1 **Global aerosol typing classification using a new hybrid algorithm**

2 **utilizing Aerosol Robotic Network data**

3 Xiaoli Wei[1,2], Qian Cui[4,5], Leiming Ma[1], Feng Zhang [2,3], Wenwen Li[2,3], Peng Liu[4]

4 [1] *Shanghai Meteorological Service 200030, China;*

5 [2] *Shanghai Qi Zhi Institute, Shanghai, 200232, China;*

6 [3] *Department of Atmospheric and Oceanic Sciences & Institute of Atmospheric Sciences, Fudan*
7 *University, Shanghai, 200438, China;*

8 [4] *School of Atmospheric Science, Nanjing University of Information Science and Technology, Nanjing*
9 *210044, China;*

10 [5] *Wuhan Meteorological Service 430000, China*

11 *Correspondence to: Feng Zhang (fengzhang@fudan.edu.cn)*

12 **Abstract**

13 Aerosols have great uncertainty owing to the complex changes in their composition in
14 different regions. The radiation properties of different aerosol types differ
15 considerably and are vital in studying aerosol regional and/or global climate effects.
16 Traditional aerosol-type identification algorithms, generally based on cluster or
17 empirical analysis methods, are often inaccurate and time-consuming. Hence, we
18 aimed to develop a new aerosol-type classification model using an innovative hybrid
19 algorithm to improve the precision and efficiency of aerosol-type identification. An
20 optical database was built using Mie scattering and a complex refractive index was
21 used as a baseline to identify different aerosol types by applying a random forest
22 algorithm to train the aerosol optical parameters obtained from the Aerosol Robotic
23 Network sites. The consistency rates of the new model with the traditional Gaussian
24 density cluster method were 90%, 85%, 84%, 84%, and 100% for dust, mixed-coarse,
25 mixed-fine, urban/industrial, and biomass burning aerosols, respectively. The
26 corresponding precision of the new hybrid algorithm (F-score and accuracy scores)
27 was 95%, 89%, 91%, and 89%. Lastly, a global map of aerosol types was generated
28 using the new model to characterize aerosol types across the five continents. This
29 study utilizing a novel approach for the classification of aerosol will help improve the
30 accuracy of aerosol inversion and determine the sources of aerosol pollution.

1. Introduction

34   Atmospheric aerosols are tiny solid or liquid particles suspended in the

35 atmosphere. Aerosols indirectly affect the energy budget and water cycle of the earth's

36 gas system by absorbing and scattering solar radiation or by changing the optical

37 properties and life cycle of the cloud as condensation nuclei of cloud droplets

38 (Redemann et al. 2000; Ramanathan et al. 2001). Additionally, desert dust, biomass

39 smog, and anthropogenic emissions of air pollutants can affect visibility, air quality,

40 and human health (Hess et al., 1998;Siomos et al., 2020;Tong et al., 2017) Evaluating

41 the impact of aerosols on radiative transfer is complex, primarily because of the

42 uncertainty of radiative forcing caused by the high spatiotemporal dynamic variation

43 of aerosol optical and physical characteristics in different regions (Kaskaoutis et al.,

44 2011;Che et al., 2018).The aerosol type embodies the long-term average

45 physicochemical properties of aerosols in a certain area (Kiehl & Briegleb, 1993;Lu et

46 al., 2023). Therefore, accurate identification of aerosol types can drive the study of the

47 climatic effects of aerosols, tracking and control of environmental pollution sources,

48 and precision of radiation transmission models.

49   Aerosol types are defined based on the radiation properties of different types of

50 aerosol particles owing to the large variation in their optical, physical, and chemical

51 properties. Currently, aerosol types are classified by two ways using two different

52 clustering techniques (Kumar et al., 2018). First, based on different sources and

53 properties at different observation points worldwide, aerosols are classified as

54 follows: dust aerosols from deserts, biomass combustion aerosols from forests or

55 grasslands, and urban/industrial (U/I) aerosols from fuel combustion in densely

56 populated urban areas (Dubovik et al., 2002;Pawar et al., 2015;Yousefi et al., 2020).

57 Second, based on the size of the radiation absorption rate, aerosols into four

58 categories: carbonaceous (fine-absorbing mode), soil dust (coarse absorption mode),

59 sulfates (nonabsorbing fine-grained mode), and sea salt aerosols (nonabsorbing

60 coarse-grained mode) (Kim et al., 2007;Levy et al., 2007). The second one is a type of

61 subcategorize anthropogenic aerosol. The first one is commonly used for aerosol

62 retrieval. Therefore, the first aerosol type classification is more common in research.

63 The optical properties of aerosols observed at ground stations are commonly used to

64 construct a two-dimensional identification space to obtain the aerosol types by

65 clustering techniques. Many combinations of optical properties and parameters are

66 available; They include $EAE_{440-870nm}$ (extinction angstrom exponent) vs. SSA_{440nm}

67 (single-scattering albedo), $AAE_{440-870nm}$ (absorption angstrom exponent) vs. $EAE_{440-870nm}$

68 $_{870nm}$, $AAE_{440-870nm}$ vs. FMF_{550nm} (fine mode fraction), and SSA_{440nm} vs. $EAE_{440-870nm}$

69 (Lee et al., 2010;Shin et al., 2019;Choi, et al., 2021). Studies have highlighted the

70 importance of selecting appropriate aerosol properties for accurate aerosol type

71 identification (Giles et al., 2012; Che et al., 2018).

72   Among the aerosol-type classification methodologies developed, those using

73 threshold and empirical analyses have the greatest potential for large-area and fixed-

74 period applications (Eck et al., 1999; Omar et al., 2005; Yang et al., 2009).

