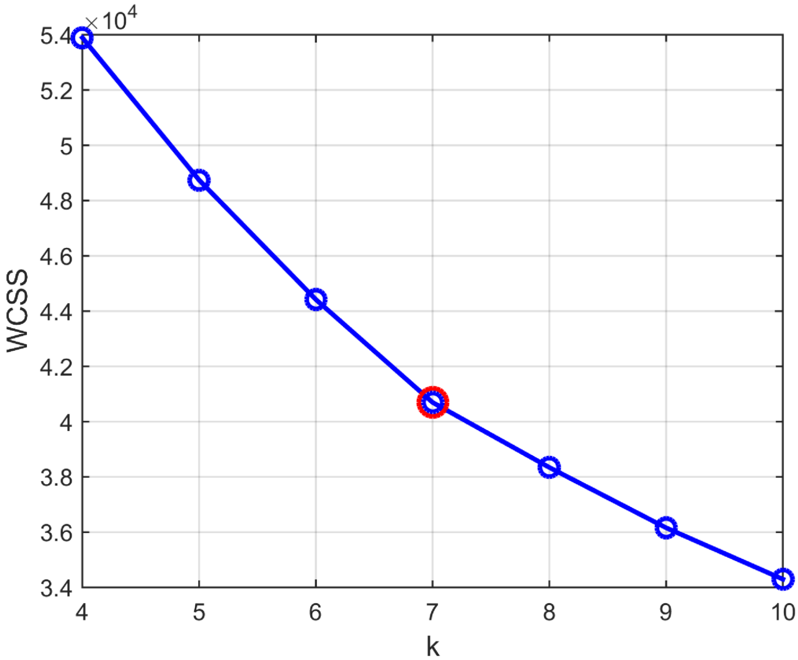

Figure 2. The WCSS calculation results of different K values.

Table 2. The centroids of different clusters.

Aerosol Types	EAE	SSA	Sphericity Factor	RRI	IRI	ER(F)[a]	ER(C)[b]	N[c]
Mixed 1	1.13	0.94	23.88	1.50	0.0059	0.167	2.08	2460
Mixed 2	1.21	0.97	62.57	1.45	0.0042	0.239	2.69	2157
Biomass Burning 1	1.66	0.93	95.23	1.49	0.0112	0.164	2.63	2739
Biomass Burning 2	1.68	0.86	85.54	1.50	0.0249	0.137	2.65	1508
Dust 1	0.42	0.91	2.62	1.42	0.0051	0.118	1.77	3351
Dust 2	0.39	0.90	2.91	1.51	0.0052	0.114	1.93	3311
Urban/Industrial	1.61	0.96	88.80	1.40	0,0046	0.173	2.45	2913

[a] ER(F): Effective Radius(fine); [b] ER(C): Effective Radius(coarse); [c] N is the number of data points in the cluster.

(1) Mixed 1. Most of the parameters of "Mixed 1" are basically at normal levels. "Mixed 1" exhibits relatively high SSA and RRI, indicating that it has relatively stronger scattering effects. However, its sphericity factor is significantly lower than other aerosol types (except for two "Dust" clusters). Given that low sphericity is the typical characteristic of dust particles, and some previous studies have classified the "Mixed" aerosol type with low sphericity as "Dust dominated mixture" (Mukhopadhyay et al., 2025), it can be concluded that "Mixed 1" contains prominent dust components.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

(2) Mixed 2. Compared with "Mixed 1", "Mixed 2" shows a further increase in SSA to the highest, and an obvious increase in sphericity factor. The significant increases in ER(F) and ER(C) also indicate larger particle sizes for each aerosol mode. Considering the geographic distribution of clustering results, "Mixed 1" mainly occurs in the northern regions of East Asia, whereas "Mixed 2" is mainly found in the southern regions. Referring to other studies, it can be inferred that the dust components of "Mixed" undergo mixing and coating with various pollutants during long-range transport, which results in an increase in SSA and more spherical particles (Khatri et al., 2014; Pan et al., 2017).

(3) Biomass Burning 1. "Biomass Burning 1" exhibits high EAE and sphericity factor, indicating it is dominated by fine-mode and spherical particles. Its IRI is obviously higher than other aerosol types(except for "Biomass Burning 2"), reflecting the strong absorption effects of particles such as black carbon generated during combustion. The parameters of "Biomass Burning 1" generally represent typical characteristics of biomass burning aerosols.

(4) Biomass Burning 2. Compared with "Biomass Burning 1", it has obviously lower SSA, and the IRI value is twice higher. These reflect stronger absorbing effects, indicating a considerably higher proportion of absorbing components in "Biomass Burning 2". This difference can also be explained by their geographic distribution. Clustering results show that "Biomass Burning 1" mainly occurs in South America, while "Biomass Burning 2" is mainly found over southern Africa. This matches the hypothesis proposed in previous studies: South American biomass burning aerosols mainly originate from smoldering fires, while those in southern Africa are flaming fires. Flaming fires release large amounts of black carbon content, and make biomass burning aerosols with stronger absorption in southern Africa (Hamill et al., 2016).

(5) Dust 1. "Dust 1" exhibits remarkably low EAE and sphericity factor, reflecting the dominance of coarse-mode and strongly non-spherical particles, which are the two most typical characteristics of dust aerosols. Its relatively low SSA is also consistent with the typical characteristics of dust aerosols.

(6) Dust 2. Compared with "Dust 1", "Dust 2" has slightly lower EAE and higher RRI and ER(C). According to the clustering results, "Dust 1" mostly occurs around the Saharan region of Africa, while "Dust 2" is mainly distributed in the Arabian region. Differences in source regions are likely responsible for these parameter discrepancies. Previous studies have also reported that aerosols with radii greater than 2 μm contribute more significantly to Arabian dust (Mateos et al., 2026).

(7) Urban/Industrial. "Urban/Industrial" exhibits relatively high EAE, SSA, and sphericity factor, reflecting that it is dominated by scattering spherical fine-mode particles. Its overall parameters are consistent with typical characteristics of urban aerosols. In addition, only one "Urban/Industrial" cluster is classified, confirming that such anthropogenic aerosols have similar characteristics even across various source regions.

The analysis of the multi-parameter characteristics of each aerosol type cluster verifies the physical rationality of the seven established multi-parameter aerosol models. These clustering results are capable of representing typical aerosol types and their subtypes. The multi-parameter aerosol models and their corresponding datasets can serve as baseline datasets for future research.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

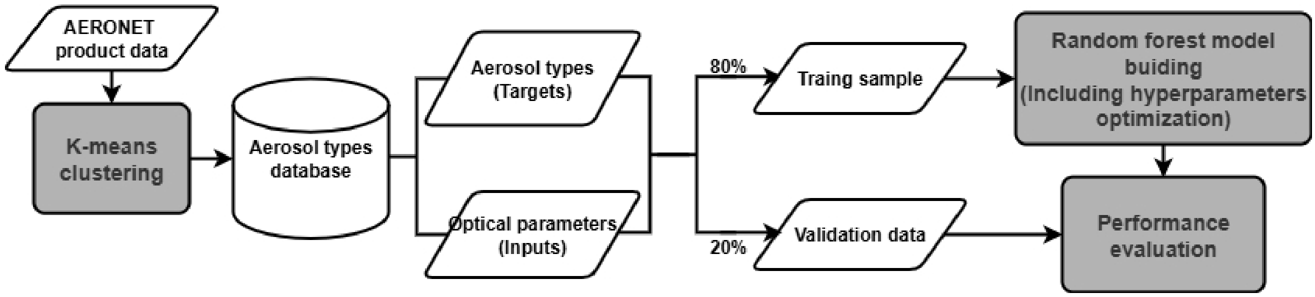

Figure 3. Schematic diagram of random forest-based aerosol types classifier construction process.

3 Aerosol types classifier

The schematic diagram of aerosol types classifier construction process is shown in Fig. 3. In this study, we use the established multi-parameter aerosol models' datasets to train the classifier. Compared to other studies that rely on simulated data or subjectively labeled samples for training, our trained classifier is considered to enable a more objective classification.