75 Traditionally, the aerosol-type classification algorithm mainly distinguishes different

76 aerosol types based on their optical properties and determines the threshold of their

77 optical properties based on clustering. However, the composition of aerosols changes

78 rapidly with time and location, owing to the combined influence of natural conditions

79 and human activities (for example, tornadoes and various anthropogenic activities)

80 (Sheridan et al., 2001). Unfortunately, determining aerosol types accurately and

81 rapidly is a challenge when using traditional methods (Bahadur et al., 2012;Shin et al.,

82 2019;Lin et al., 2021). Nevertheless, with advancements in data science, artificial

83 intelligence techniques have aided the accurate and rapid recognition of different

84 aerosol types.

85   Artificial intelligence algorithms can receive multiple aerosol characteristic

86 parameters as input, thus preventing the sole reliance of aerosol classification on a

87 limited number of features (Li et al., 2022). For example, Boselli (2012) performed a

88 k-means clustering analysis of single scattering albedo (SSA), aerosol optical depth

89 (AOD), electrical asymmetry effect (EAE), and asymmetry parameter (g) datasets for

90 the central Mediterranean Sea for the classification of aerosol into four: dusty,

91 continental, oceanic, or mixed aerosols. Nicolae (2018) developed a neural network

92 algorithm to estimate the aerosol typing of Lidar data and Hamill (2016) introduced

93 the Mahalanobis Distance for aerosol classification to determine a specific aerosol

94 type for each reference cluster. Li (2022) generated spatial contiguous aerosol type

95 map in China with an empirical aerosol type retrieval algorithm.Overall, limited

96 information on the optical properties of aerosols can reasonably determine the type of

97 aerosol (Hamill et al., 2016). However, some challenges remain in identifying aerosol

98 types through machine learning. First, the amount of valid ground aerosol property

99 data that can be used for training is less due to cloud removal and quality control.

100 Second, the accuracy of machine learning depends on the labeled aerosol typing

101 dataset, and finding a suitable classification method to classify the dataset is

102 challenging. Third, evaluating the accuracy of the final trained model is also tedious

103 (Zhang & Li, 2019;Siomos et al., 2020; Choi, et al., 2021a,b)

104 The traditional aerosol type identification methods are easily limited by time and

105 space, and most of them only classify aerosol types using two optical property

106 parameters, limiting the complete characterization of aerosols. Considering these

107 limitations, we aimed to (1) develop a new algorithm that can accurately and quickly

108 identify aerosol types to overcome existing problems such as low accuracy,

109 insufficient data, and difficulty in setting labels; (2) investigate the characteristics of

110 the regional spatial distribution of global aerosol types obtained using the new

111 machine learning algorithms, considering the large regional differences in aerosol

112 types. To achieve this, we propose a new aerosol-type classification algorithm based

113 on a Gaussian cluster and random forest algorithm to generate an aerosol-typing map

114 over several representative regions of the world.

2. Study area and data

116 Figure 1 shows the study area and the Aerosol Robotic Network (AERONET) site

117 distribution, which covers major regions of the world, to ensure the generalizability of

118 the research algorithm. We used 47 aerosol sites as marked on the map that were

119 distributed over five continents to train and verify machine learning by literature

120 review. The 47 sites represent different aerosol-type properties of different aerosol

121 source regions, including dust, mixed (mixed coarse and mixed fine aerosols), U/I,

122 and biomass-burning (BB) aerosols (Table 1 and Figure 1). Marine aerosols were not

123 considered because their low optical thickness values (generally <0.4) can result in a

124 less valid data scale that would not meet the study requirements. Here, the aerosol

125 source region refers to the area affected by one dominant emission source, where the

126 aerosol types are fixed and not easily confused (Giles et al., 2012;Hamill et al., 2016).

127 Table 2 presents the optical properties and microphysical characteristic parameters of

128 aerosols at four bands of AERONET (440, 675, 870, and 1020 nm). These parameters

129 were used to construct a database of SSA, AOD, and asymmetry parameters. Further,

130 typical sites dominated by different aerosol types worldwide were selected for

131 compositional analysis using the new model. The selected sites are distributed across

132 different regions of the world and represent a specific aerosol-dominated type and

133 aerosol source region.

134   For dust aerosols, five AERONET sites, namely Banizoumbou, Cape Verde,

135 Dakar, and Ouagadougou in Africa and Solar Village in West Asia, influenced by the

136 Saharan Desert, were considered. The Dakar and Cape Verde sites are located at the

137 tip of the Cape Verde Peninsula—the westernmost part of Africa, bordering the

138 Atlantic Ocean. Although these two sites are located in the ocean, they are dominated

139 by dust aerosols influenced by aerosol plumes in the Saharan Desert. Moreover, the

140 Banizoumbou and Ouagadougou sites are in the middle of Africa. Here, the

141 northeasterly winds prevail in winter, and northwesterly winds prevail in summer,

142 which can bring dust aerosols from the Saharan Desert. For mixed aerosols, the

143 AERONET sites Ilorin, Kanpur, Sede Boker, and XiangHe were selected. For U/I

144 aerosols, the AERONET sites GSFC, Ispra, Mexico City, and Moldova were selected.

145 Four AERONET sites, namely, AltaFloresta, Abracos Hill, LakeArgyle, and Mongu,

146 were selected as BB aerosol-dominant sites

147

Figure 1. Study area and 47 AERONET sites selected by literature review.

149 **Table 1**. 47 AERONET sites selected by literature review.

Aerosol Type	Sites for Training	Sites for Testing
Dust	Agoufou,Capo_Verde,Dakar, Mezaira, Mussafa, Ouagadougou,	Banizoumbou,Solar_Village, Blida
Mixed	Anmyon,Beijing, Chen-Kung_Univ,Ilorin,Kanpur, Sede_Boker, Gosan_SUN, Pune, Taipei_CWB	Osaka, XiangHe, Pokhara,
Urban/Industry	Brookhaven,Billerica,Belsk,GSFC,Ispra,UMBC,Lille, Mexcio_City,Moldova,MD_Science_Center,Wallops	Athens-Noa,Shirahama, Leipzig
Biomass Burning	Abracos_Hill,Alta_Floresta,Cuiaba,Concepcion Los_Fieros,Mongu,Senanga, Skukuza,Zambezi	Bonanza_Creak, Etosha_Pan, Lake_Argle

150

151 **Table 2**. The optical and microphysical properties for aerosol type identification.

	Parameters	Variables (band waves)
Optical Properties	Ångström Exponent (AE)	EAE (440-870)[1]
	Aerosol Optical Depth (AOD)	AOD (440,675,870,1020)[1]
	Single Scattering Albedo (SSA)	SSA (440,675,870,1020)[1]
	Asymmetry Parameter	g (440,675,870,1020)[1]
	Imaginary Part of the Complex Refractive Index	REFI (440,675,870,1020)[1]
	Real Part of the Complex Refractive Index	REFR(440,675,870,1020)[1]
Microphysical Properties	Effective Radius	EffRad-F[2], EffRad-C[2]
	Standard Deviation of Effective Radius	StaDev-F[2], StaDev-C[2]
	Size Distribution	Vol-Con (0.05-15μm)

152 Note：[1] refers to wavelength in nm; [2] refers to different modes; EAE is Extinction Ångström Exponent; REFI is Imaginary Part of the

153 Complex Refractive Index; REFR is Real Part of the Complex Refractive Index; F refers to fine mode; C refers to coarse mode; EffRad is

154 Effective Radius; StaDev is standard deviation; Vol-Con is Volume concentration

3. Methods

A new aerosol classification typing hybrid approach that provides insight into spatiotemporal variations in aerosol pollution and climate impacts on a global scale is proposed in this study. In this approach, an aerosol optical properties database using the Mie scattering model was built for calculating rapidly unique aerosol-type features. Additionally, the approach introduced, for the first time, the median value of the complex refractive index (CRI) as the criterion for identifying the aerosol type. Further, we have selected the aerosol classification based on the source (as described in Section 1), according to the parameters applied in this study and the requirements for AOD retrieval. Figure 2 shows the working flowchart of the hybrid aerosol-type identification approach, including three stages: aerosol typing preliminary classification, aerosol optical database generation, and global aerosol typing identification and validation. The details of these three stages are as follows.
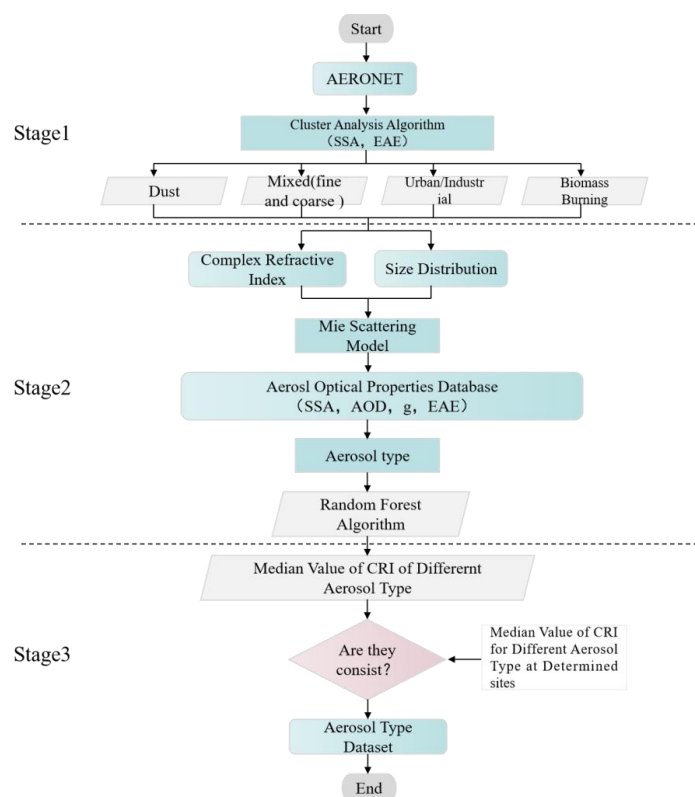
Figure 2. Flow chart of the new hybrid algorithm in aerosol type identification.

3.1 Aerosol typing preliminary classification (Stage 1)

Stage 1 aimed to solve the problem of obtaining a feature parameter dataset for the baseline aerosol type. In previous studies, the Gaussian kernel density clustering method showed great potential for distinguishing the optical properties of different aerosol types and determining their corresponding thresholds rapidly (Kalapureddy et al. 2009; Pathak et al. 2012). The high concentration value in each cluster generally represents the dominant pattern of a specific aerosol type, particularly the data within the window, taking the cluster centroid as the center and a specific distance as the radius. Preliminary aerosol-type datasets can be generated by digging deep into the distribution information of the effective radius, variance, and refractive index of the data within the window. The spectral absorbability and particle size of aerosols guide the identification of dust, carbonaceous, or hygroscopic aerosols; SSA indicates the absorption of aerosol particles; and EAE describes aerosol particle size (Giles et al., 2012). Consequently, in this study, SSA_{440nm} and $EAE_{440-870nm}$ of 47 AERONET sites and the Gaussian kernel density method were used to estimate the relative densities and determine the primary patterns of the dominant aerosol types; here, the aerosol type was classified as a dust aerosol. Eqs. (1) and (2) represent the kernel density and Gaussian kernel density methods (Rosenblatt, 1956)

$$f_{X(v)} = \frac{1}{L}\sum_{i=1}^{L} k_\sigma(\frac{\vec{x}-\vec{x}_i}{\sigma}) \ , \qquad (1)$$

where $f_{X(v)}$ denotes the kernel density and k_σ indicates the kernel function. x_1, x_2... x_L are the sample points of independent identical distribution. Mathematically, kernel functions are symmetric, normalized, and sample-centric when used for density estimation; this is best described by the Gaussian kernel equation given by Eq. (2).