3.1 Input parameters

We want to construct an aerosol types classifier based on active-passive remote sensing directly retrieved optical parameters. Some easily available and representative optical parameters were selected as inputs. Since complex retrieved aerosol properties are mostly derived from these directly retrieved aerosol optical parameters, it can be considered that the classifier based on these directly retrieved optical parameters is feasible. These optical parameters are listed below:

(1) Aerosol Optical Depth (AOD). AOD indicates the aerosol extinction of the entire atmospheric column. AOD is related to aerosol absorption capacity and is frequently used in aerosol types classification (Alemu and Raju, 2024; Kumar et al., 2015; Mao et al., 2019; Tan et al., 2015). As a widely used basic aerosol optical parameter, AOD can be easily obtained from sun photometer or other passive optical remote sensing instruments.

(2) Extinction Angstrom Exponent (EAE). As is mentioned in Sect. 2.1, EAE is related to aerosol particle size characteristics, and it can be directly calculated from AOD at 2 different wavelengths.

(3) Lidar Ratio (LR). LR is the ratio of the aerosol extinction coefficient(σ_e) to the aerosol backscattering coefficient(β_s):

$$LR = \sigma_e / \beta_s \tag{6}$$

LR can describe the relative magnitude of aerosol extinction and backscattering capacity, which is related not only to aerosol absorbing and scattering capacity but also to aerosol particle size characteristics. LR is also frequently used in classification (Groß et al., 2013; Noh et al., 2007; Papagiannopoulos et al., 2018). LR can be detected directly from Raman lidar or High

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Table 3. The final determined hyperparameters of the random forest classifier.

Hyperparameters	Value
n_estimators	450
max_depth	25
max_features	10
min_samples_split	5
min_samples_leaf	2

Spectral Resolution Lidar(HSRL) (Haarig et al., 2025; Yao et al., 2023). It can also be obtained from Mie-scattering lidar data by introducing the iterative procedure (Mortier et al., 2013; Yang et al., 2026; Zhang et al., 2022).

(4) Depolarization Ratio (DR). DR describes the aerosol depolarizing ability. Many studies have used DR to distinguish spherical from non-spherical aerosols (Burton et al., 2012; Lin et al., 2022). DR can be obtained from the lidar equipped with depolarization channels (Floutsi et al., 2023; Noh et al., 2017).

All these input parameters can be directly provided by AERONET Level 2.0 aerosol products. However, except for EAE, other optical parameters here were provided at 4 various wavelengths(440, 675, 870 and 1020nm) in AERONET. Since random forest algorithm result is not sensitive to strongly correlated inputs, AOD, LR, DR at 4 wavelengths were all used in this study, namely, a total of 13 parameters were taken as inputs.

3.2 Classifier training

Random forest algorithm is employed in this study to train the aerosol types classifier. Random forest classifier is composed of an ensemble of decision trees, which were constructed through training with bootstrap sampling and random feature selection. Its classification results are determined by majority voting. Random forest classifier exhibits good robustness when processing high-dimensional data, and it can effectively overcome the problem of overfitting in single decision tree classifier (Breiman, 2001; Lan et al., 2020).

The hyperparameters of the random forest classifier can affect its performance seriously. In this study the hyperparameters were optimized by using grid-searching method. The final determined hyperparameters are listed in Table 3. Where "n_estimators" refers to the number of decision tree, 'max_depth' is the maximum depth of decision tree, "max_features" is the maximum number of features considered, "min_samples_split" is the minimum number of samples required to split a node, and "min_samples_leaf" is the minimum number of samples for nodes.

In this study, 80% of the database data were used as training samples, and the remaining 20% were stored separately as validation data. According to the hyperparameters determined above, a random forest-based aerosol type classifier is constructed. The out-of-bag accuracy of the constructed classifier is calculated to be 78.27%.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Table 4. The MDI of each input parameter.

Input parameters	Importance value (MDI)
DR (1020nm)	0.151
EAE	0.148
DR (440nm)	0.135
LR (440nm)	0.109
LR (675nm)	0.109
LR (1020nm)	0.075
DR (870nm)	0.071
LR (870nm)	0.068
DR (675nm)	0.047
AOD (440nm)	0.032
AOD (675nm)	0.02
AOD (1020nm)	0.019
AOD (870nm)	0.016

3.3 Parameter importance analysis

To identify which input parameters have greater influence on aerosol types classification, an importance analysis is conducted by calculating the Mean Decrease Impurity(MDI) for each input parameter. A higher MDI indicates that this feature plays a more significant role in classification. The normalized MDI for all input parameters were calculated and presented in Table 4, where the input parameters have been sorted in descending order of MDI.

As shown in Table 4, all inputs are informative for the classification. DR, EAE, and LR exhibit greater influence on classification, whereas AOD contributes relatively less to the classification performance. These analysis results may provide an important reference for further research improvements.

3.4 Evaluation

In this study, only 80% of the database data were randomly selected as training samples; the remaining 20% were stored as an independent validation dataset. The classification performance of the random forest classifier can be evaluated based on its outputs of this validation dataset. The confusion matrix is shown in Fig. 4. The accuracies of various aerosol types are given in Table 5.

From the evaluation results, it can be seen that the proposed random forest-based aerosol types classifier has good performance in each aerosol type. Even the smallest accuracy of these aerosol types is more than 70%, and the overall accuracy is nearly 80%. This classifier is considered to be effective, exhibiting good generalization ability and strong robustness. In addition, these selected inputs have also been proven to be capable of capturing aerosol characteristics.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

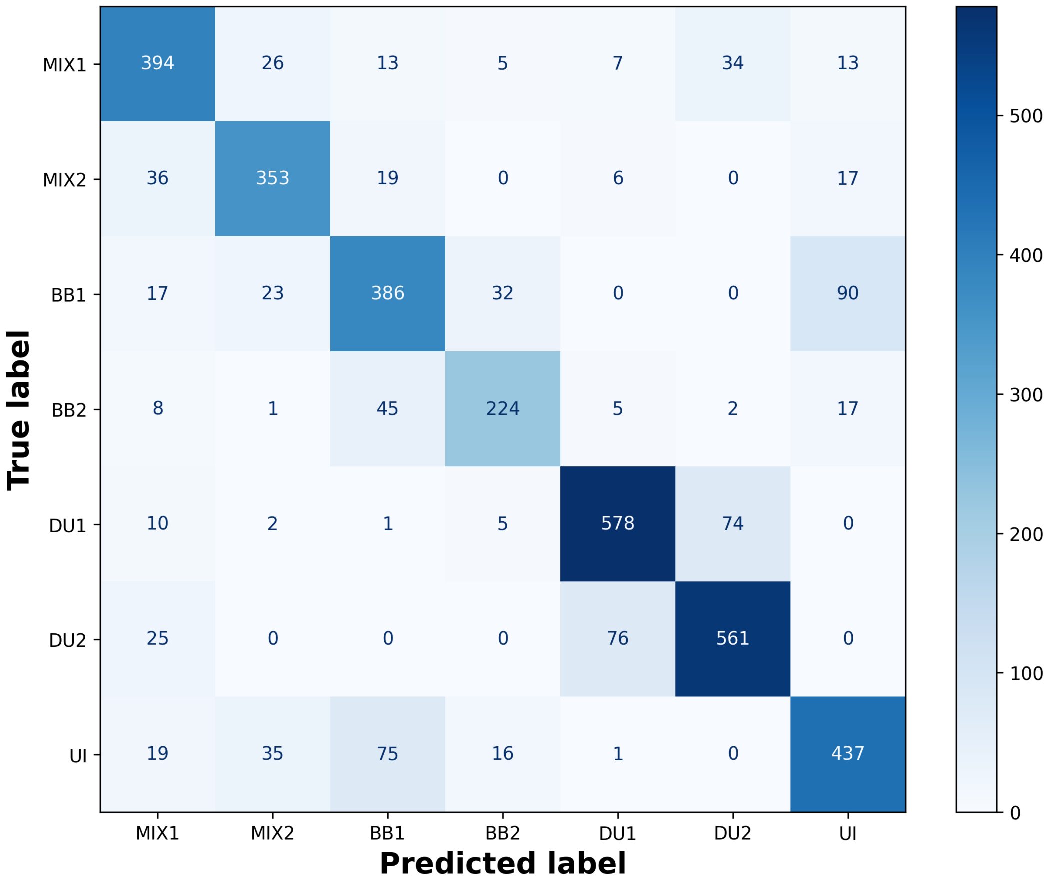

Figure 4. Confusion matrix of random forest-based aerosol types classifier. Here MIX1 denotes "Mixed 1"; MIX2 denotes "Mixed 2"; BB1 denotes "Biomass Burning 1"; BB2 denotes "Biomass Burning 2"; DU1 denotes "Dust 1"; DU2 denotes "Dust 2"; and UI denotes "Urban/Industrial".