$$k_\sigma = \frac{1}{\sqrt{2\pi}\sigma}\exp(\frac{-|\vec{x}-\vec{x}_i|^2}{2\sigma^2}), \qquad (2)$$

where σ is the kernel size used as a smoothing factor (Moraes et al., 2021).

The mixed aerosols comprised fine- and coarse-mode aerosols, indicated by $EAE > 0.8$ and $EAE \leqslant 0.8$, respectively. Figure 3 shows the clustering distribution of

8

197 EAE and SSA using the Gaussian kernel density method for different aerosol types at

198 the 47 AERONET sites. For the dust aerosol cluster, the density core area EAE was

199 0.1–0.3, and SSA was 0.89–0.94, implying that it contained many coarse aerosol

200 particles with moderate absorptivity. Furthermore, the mixed aerosols had two distinct

201 centers: one for the coarse-mode aerosols with a median EAE value of 0.4, indicating

202 that the cluster contained massive high-absorption aerosols, and the other for fine-

203 mode aerosols with a median EAE value of 1.3. Low-absorption aerosols were

204 dominant in the cluster, similar to U/I aerosols. Additionally, the density core region

205 EAE of U/I aerosol was 1.5–1.8, and SSA was 0.94–0.97, implying the dominance of

206 fine and low-absorption aerosols. Conversely, BB aerosols had two indistinct centers.

207 This is because, during biomass combustion, gas and particulate matter emissions are

208 limited by the combustion conditions, divided into combustion and simmering.

209 Combustion produces black smoke, and simmering produces white smoke.

210 Combustion, such as burning flames (grass) with high black carbon content, has a

211 strong absorption capacity, resulting in a low SSA. Simmering, such as burning wood

212 (i.e., trees), tends to be smoldering, lasts longer, has a weaker absorption capacity, and

213 has a higher SSA value. Therefore, despite possessing different absorption

214 characteristics, BB aerosols are defined as one aerosol type with an unseparated

215 center of combustion and simmering.
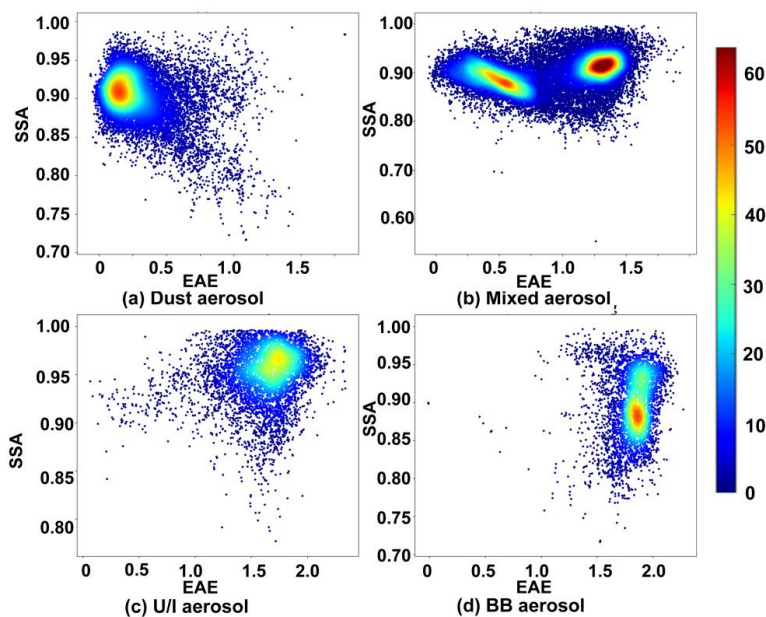
Figure 3. The clustering distribution of EAE and SSA using the Gaussian kernel density method for different aerosol types.

3.2 Aerosol optical database generation (Stage 2)

In stage 2, the aerosol optical parameter database was built using the aerosol size distribution parameters, CRI, and Mie scattering model. The main reasons for constructing an aerosol optical parameter database instead of using the AERONET data directly are as follows: 1) many data are missed in AERONET, particularly those for sites dominated by biomass combustion, which does not meet the requirements of machine learning methods or traditional aerosol type identification algorithms; 2) Calculating the optical properties of aerosols based on a fixed refractive index can accurately determine aerosol types. Therefore, once the aerosol spectral distribution parameters, such as effective radius, variance, and refractive index of the five aerosol types, are determined in stage 1, the aerosol optical parameter database can be constructed using the Mie scattering model in stage 2, assuming that aerosols are spherical particles. The Mie scattering model is a simple, practical, and ideal spherical particle model commonly used in radiation transport models (Michael et al., 1994). Figure 4 shows the details involved in the building aerosol optical database. The

10

234   aerosol optical database has four major parameters (AOD, EAE, SSA, and g) at four

235   wavelengths (440, 675, 870, and 1020 nm, respectively).
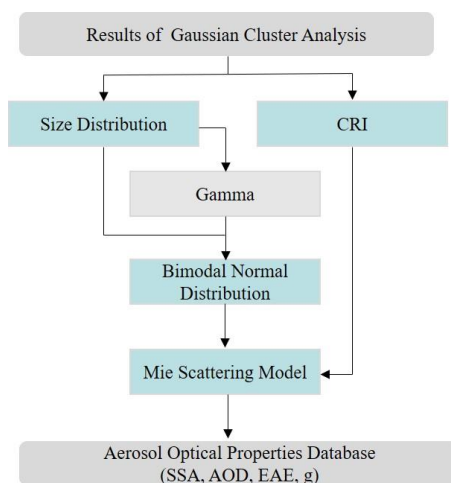
236

237   **Figure 4**. The diagram of building aerosol optical database.

238   As shown in Figure 4, size distribution is a major parameter in building aerosol

239   optical databases. Table 3 presents the aerosol size distribution parameters, including

240   the effective radius and standard deviation range for the five aerosol types in the

241   coarse and fine modes, which were calculated using the data in the window

242   determined by the Gaussian kernel density algorithm. These aerosol size distribution

243   parameters were used to build the aerosol optical database for the Mie scattering

244   model.

245   **Table 3**. Size distribution parameters of five aerosol types in coarse and fine mode (unit: μm)

Aerosol type	REFF-fine	REFF-coarse	Std-fine	Std-coarse
Dust	0.05-0.42	1.3-2.65	0.5-0.8	0.4-0.7
Mixed-coarse	0.05-0.25	1.25-3.5	0.4-0.8	0.4-0.7
Mixed-fine	0.05-0.27	1.2-4.5	0.3-0.6	0.5-0.8
U/I	0.05-0.26	1.45-3.5	0.3-0.6	0.5-0.8
BB	0.05-0.17	1.35-4.5	0.3-0.5	0.5-0.8

246   The Mie scattering model has various size distribution functions, including log-

247   normal, power-law, and bimodal log-normal distributions, which describe the aerosol

248   type. According to the particle radii provided by AERONET, the size distributions of

11

249 different aerosol types can be divided into coarse and fine modes. The bimodal log-

250 normal function [Eq. (3)] is reportedly the most suitable size distribution function for

251 modeling aerosol particle size distribution (Remer et al., 2009)：

252
$$n(r) = cons\tan t \times r^{-4}\{\exp(-\frac{(\ln r - \ln r_{g1})^2}{2\ln^2 \sigma_{g1}}) + \gamma \exp(-\frac{(\ln r - \ln r_{g2})^2}{2\ln^2 \sigma_{g2}})\} \ , \qquad (3)$$

253 where n(r) is the number of particles at different radii; constant is obtained by fitting;

254 While r_{g1} and r_{g2} denote the radii, σ_{g1} and σ_{g2} denote the variances of the aerosol in the

255 coarse and fine modes, respectively; and γ is determined by the volume distribution.

256 In the bimodal normal distribution model, γ is the ratio of coarse to fine modes, which

257 can be fitted by the volume distribution from AERONET; notably, volume distribution

258 is the average of the standard aerosols obtained after clustering at the training sites.

259   Figure 5 shows the volume distributions of the different aerosol types. The

260 aerosol volume distribution of dust aerosol-dominant sites focuses on the large radius;

261 the peak value of γ was 8.1, and the radius of dust aerosols was 1.5–2.0 μm.

262 Additionally, the mixed-coarse aerosol with the radius in the range of 0.04–0.2 μm

263 and 4.9 as the maximum value of γ. The mixed-fine aerosol had two obvious peaks:

264 one with a large radius, namely the coarse mode, with a radius of 2.2–3 μm and 2.1 as

265 the peak point of γ; a second with a small radius of 0.1–0.22 μm and 0.14 as the peak

266 point of γ. Moreover, the volume distributions of U/I and BB aerosols were similar.

267 Both had a relatively low range of γ values at large radii and relatively high values at

268 small radii, with peak values of 0.81 and 0.7 for U/I and BB aerosols, respectively.
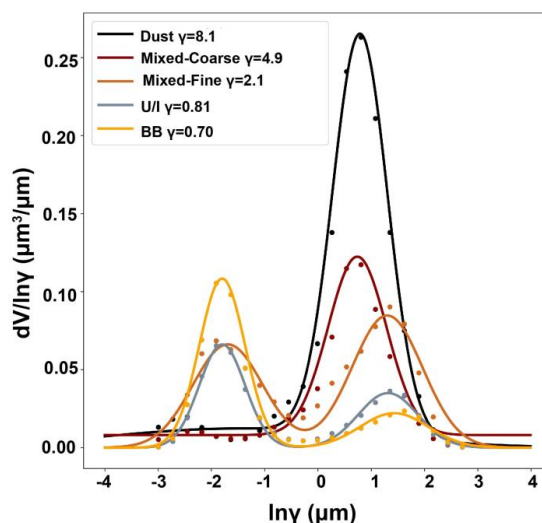
Figure 5. Volume distribution of the five aerosol types.

The CRI of aerosols is another key parameter among aerosol optical properties; it determines inherent optical properties of aerosols, such as scattering and absorption (Raut and Chazette, 2008). The CRI is vital for determining aerosols' chemical and physical compositions (Dubovik and King, 2000). Aerosols in the real atmosphere are usually mixed with different types of particles, which a single refractive index cannot identify; however, the CRI represents the entire aerosol model in the atmosphere (Redemann et al., 2000). Ideally, the CRI and aerosol components can be mutually determined (Wu et al., 2021). Table 4 depicts the CRI standard values for the five aerosol types obtained by calculating the median value of the CRI of the dominant aerosol type after Gaussian density clustering. These values were used as a baseline for identifying the aerosol types in subsequent studies. As presented in Table 4, the minimum imaginary index part is represented by the dust aerosol with CRI of 0.00374, 0.000847, 0.000847, and 0.000820 at 440, 675, 870, and 1020 nm, respectively, owing to the weakest absorption of dust aerosols. Moreover, the imaginary index part of the mixed-fine aerosols (0.01) was close to that of the U/I (0.07) and BB aerosols (0.02) because of their similar absorption properties.