4 Experiment and results

To verify the practical feasibility and generalization ability of the proposed aerosol types classifier, classification experiments were conducted using AERONET Version 3 Level 2.0 data from 27 typical global sites (up to the end of 2025). These sites include not only the 19 typical sites used in clustering, but also 8 additional typical sites that are also recognized as dominated by single aerosol types, with sufficient high-quality data. The selection of these 8 new sites also considered their representativeness of typical aerosol types and source regions. For instance, "Osaka" and "XiangHe" in East Asian source region were chosen as representative sites for "Mixed"; "Concepcion" in South America and "Lake_Argyle" in Australia were selected for "Biomass Burning". No additional sites in South Africa were included, because there remains controversy over

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Table 5. The accuracies of various aerosol types.

Aerosol types	Accuracy
Mixed 1	80.08%
Mixed 2	81.90%
Biomass Burning 1	70.44%
Biomass Burning 2	74.17%
Dust 1	86.27%
Dust 2	84.74%
Urban/Industrial	74.96%
All types	79.53%

whether the other high-quality sites in this region can represent typical "Biomass Burning" (Eck et al., 2003). "Agoufou"(North Africa) and "Mezaira"(Arabia) were selected for "Dust". "Paris"(Europe) and "Wallops"(North America) were adopted as "Urban/Industrial" sites. The classification results of these 27 sites are presented in the Table 6.

As shown by classification results, the dominant aerosol types of 27 AERONET sites were all classified correctly.

(1) The dominant aerosol types at "Mixed" sites in East Asia were classified as either "Mixed 1" or "Mixed 2". Except for "Chen-Kung_Univ", which was classified as "Mixed 2" dominated, the remaining sites were all classified as "Mixed 1" dominated. Their classification results are also shown in Fig. 5. According to the aerosol model analysis in Sect. 2.3, the northern East Asian sites("Mixed 1") are more affected by dust components, which is also reported by pervious study (Nie et al., 2025). The additional sites "Osaka" and "XiangHe" are both located in relatively northern regions; it is reasonable that they are classified as "Mixed 1" dominated.

(2) The dominant aerosol types at "Biomass Burning" sites in South America were classified as "Biomass Burning 1", in South Africa were classified as "Biomass Burning 2". Besides, the Australian site "Lake_Argyle" was classified as "Biomass Burning 2" dominated. Their classification results are also shown in Fig. 6, Fig. 7, and Fig. 8, respectively. These classification results reveal obvious differences in dominant aerosol types among various source regions; this is consistent with the hypothesis cited in the aerosol model analysis that flaming fires lead to stronger absorption of aerosols in southern Africa. The Australian site results may indicate that biomass burning aerosols in this region originate mainly from flaming fires.

(3) For "Dust" sites, those in North Africa were classified as "Dust 1", sites in Arabia were classified as "Dust 2". Their classification results are also shown in Fig. 9 and Fig. 10. There are also obvious differences in dominant aerosol types between these two source regions. Some previous observations have reported that the North African dust have significantly higher lidar ratio value than Arabian dust (Mamouri et al., 2013; Müller et al., 2007). Since LR has been used as one of the inputs of the classifier, it can effectively distinguish "Dust 1" from "Dust 2".

(4) "Urban/Industrial" sites in both North America and Europe were all classified as "Urban/Industrial" dominated. The classification results are shown in Fig. 11 and Fig. 12. As mentioned in the aerosol model analysis, such anthropogenic

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Table 6. Aerosol types classification results of 27 AERONET sites.

Sites	Mixed 1	Mixed 2	Biomass Burning 1	Biomass Burning 2	Dust 1	Dust 2	Urban/Industrial
Anmyon	39.75%*	23.43%	5.60%	0.39%	5.13%	4.73%	20.98%
Beijing	29.02%*	18.47%	21.22%	8.68%	3.97%	10.14%	8.50%
Chen-Kung_Univ	17.68%	43.50%*	7.18%	0.54%	0.27%	0	30.83%
Gosan_SNU	36.09%*	16.13%	4.64%	0.20%	13.31%	7.26%	22.38%
Abracos_Hill	0	0	88.11%*	4.51%	0	0	7.38%
Alta_Floresta	1.30%	0	78.55%*	4.20%	0	0	15.94%
Cuiaba	1.19%	0	83.33%*	10.71%	0	0	4.76%
Mongu	0	0.08%	26.33%	71.73%*	0	0	1.86%
Senanga	0	0	15.20%	84.00%*	0	0	0.80%
Zambezi	0	0	4.44%	95.56%*	0	0	0
Banizoumbou	3.75%	1.26%	0.04%	2.18%	69.83%*	22.94%	0
Capo_Verde	3.50%	0.95%	0	2.44%	74.13%*	18.66%	0.32%
Mussafa	28.43%	0.99%	0.17%	0.50%	17.19%	43.64%*	9.09%
Solar_Village	6.63%	0.16%	0.28%	0.28%	17.38%	74.99%*	0.28%
GSFC	3.91%	16.02%	31.64%	0.26%	0	0	48.18%*
MD_Science_Center	3.47%	9.06%	18.88%	1.36%	0	0.15%	67.07%*
Ispra	9.98%	21.39%	19.16%	3.92%	0.89%	0.71%	43.94%*
Leipzig	13.16%	13.82%	17.11%	0	3.62%	1.32%	50.99%*
Lille	11.29%	23.04%	24.88%	1.15%	4.84%	1.38%	33.41%*
Osaka	38.60%*	10.46%	10.68%	1.67%	4.12%	4.89%	29.59%
XiangHe	26.92%*	16.74%	21.85%	6.73%	5.30%	6.90%	15.56%
Concepcion	0	0	58.97%*	5.13%	0	0	35.90%
Lake_Argyle	1.77%	0	14.34%	60.71%*	1.06%	0.71%	21.42%
Agoufou	2.54%	0.66%	0	1.22%	58.78%*	36.35%	0.44%
Mezaira	14.36%	1.09%	0.31%	1.68%	12.80%	67.53%*	2.22%
Paris	9.35%	23.58%	13.01%	2.44%	2.44%	2.85%	46.34%*
Wallops	2.55%	11.28%	33.62%	0.21%	0	0	52.34%*

$*$: The dominant aerosol type.

280 aerosols have similar characteristics across different source regions. It is reasonable that all these sites are classified as "Urban/Industrial" dominated.