288 **Table 4**. Real and imaginary index of CRI for the five aerosol types (Bands:440/675/870/1020
289 nm).

Aerosol Type	Imaginary Index	Real Index
Dust	0.00374/0.000847/0.000847/0.000820	1.4671/1.4778/1.4622/1.4504
Mixed-coarse	0.005349/0.002444/0.00204657/0.001845	1.4672/1.5088/1.5049/1.495
Mixed-fine	0.01449/0.01001/0.01009/0.009589	1.5075/1.5203/1.5243/1.516
U/I	0.007185/0.007166/0.007536/0.007552	1.4497/1.4397/1.4383/1.4346
BB	0.01961/0.01906/0.01903/0.01854	1.5133/1.5261/1.531/1.5282

290 Lastly, by fixing the CRI, changing the size distribution, and using the Mie

291 scattering model, we generated the aerosol optical property database for five aerosols,

292 including the data for AOD, EAE, SSA, and g. In the aerosol optical property

293 database, AOD is the value obtained after eliminating the influence of the aerosol

294 concentration. The AOD was obtained from the extinction cross section (C_{ext})

295 calculated using the Mie scattering model in Eqs. (3) and (4), where β_e is the

296 extinction coefficient, n(r) is the aerosol spectral distribution, and N(z) is the variation

297 of aerosol concentration with height. Notably, the effect of aerosol concentration

298 needs to be removed from the AOD when referring to aerosol optical properties. The

299 AOD was normalized by dividing the aerosol optical thickness at the four

300 wavelengths by the optical thickness at 440 nm. The other parameters (EAE, SSA,

301 and g) were calculated using Eqs. (6) – (8).

302
$$\beta_{e/s} = \int_{r\min}^{r\max} C_{ext/sca} n(r) dr \quad , \tag{4}$$

303
$$\tau_{e/s} = \int_0^{Z_{top}} \beta_{e/s} N(z) dz \quad , \tag{5}$$

304
$$EAE_{440-870nm} = -\frac{\ln(\tau_{440nm}) - \ln(\tau_{870nm})}{\ln(440) - \ln(870)} \quad , \tag{6}$$

305
$$SSA = \frac{\tau_s}{\tau_e} \quad , \tag{7}$$

306 and

307
$$g = <\cos\Theta> = \frac{1}{2} \int_{-1}^{1} p(\cos\Theta) \cos\Theta \, d\cos\Theta \quad , \tag{8}$$

308   where τ_{440} and τ_{870} are the extinction optical depths of the aerosol at 440 and 870 nm,

309   respectively, $EAE_{440\text{-}870}$ nm is the extinction Ångström index from the 440 to 870 nm

310   band, and Θ denotes the scattering angle.

311   The amount of data for the five aerosol types calculated using the Mie scattering

312   model is presented in Table 5. The least amount of data was observed for the mixed-

313   fine aerosol owing to its small distribution range of effective variance. The largest

314   data was observed for dust and mixed aerosols owing to their widely distributed

315   effective radii. A total of 326400 datasets were present in the aerosol optical database,

316   which meets the requirements for random forest algorithm.

317   **Table 5**. The data size of optical database simulated by Mie scattering model.

Total	Dust	Mixed-coarse	Mixed-fine	U/I	BB
326400	88200	96000	42000	51840	48360

318   **3.3 Global aerosol type identification and validation (Stage 3)**

319   In stage 3, the random forest model was introduced to the aerosol-type

320   identification algorithm. The random forest model is an integrated model based on

321   classification and regression trees, in which multiple trees are aggregated using

322   majority voting and averaging for classification and regression (Breiman, 2001). The

323   model has a high prediction accuracy, excellent tolerance for abnormal values and

324   noise, and a hard overfit. In a comparison by Fernandez (2014), the random forest

325   algorithm performed the best among 179 classification algorithms.

326   In this study, the input parameters for random forest model training, including

327   SSA_{440nm}, SSA_{675nm}, SSA_{870nm}, SSA_{1020nm}, g_{440nm}, g_{675nm}, g_{870nm}, g_{1020nm}, normalized

328   AOD_{440nm}, AOD_{675nm}, AOD_{870nm}, AOD_{1020nm}, and $EAE_{440\text{-}870nm}$, were selected from the

329   aerosol optical property database, and the expected output values were the specific

330   aerosol types. The random forest model was optimized and the parameters were

331   determined using the grid-searching method. The parameters, including n_estimators

332   (classifier), max_features (maximum feature value), and min_samples_leaf (minimum

333   number of samples for nodes), were set as 160, 10, 12, and 12, respectively. Then,

334   based on the trained and optimized model, aerosol typing of any AERONET site in

335 different regions of the world can be identified quickly. Generating the aerosol type

336 distribution map on a global scale is vital for regional and global climate studies and

337 ground remote sensing.