These classification results are all reasonable. This demonstrates that the proposed classifier can not only accurately classify the dominant aerosol types in different regions based on real detection data, but also is capable of further classifying basic

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

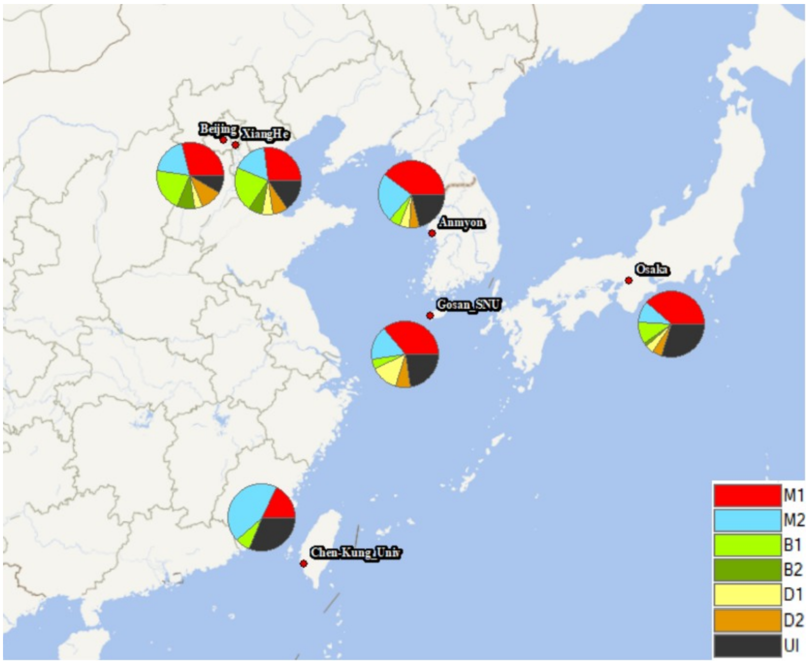

Figure 5. Classification results of sites in East Asian source region. M1 denotes "Mixed 1"; M2 denotes "Mixed 2"; B1 denotes "Biomass Burning 1"; B2 denotes "Biomass Burning 2"; D1 denotes "Dust 1"; D2 denotes "Dust 2"; and UI denotes "Urban/Industrial".

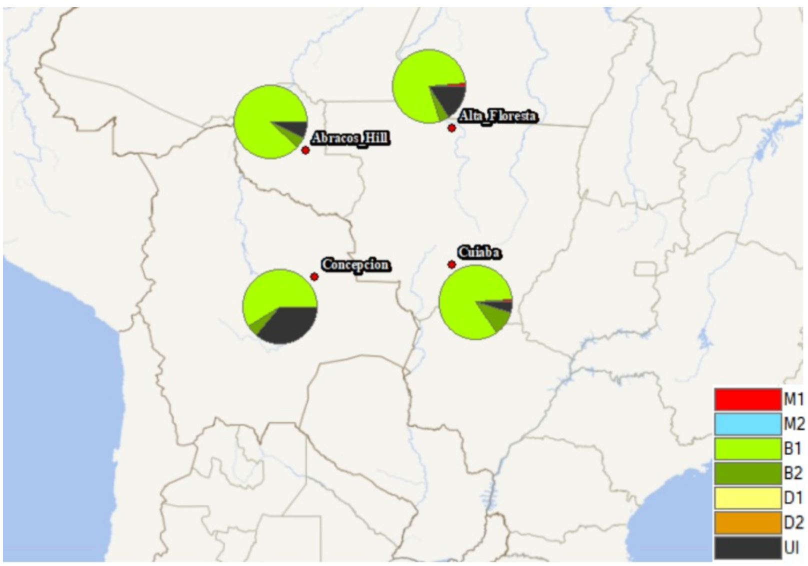

Figure 6. Classification results of sites in South American source region.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

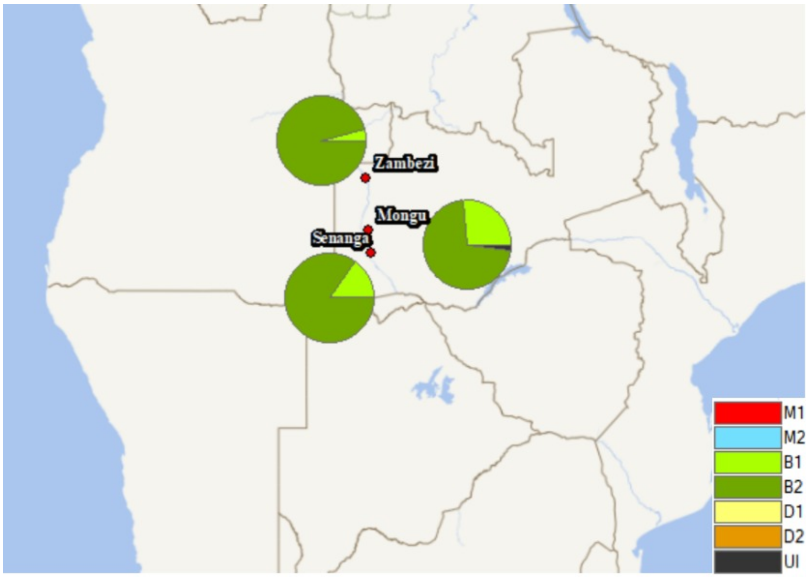

Figure 7. Classification results of sites in South African source region.

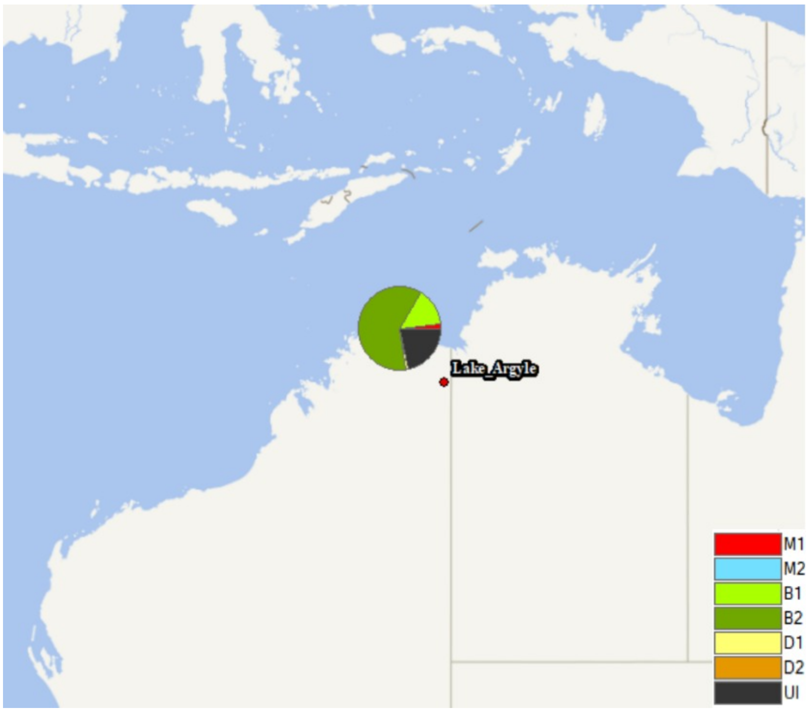

Figure 8. Classification results of sites in Australian source region.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

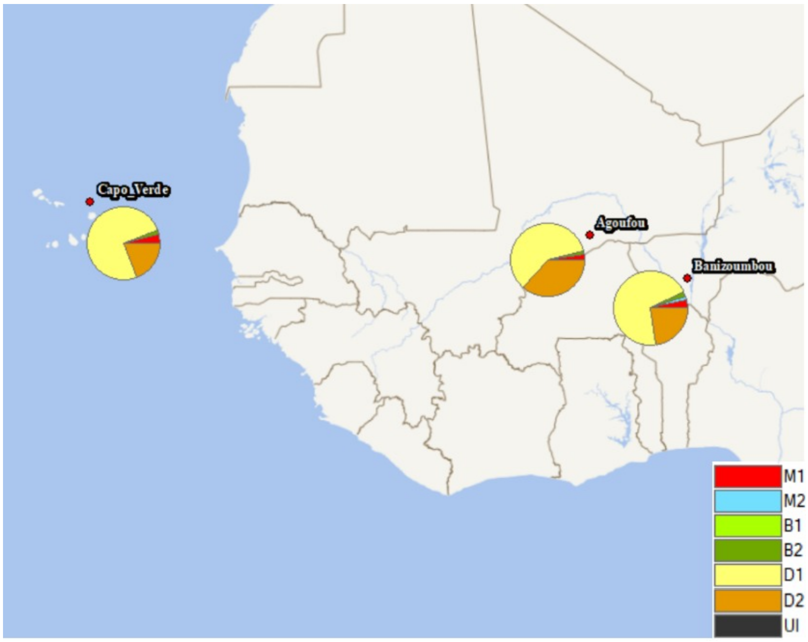

Figure 9. Classification results of sites in North African source region.