338 **4 Results**

339 **4.1 Algorithm comparison**

340 To demonstrate the effectiveness of the new hybrid algorithm, its performance

341 was compared with that of Gaussian density clustering algorithm. Figure 6 shows the

342 confusion matrix between the new hybrid and Gaussian density clustering algorithms

343 in identifying aerosol types. The results of the hybrid algorithm showed 90%

344 consistency with that from the Gaussian density clustering algorithm, in delineating

345 dusty aerosols, indicating that its efficiency in identifying dust. For mixed-coarse

346 aerosols, the consistency reached 85%, with 14% identified as mixed-fine aerosols,

347 1% as dust by the hybrid algorithm, and 15% as mixed-coarse aerosols by the

348 Gaussian density clustering algorithm. Similarly, for mixed-fine aerosols, both

349 algorithms showed 84% consistency, with 14% identified as a mixed-coarse aerosol

350 by the hybrid algorithm and as a mixed-fine aerosol by the Gaussian density cluster

351 algorithm. Furthermore, both algorithms identified 84% of U/I aerosols correctly, with

352 the remaining 16% identified as mixed aerosols (fine and coarse). Lastly, the

353 classification of BB aerosols using these two methods was the same. Overall, the

354 Gaussian density clustering and hybrid algorithms were consistent in dust, mixed-

355 coarse, U/I, and BB aerosol identification.
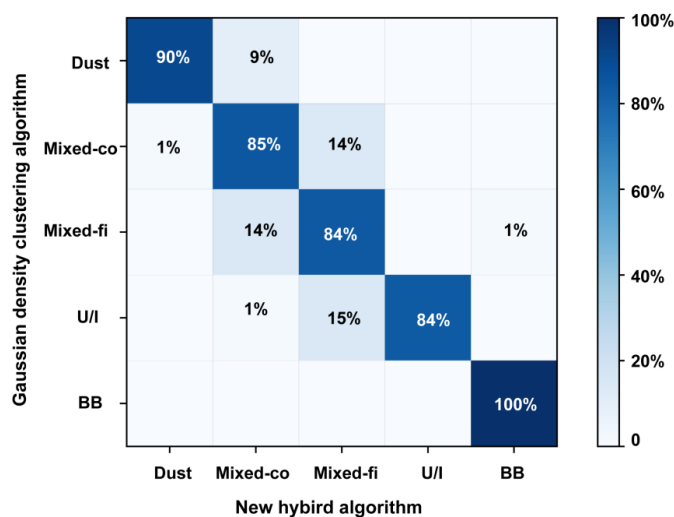
Figure 6. The confusion matrix between Gaussian density clustering and hybrid algorithm.

As described in the Methods section, a specific aerosol type theoretically has a fixed CRI owing to its constant composition. The CRI characterizes the mixture composition of aerosol particles and is a key parameter controlling the inherent scattering and absorption characteristics of aerosol particles. To further analyze the accuracy of the new algorithm, the aerosol CRI was applied as a key criterion for aerosol identification. The CRI has two parts: imaginary and real. The imaginary part indicates radiation absorption by aerosols, with a small value signifying a small absorption. Because the radiation of aerosols is more dependent on the imaginary than the real part, the imaginary part is essential for inferring the optical properties and aerosol types. Hence, we compared the real and imaginary parts of the CRI calculated using the hybrid and Gaussian density clustering algorithms.

Figure 7 shows box plots of the aerosol CRI for dust, mixed-coarse, mixed-fine, U/I, and BB aerosols using the hybrid classification and Gaussian density clustering algorithms. Based on the principle that the CRI of aerosols is fixed under ideal conditions, the closer the median value of the CRI of the identified aerosol type is to the median value of the benchmark CRI, the more accurate is the identification method.

375   As shown in Figures 7 (a) and (f), the median values of the CRI real part for dust

376   aerosol are in the range 1.45–1.53 at four bands, and those of the imaginary part are

377   0.003–0.004 at 440 nm; further, the values in other bands decrease rapidly as

378   wavelength increases. The imaginary part of CRI represents the absorption of light by

379   the aerosol, with a small absorption indicating strong scattering. The results of the

380   imaginary part are consistent with the spectral dependence properties of dust-based

381   aerosols according to the wavelength. This is primarily because dust aerosols,

382   composed of clay, quartz, and hematite, exhibit strong absorption in the blue band

383   (440 nm) and low absorption in the visible and near-infrared bands. For the dust

384   aerosols, the CRI determined by the two methods did not differ much. However, the

385   median value of the CRI obtained using the hybrid algorithm was slightly closer to the

386   benchmark CRI than that obtained using the Gaussian density clustering algorithm for

387   dust aerosols. Therefore, the hybrid algorithm was concluded to be more accurate in

388   identifying dust aerosol.

389   Figures 7 (b) and (g) show the median values of the CRI real part for mixed-

390   coarse aerosol is 1.47–1.55 at four bands using the new hybrid algorithm, but the

391   imaginary part is 0.004–0.009 at 440 nm. However, the real part is 1.44-1.50 at four

392   bands determined by Gaussian density clustering algorithm, and the imaginary part is

393   0.006–0.009 at 440nm. The median value of the hybrid algorithm was closer to the

394   baseline median value than that of the Gaussian density clustering algorithm for both

395   the real and imaginary parts.