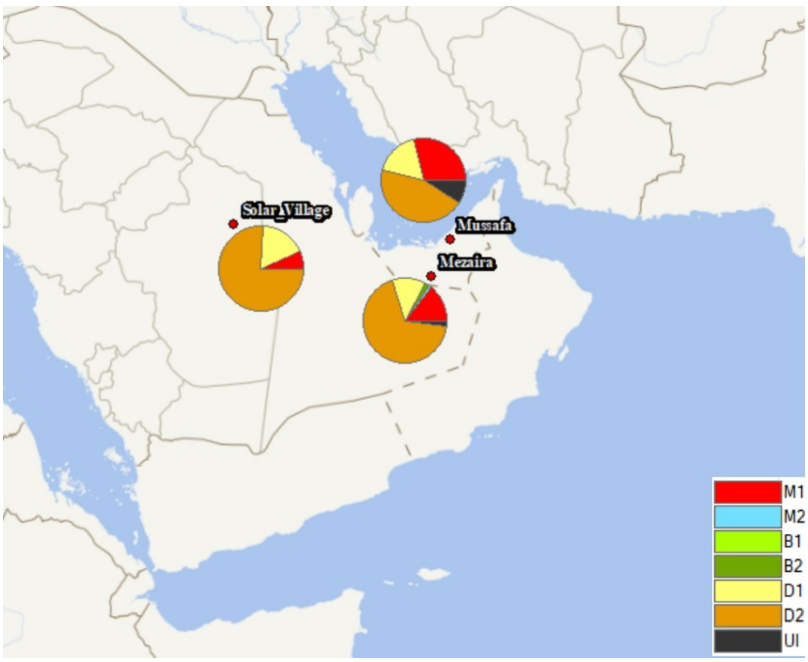

Figure 10. Classification results of sites in Arabian source region.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

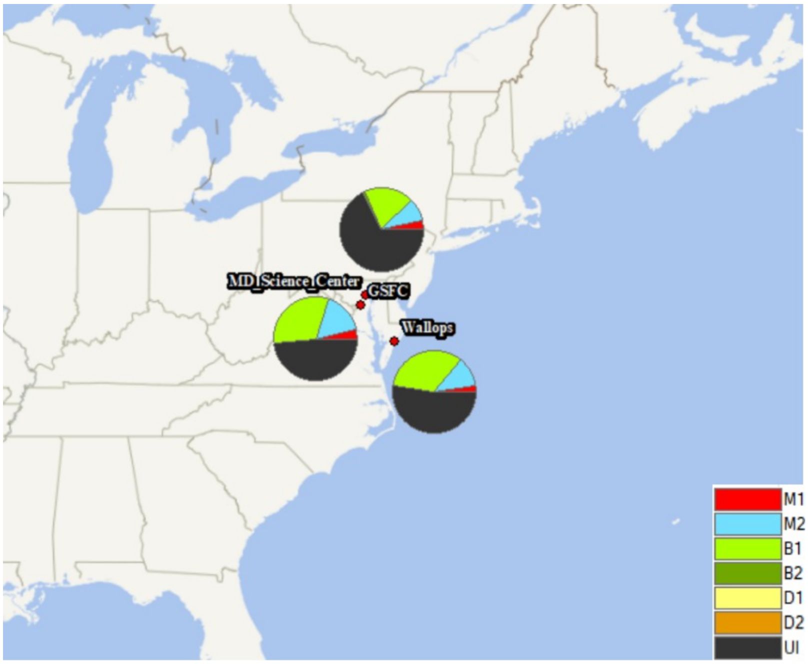

Figure 11. Classification results of sites in North American source region.

aerosol types into subtypes. The proposed classifier has been proven to be feasible in practice, with satisfactory classification performance and generalization ability.

5 Conclusions

Conventional aerosol types classification methods mostly suffer from limited remote sensing instruments and empirical classification schemes. To address these limitations, this study proposes a systematic aerosol types classification method based on active-passive joint optical remote sensing measurements and machine learning-enhanced classification scheme. By building global multi-parameter aerosol models and developing an aerosol types classifier that takes active-passive remote sensing directly retrieved parameters as inputs, this study achieves a more universally applicable, comprehensive, and accurate aerosol types classification. Compared with previous machine learning-enhanced aerosol types classification studies, this study requires only active-passive remote sensing directly retrieved parameters that are easily available, rather than complex retrieved aerosol properties, so that it can be directly used in automated classification of field observations; Furthermore, this study is based on the global comprehensive multi-parameter aerosol model database established by clustering quality-ensured and representative real measurement data, rather than empirically labeled data or simulated data, which further enhances its classification accuracy and objectivity. Evaluated with validatioin data divided from database, the overall accuracy of the classifier is nearly 80%, demonstrating good generalization ability and strong robustness. The AERONET version 3 Level 2.0 data from 27 typical sites

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

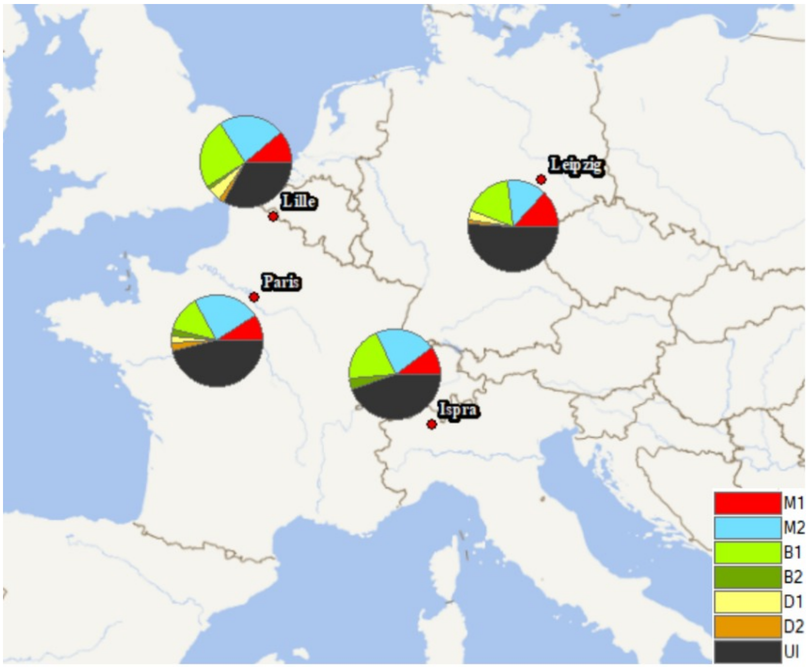

Figure 12. Classification results of sites in European source region.

were also used to test the practical feasibility of the proposed classifier. Classification results were reasonable and proved its classification capability. This systematic aerosol types classification method has proven to be effective and is expected to be directly applied in multi-device field detection studies, offering a valuable tool for advancing atmospheric aerosol research. It has significant implications for atmospheric physics, especially in improving radiative forcing calculations and atmospheric radiative transfer modeling.

Atmospheric aerosols are generally continuous mixtures, and the seven classes established in this study represent a discrete simplification of this continuum. These classes are intend to describe the dominant optical state of the aerosol rather than its chemical categories or mixing state. Such discrete representations are the commonly adopted operational approach for regional source tracking and aerosol model selection in atmospheric remote sensing research (Levy et al., 2013; Omar et al., 2009). Accordingly, the proposed classifier does not aim to quantify the aerosol mixing state; rather, it provides a consistent descriptor of the dominant aerosol optical type derived directly from active and passive remote sensing parameters. It is complementary to continuous microphysical aerosol characterizations.

In this study, we only use quality-ensured AERONET Level 2.0 product data for database establishment, which led to the exclusion of the aerosol type "Maritime". Although this exclusion is commonly adopted in many similar studies, it indeed limits the geographical applicability of our research. How to improve this point in the future is worthy of further study.

Data availability. The data used in this study are publicly available from the NASA Aerosol Robotic Network website(https://aeronet.gsfc.nasa.gov/).

Author contributions. ZY: Conceptualization, Methodology, Software, Formal analysis, Validation, Writing- Original draft preparation. QX: Writing - review & editing, Supervision. CC: Supervision, Funding acquisition.

Competing interests. The authors have no competing interests to declare.

Acknowledgements. This research was funded by the National Natural Science Foundation of China, grant number 42105082. The authors would like to thank the AERONET team for providing the data used in this study.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

References

Alemu, A. A. and Raju, J. P.: Temporal distributions of aerosols over the Horn of Africa–Ethiopia using MODIS satellite data: Part 01, J. Quant. Spectrosc. Radiat. Transfer, 325, 109 085, https://doi.org/10.1016/j.jqsrt.2024.109085, 2024.

Angström, A.: The parameters of atmospheric turbidity, Tellus A, 16, 64–75, https://doi.org/10.1111/J.2153-3490.1964.TB00144.X, 1964.

Annapurna, S., Anitha, M., and Kumar, L. S.: Composition and source based aerosol classification using machine learning algorithms, Adv. Space Res., 73, 474–497, https://doi.org/10.1016/j.asr.2023.09.068, 2024.

Assiri, M. E.: Aerosol Types Identification over the Arabian Peninsula Using AERONET Products: Evaluation with Multisource Datasets, Earth Syst. Environ., 8, 483 – 499, https://doi.org/10.1007/s41748-024-00389-x, 2024.

Awais, M. and Wang, L.: Global classification of aerosols based on ground observations: A comparison between an empirical method and machine learning algorithms, Phys Chem Earth, 140, 104 017, https://doi.org/10.1016/j.pce.2025.104017, 2025.

Barnaba, F. and Gobbi, G. P.: Aerosol seasonal variability over the Mediterranean region and relative impact of maritime, continental and Saharan dust particles over the basin from MODIS data in the year 2001, Atmos. Chem. Phys., 4, 2367–2391, https://doi.org/10.5194/acp-4-2367-2004, 2004.

Breiman, L.: Random Forests, Mach. Learn., 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Burton, S. P., Ferrare, R. A., Hostetler, C. A., Hair, J. W., Rogers, R. R., Obland, M. D., Butler, C. F., Cook, A. L., Harper, D. B., and Froyd, K. D.: Aerosol classification using airborne High Spectral Resolution Lidar measurements—Methodology and examples, Atmos. Meas. Tech., 5, 73–98, https://doi.org/10.5194/amt-5-73-2012, 2012.

Carslaw, K. S., Lee, L. A., Reddington, C. L., Pringle, K. J., Rap, A., Forster, P. M., Mann, G. W., Spracklen, D. V., Woodhouse, M. T., and Regayre, L. A.: Large contribution of natural aerosols to uncertainty in indirect forcing, Nature, 503, 67–71, https://doi.org/10.1038/nature12674, 2013.

Cattrall, C., Reagan, J., Thome, K., and Dubovik, O.: Variability of aerosol and spectral lidar and backscatter and extinction ratios key aerosol types derived from selected Aerosol Robotic Network locations, J. Geophys. Res.: Atmos., 110, 2004JD005 124, https://doi.org/10.1029/2004JD005124, 2005.

Che, H., Qi, B., Zhao, H., Xia, X., Eck, T. F., Goloub, P., Dubovik, O., Estelles, V., Cuevas-Agulló, E., Blarel, L., Wu, Y., Zhu, J., Du, R., Wang, Y., Wang, H., Gui, K., Yu, J., Zheng, Y., Sun, T., Chen, Q., Shi, G., and Zhang, X.: Aerosol optical properties and direct radiative forcing based on measurements from the China Aerosol Remote Sensing Network (CARSNET) in eastern China, Atmos. Chem. Phys., 18, 405–425, https://doi.org/10.5194/acp-18-405-2018, 2018.

Chen, Q., Yuan, Y., Huang, X., He, Z., and Tan, H.: Assessment of column aerosol optical properties using ground-based sun-photometer at urban Harbin, Northeast China, J. Environ. Sci., 74, 50–57, https://doi.org/10.1016/j.jes.2018.02.003, 2018.

Choi, W., Kang, H., Shin, D., and Lee, H.: Satellite-Based Aerosol Classification for Capital Cities in Asia Using a Random Forest Model, Remote Sens., 13, 2464, https://doi.org/10.3390/rs13132464, 2021a.

Choi, W., Lee, H., Kim, D., and Kim, S.: Improving Spatial Coverage of Satellite Aerosol Classification Using a Random Forest Model, Remote Sens., 13, 1268, https://doi.org/10.3390/rs13071268, 2021b.

Del Aguila, A., Ortiz-Amezcua, P., Tabik, S., Bravo-Aranda, J. A., Fernandez-Carvelo, S., and Alados-Arboledas, L.: Aerosol type classification with machine learning techniques applied to multiwavelength lidar data from EARLINET, Atmos. Chem. Phys., 25, 12 549–12 567, https://doi.org/10.5194/acp-25-12549-2025, 2025.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D., and Slutsker, I.: Variability of Absorption and Optical Properties of Key Aerosol Types Observed in Worldwide Locations, J. Atmos. Sci., 59, 590–608, https://doi.org/10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2, 2002.

Eck, T. F., Holben, B. N., Ward, D. E., Mukelabai, M. M., Dubovik, O., Smirnov, A., Schafer, J. S., Hsu, N. C., Piketh, S. J., Queface, A., Roux, J. L., Swap, R. J., and Slutsker, I.: Variability of biomass burning aerosol optical characteristics in southern Africa during the SAFARI 2000 dry season campaign and a comparison of single scattering albedo estimates from radiometric measurements, J. Geophys. Res.: Atmos., 108, 8477, https://doi.org/10.1029/2002JD002321, 2003.

Elmourssi, D. M., El-Assy, A. M., and Amer, H. M.: Clustering and machine learning techniques identify air pollution regimes in Greater Cairo, Sci. Rep., 16, 14 038, https://doi.org/10.1038/s41598-026-49777-5, 2026.

Evgenieva, T., Gurdev, L., Toncheva, E., and Dreischuh, T.: Optical and Microphysical Properties of the Aerosol Field over Sofia, Bulgaria, Based on AERONET Sun-Photometer Measurements, Atmosphere, 13, 884, https://doi.org/10.3390/atmos13060884, 2022.

Falah, S., Mhawish, A., Omar, A. H., Sorek-Hamer, M., Lyapustin, A. I., Banerjee, T., Kizel, F., and Broday, D. M.: Intercomparison of Aerosol Types Reported as Part of Aerosol Product Retrieval over Diverse Geographic Regions, Remote Sens., 14, 3667–3667, https://doi.org/10.3390/rs14153667, 2022.

Fan, Y., Sun, X., Huang, H., Ti, R., and Liu, X.: The primary aerosol models and distribution characteristics over China based on the AERONET data, J. Quant. Spectrosc. Radiat. Transfer, 275, 107 888, https://doi.org/10.1016/j.jqsrt.2021.107888, 2021.

Filioglou, M., Giannakaki, E., Backman, J., Kesti, J., Hirsikko, A., Engelmann, R., O'Connor, E., Leskinen, J. T. T., Shang, X., Korhonen, H., Lihavainen, H., Romakkaniemi, S., and Komppula, M.: Optical and geometrical aerosol particle properties over the United Arab Emirates, Atmos. Chem. Phys., 20, 8909–8922, https://doi.org/10.5194/acp-20-8909-2020, 2020.

Floutsi, A. A., Baars, H., Engelmann, R., Althausen, D., Ansmann, A., Bohlmann, S., Heese, B., Hofer, J., Kanitz, T., Haarig, M., Ohneiser, K., Radenz, M., Seifert, P., Skupin, A., Yin, Z., Abdullaev, S. F., Komppula, M., Filioglou, M., Giannakaki, E., Stachlewska, I. S., Janicka, L., Bortoli, D., Marinou, E., Amiridis, V., Gialitaki, A., Mamouri, R.-E., Barja, B., and Wandinger, U.: DeLiAn – a growing collection of depolarization ratio, lidar ratio and Angstrom exponent for different aerosol types and mixtures from ground-based lidar observations, Atmos. Meas. Tech., 16, 2353–2379, https://doi.org/10.5194/amt-16-2353-2023, 2023.

Giles, D. M., Holben, B. N., Eck, T. F., Sinyuk, A., Smirnov, A., Slutsker, I., Dickerson, R. R., Thompson, A. M., and Schafer., J. S.: An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source regions, J. Geophys. Res.: Atmos., 117, D17 203, https://doi.org/10.1029/2012JD018127, 2012.

Groß, S., Esselborn, M., Weinzierl, B., Wirth, M., Fix, A., and Petzold, A.: Aerosol classification by airborne high spectral resolution lidar observations, Atmos. Chem. Phys., 13, 2487–2505, https://doi.org/10.5194/acp-13-2487-2013, 2013.

Haarig, M., Engelmann, R., Baars, H., Gast, B., Althausen, D., and Ansmann, A.: Discussion of the spectral slope of the lidar ratio between 355 and 1064 nm from multiwavelength Raman lidar observations, Atmos. Chem. Phys, 25, 7741–7763, https://doi.org/10.5194/acp-25-7741-2025, 2025.

Hamill, P., Giordano, M., Ward, C., Giles, D., and Holben, B.: An AERONET-based aerosol classification using the Mahalanobis distance, Atmos. Environ., 140, 213–233, https://doi.org/10.1016/j.atmosenv.2016.06.002, 2016.

Han, K. M., Jung, C. H., Song, C. H., Koo, J. H., Yoon, Y. J., Lee, B. Y., Kim, H. S., and Seo, S.: Trends and classification of aerosol observed from MODIS sensor over Northern Europe and the Arctic, Atmos. Pollut. Res., 16, 102 329, https://doi.org/10.1016/j.apr.2024.102329, 2025.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Khatri, P., Takamura, T., Shimizu, A., and Sugimoto, N.: Observation of low single scattering albedo of aerosols in the downwind of the East Asian desert and urban areas during the inflow of dust aerosols, J. Geophys. Res.: Atmos., 119, 787 – 802, https://doi.org/10.1002/2013JD019961, 2014.

Kumar, K. R., Yin, Y., Sivakumar, V., Kang, N., Yu, X., Diao, Y., Adesina, A. J., and Reddy, R.: Aerosol climatology and discrimination of aerosol types retrieved from MODIS, MISR and OMI over Durban (29.88°S, 31.02°E), South Africa, Atmos. Environ., 117, 9–18, https://doi.org/10.1016/j.atmosenv.2015.06.058, 2015.

Lan, T., Hu, H., Jiang, C., Yang, G., and Zhao, Z.: A comparative study of decision tree, random forest, and convolutional neural network for spread-F identification, Adv. Space Res., 65, 2052–2061, https://doi.org/10.1016/j.asr.2020.01.036, 2020.

Lee, J., Kim, J., Song, C., Kim, S., Chun, Y., Sohn, B., and Holben, B.: Characteristics of aerosol types from AERONET sunphotometer measurements, Atmos. Environ., 44, 3110–3117, https://doi.org/10.1016/j.atmosenv.2010.05.035, 2010.

Levy, R. C., Mattoo, S., Munchak, L. A., Remer, L. A., Sayer, A. M., Patadia, F., and Hsu, N. C.: The Collection 6 MODIS aerosol products over land and ocean, Atmos. Meas. Tech., 6, 2989–3034, https://doi.org/10.5194/amt-6-2989-2013, 2013.

Li, D., Cohen, J. B., Qin, K., Xue, Y., and Rao, L.: Absorbing Aerosol Optical Depth From OMI/TROPOMI Based on the GBRT Algorithm and AERONET Data in Asia, IEEE Trans. Geosci. Remote Sens., 61, 1–10, https://doi.org/10.1109/TGRS.2022.3231699, 2023.

Li, J., Hendricks, J., Righi, M., and Beer, C. G.: An aerosol classification scheme for global simulations using the K-means machine learning method, Geosci. Model Dev., 15, 509–533, https://doi.org/10.5194/gmd-15-509-2022, 2022.

Lin, Y., Tian, P., Tang, C., Pang, S., and Zhang, L.: Combining CALIPSO and AERONET Data to Classify Aerosols Globally, IEEE Trans. Geosci. Remote Sens., 60, 1–12, https://doi.org/10.1109/TGRS.2021.3138085, 2022.

Mallet, M. D., Cravigan, L. T., Milic, A., Alroe, J., Ristovski, Z. D., Ward, J., Keywood, M., Williams, L. R., Selleck, P., and Miljevic, B.: Composition, size and cloud condensation nuclei activity of biomass burning aerosol from northern Australian savannah fires, Atmos. Chem. Phys., 17, 3605–3617, https://doi.org/10.5194/acp-17-3605-2017, 2017.

Mamouri, R. E., Ansmann, A., Nisantzi, A., Kokkalis, P., Schwarz, A., and Hadjimitsis, D.: Low Arabian dust extinction-to-backscatter ratio, Geophys. Res. Lett., 40, 4762–4766, https://doi.org/10.1002/grl.50898, 2013.

Mao, Q., Huang, C., Chen, Q., Zhang, H., and Yuan, Y.: Satellite-based identification of aerosol particle species using a 2D-space aerosol classification model, Atmos. Environ., 219, 117 057, https://doi.org/10.1016/j.atmosenv.2019.117057, 2019.

Mateos, D., Toledano, C., Calle, A., Román, R., Herreras-Giralda, M., González, R., Herrero-Anta, S., González-Fernández, D., Herrero-del Barrio, C., Nisantzi, A., Mamouri, R. E., Groß, S., Cachorro, V. E., de Frutos, A. M., and Weinzierl, B.: Saharan and Arabian dust optical properties registered by sun photometry during A-LIFE field experiment in Cyprus, Atmos. Chem. Phys., 26, 1993–2005, https://doi.org/10.5194/acp-26-1993-2026, 2026.

Mortier, A., Goloub, P., Podvin, T., Deroo, C., Chaikovsky, A., Ajtai, N., Blarel, L., Tanre, D., and Derimian, Y.: Detection and characterization of volcanic ash plumes over Lille during the Eyjafjallajökull eruption, Atmos. Chem. Phys., 13, 3705–3720, https://doi.org/10.5194/acp-13-3705-2013, 2013.

Mukhopadhyay, S., Ningombam, S. S., Amoghavarsha, A., Madhavan, B., Eck, T. F., Dumka, U. C., Khatri, P., and Gupta, P.: Classification of global aerosol types and its radiative effects using Aerosol Robotic Network (AERONET) data, Atmos. Environ., 362, 121 530, https://doi.org/10.1016/j.atmosenv.2025.121530, 2025.

Müller, D., Ansmann, A., Mattis, I., Tesche, M., Wandinger, U., Althausen, D., and Pisani, G.: Aerosol-type-dependent lidar ratios observed with Raman lidar, J. Geophys. Res.: Atmos., 112, D16 202, https://doi.org/10.1029/2006JD008292, 2007.

Nie, X., Yu, L., Mao, Q., and Zhang, X.: Study on global atmospheric aerosol type identification from combined satellite and ground observations, Atmos. Environ., 347, 121 100, https://doi.org/10.1016/j.atmosenv.2025.121100, 2025.

Noh, Y., Müller, D., Lee, K., Kim, K., Lee, K., Shimizu, A., Sano, I., and Park, C. B.: Depolarization ratios retrieved by AERONET sun–sky radiometer data and comparison to depolarization ratios measured with lidar, Atmos. Chem. Phys., 17, 6271–6290, https://doi.org/10.5194/acp-17-6271-2017, 2017.

Noh, Y. M., Kim, Y. J., Choi, B. C., and Murayama, T.: Aerosol lidar ratio characteristics measured by a multi-wavelength Raman lidar system at Anmyeon Island, Korea, Atmos. Res., 86, 76–87, https://doi.org/10.1016/j.atmosres.2007.03.006, 2007.

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

Omar, A. H., Winker, D. M., Vaughan, M. A., Hu, Y., Trepte, C. R., Ferrare, R. A., Lee, K.-P., Hostetler, C. A., Kittaka, C., Rogers, R. R., Kuehn, R. E., and Liu, Z.: The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm, J ATMOS OCEAN TECH, 26, 1994 – 2014, https://doi.org/10.1175/2009JTECHA1231.1, 2009.

Pan, X., Uno, I., Wang, Z., Nishizawa, T., Sugimoto, N., Yamamoto, S., Kobayashi, H., Sun, Y., Fu, P., Tang, X., and Wang, Z.: Real-time observational evidence of changing Asian dust morphology with the mixing of heavy anthropogenic pollution, Sci. Rep., 7, 335, https://doi.org/10.1038/s41598-017-00444-w, 2017.

Papagiannopoulos, N., Mona, L., Amodeo, A., D'Amico, G., Gumà Claramunt, P., Pappalardo, G., Alados-Arboledas, L., Guerrero-Rascado, J. L., Amiridis, V., Kokkalis, P., Apituley, A., Baars, H., Schwarz, A., Wandinger, U., Binietoglou, I., Nicolae, D., Bortoli, D., Comerón, A., Rodríguez-Gómez, A., Sicard, M., Papayannis, A., and Wiegner, M.: An automatic observation-based aerosol typing method for EARLINET, Atmos. Chem. Phys., 18, 15 879–15 901, https://doi.org/10.5194/acp-18-15879-2018, 2018.

Rizza, U., Mancinelli, E., Morichetti, M., Passerini, G., and Virgili, S.: Aerosol Optical Depth of the Main Aerosol Species over Italian Cities Based on the NASA/MERRA-2 Model Reanalysis, Atmosphere, 10, 709, https://doi.org/10.3390/atmos10110709, 2019.

Roger, J.-C., Vermote, E., Skakun, S., Murphy, E., Dubovik, O., Kalecinski, N., Korgo, B., and Holben, B.: Aerosol models from the AERONET database: application to surface reflectance validation, Atmos. Meas. Tech, 15, 1123–1144, https://doi.org/10.5194/amt-15-1123-2022, 2022.

Russell, P. B., Bergstrom, R. W., Shinozuka, Y., Clarke, A. D., DeCarlo, P. F., Jimenez, J. L., Livingston, J. M., Redemann, J., Dubovik, O., and Strawa, A.: Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition, Atmos. Chem. Phys., 10, 1155–1169, https://doi.org/10.5194/acp-10-1155-2010, 2010.

Sabetghadam, S., Khoshsima, M., and Pierleoni, A.: Aerosol climatology and determination of different types over the semi-arid urban area of Tehran, Iran: Application of multi-platform remote sensing satellite data, Atmos. Pollut. Res., 11, 1625–1636, https://doi.org/10.1016/j.apr.2020.06.029, 2020.

Shekhar, J. S., S., M. K. C. V. S., and Rao, N. R.: Flexible atmospheric compensation technique (FACT): a 6S based atmospheric correction scheme for remote sensing data, Geocarto Int., 36, 28–46, https://doi.org/10.1080/10106049.2019.1588391, 2021.

Sinyuk, A., Holben, B. N., Eck, T. F., Giles, D. M., Slutsker, I., Korkin, S., Schafer, J. S., Smirnov, A., Sorokin, M., and Lyapustin, A.: The AERONET Version 3 aerosol retrieval algorithm, associated uncertainties and comparisons to Version 2, Atmos. Meas. Tech., 13, 3375–3411, https://doi.org/10.5194/amt-13-3375-2020, 2020.

Sreekanth, V.: On the classification and sub-classification of aerosol key types over south central peninsular India: MODIS–OMI algorithm, Sci. Total Environ., 468-469, 1086–1092, https://doi.org/10.1016/j.scitotenv.2013.09.038, 2014.

https://doi.org/10.5194/egusphere-2026-4776
Preprint. Discussion started: 21 September 2026

Tan, F., San Lim, H., Abdullah, K., Yoon, T. L., and Holben, B.: AERONET data–based determination of aerosol types, Atmos. Pollut. Res., 6, 682–695, https://doi.org/10.5094/APR.2015.077, 2015.

Taylor, M., Kazadzis, S., Amiridis, V., and Kahn, R.: Global aerosol mixtures and their multiyear and seasonal characteristics, Atmos. Environ., 116, 112–129, https://doi.org/10.1016/j.atmosenv.2015.06.029, 2015.

Wang, S., Wang, T., Jiao, Y., Dong, Y., Li, J., Bi, J., Huo, Y., Amonov, M. O., and Abdullaev, S. F.: Global aerosol models considering their spatial heterogeneities based on AERONET measurements, Atmos. Res., 308, 107 521, https://doi.org/10.1016/j.atmosres.2024.107521, 2024.

Wei, X., Cui, Q., Ma, L., Zhang, F., Li, W., and Liu, P.: Global aerosol-type classification using a new hybrid algorithm and Aerosol Robotic Network data, Atmos. Chem. Phys., 24, 5025–5045, https://doi.org/10.5194/acp-24-5025-2024, 2024.

Yang, Z. and Xu, C.: Development of bioaerosol lidar system, in: Fifth International Conference on Mechatronics and Computer Technology Engineering (MCTE 2022), vol. 12500, p. 71, SPIE, https://doi.org/10.1117/12.2660688, 2022.

Yang, Z., Xu, Q., and Cheng, C.: A method of aerosol types classification combining MODIS data and Mie-scattering lidar detection data, Opt. Commun., 600, 132 694, https://doi.org/10.1016/j.optcom.2025.132694, 2026.

Yao, N., Zhang, M., Bu, L., Gao, H., and Wang, Q.: Aerosol Classification Based on Airborne High Spectral Resolution Lidar, Acta Opt. Sin., 43, 2428 005, https://doi.org/10.3788/AOS230519, 2023.

Youroi, A. S., Borgohain, A., Kundu, A., Gautam, R., Kundu, S. S., Gohain Barua, A., and Aggarwal, S.: Aerosol characterization using dynamic columnar properties over a remote station in the Meghalaya plateau, Atmos. Pollut. Res., 17, 102 934, https://doi.org/10.1016/j.apr.2026.102934, 2026.

Yuan, C., Zhang, M., Wang, L., Ma, Y., and Gong, W.: Influence of aerosol on photosynthetically active radiation under haze conditions, J. Quant. Spectrosc. Radiat. Transfer, 311, 108 778, https://doi.org/10.1016/j.jqsrt.2023.108778, 2023.

Zhang, C., Wu, J., Fan, W., Zhao, J., Yang, Q., Zha, J., Zhao, D., and Sun, P.: A Method of Inversing Dynamic Aerosol Extinction-to-Backscattering Ratio Based on Lidar Echo Signal and Ground Aerosol Extinction Coefficient or Aerosol Optical Depth, IEEE Trans. Geosci. Remote Sens., 60, 1–15, https://doi.org/10.1109/TGRS.2022.3152902, 2022.

Zhang, L. and Li, J.: Variability of Major Aerosol Types in China Classified Using AERONET Measurements, Remote Sens., 11, 2334, https://doi.org/10.3390/rs11202334, 2019.

Zhou, M., Chang, J., Chen, S., Meng, Y., and Dai, T.: Aerosol Type Recognition Model Based on Naive Bayesian Classifier, Acta Opt. Sin., 42, 1801 006, https://doi.org/10.3788/AOS202242.1801006, 2022.