396   Figures 7 (c) and (h) show the median value of the CRI real part for mixed-fine

397   aerosols determined using the new hybrid and Gaussian density clustering algorithms,

398   which was 1.42–1.51 at four bands. This result is close to the range (1.44–1.52)

399   reported by Wu (2021) in Beijing using a random forest algorithm. The median CRI

400   of the real part at four bands and imaginary part at the (675-870-1020 nm) bands were

401   close to the baseline median value for the new algorithm. Additionally, the median

402   value of the imaginary part was lower than that of the new hybrid algorithm and

403   further from baseline data for the identifying aerosol type results mixed with 14%

404   coarse aerosols. Mixed coarse aerosols result in weaker absorption. Hence, the new
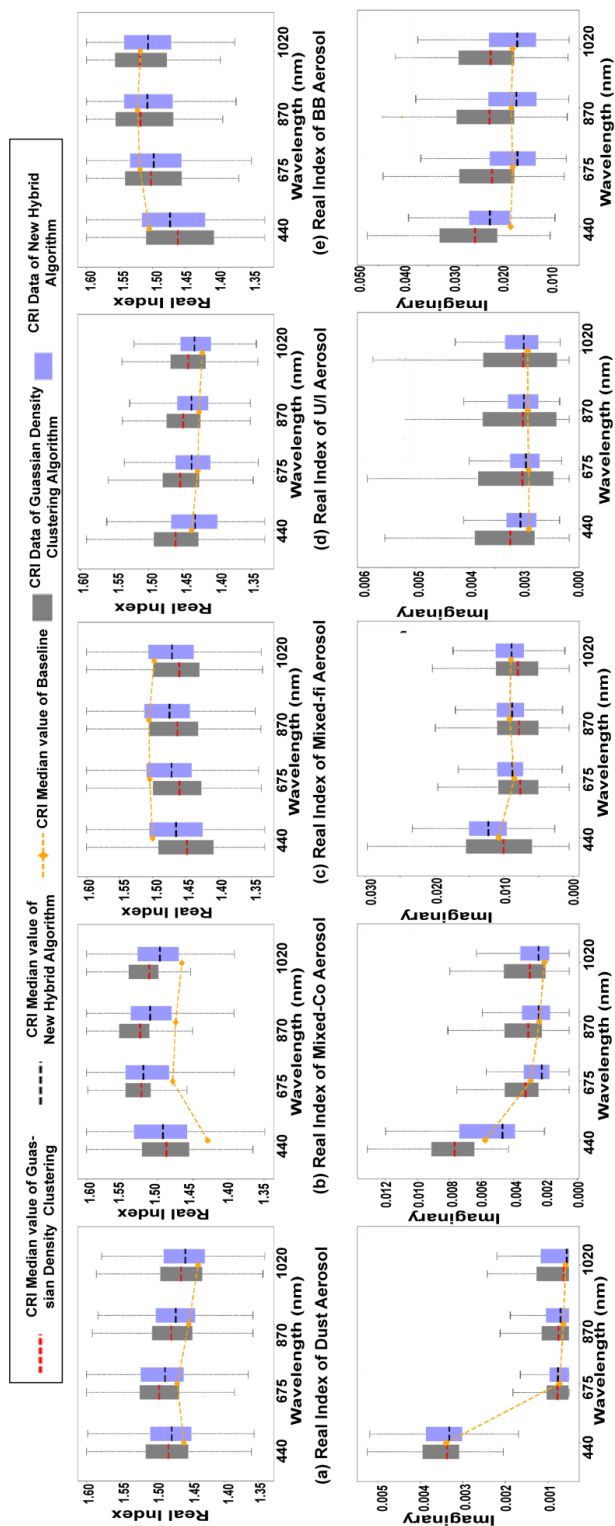
Figure 7. Box plots of the real index and the imaginary index of the CRI for (a) dust, (b)mixed-coarse, (c) mixed-fine aerosol, (d)U/I, and (e) BB aerosol identified by the Gaussian density clustering algorithm and new hybrid algorithm, respectively (the upper line is the real part, and the bottom line is the imaginary part).

19

409 hybrid algorithm performed better at identifying mixed-fine aerosols than the

410 Gaussian density clustering algorithm.

411   Similarly, as seen in Figures 7 (d) and (i), the median value of the CRI real part

412 for U/I aerosol identified using the new hybrid algorithm was 1.39–1.47. This median

413 value was lower than that of the mixed-fine aerosols. This is because the real part

414 indicates the absorption ability of aerosols, and the absorption ability of U/I aerosols

415 was less than that of mixed-fine aerosols. For the imaginary part also, the new hybrid

416 algorithm performed slightly better than the Gaussian density clustering algorithm at

417 the four bands.

418   For BB aerosols, the median value of the real part generated using the new hybrid

419 algorithm differed slightly from that generated by the Gaussian density clustering

420 algorithm. Additionally, the median obtained using the Gaussian density clustering

421 algorithm was closer to the baseline. Furthermore, when analyzing the imaginary part,

422 the new hybrid algorithm performed much better than the Gaussian density clustering

423 algorithm. Even with a 100% concordance rate between the new hybrid and Gaussian

424 density clustering algorithms in identifying BB aerosols, the refractive index still

425 differed. This result indicates that 1% of mixed-fine aerosols classified using the

426 Gaussian density clustering algorithm were correctly identified as BB aerosols by the

427 new algorithm. Overall, these results demonstrate that the new algorithm is reliable.

428 **4.2 Aerosol type determination for typical sites**

429 **4.2.1 Dust aerosol**

430   Figure 8 shows the aerosol types obtained using the new hybrid algorithm for the

431 five sites selected for dust aerosol identification. According to the prediction by the

432 new hybrid algorithm, the aerosols at these five sites mainly contained dust aerosols

433 along with a small amount of U/I, mixed-fine, and BB aerosols, and a large amount of

434 mixed coarse aerosols. This shows that other types of aerosols invaded these areas

435 besides dust aerosol. BB aerosols may have been transferred from the southern

436 African savannah. Additionally, U/I aerosols could be from industrial cities, such as

437 Dakar, Abidjan, and Lagos, which are dominated by anthropogenic aerosols and are

438   close to the AERONET sites.
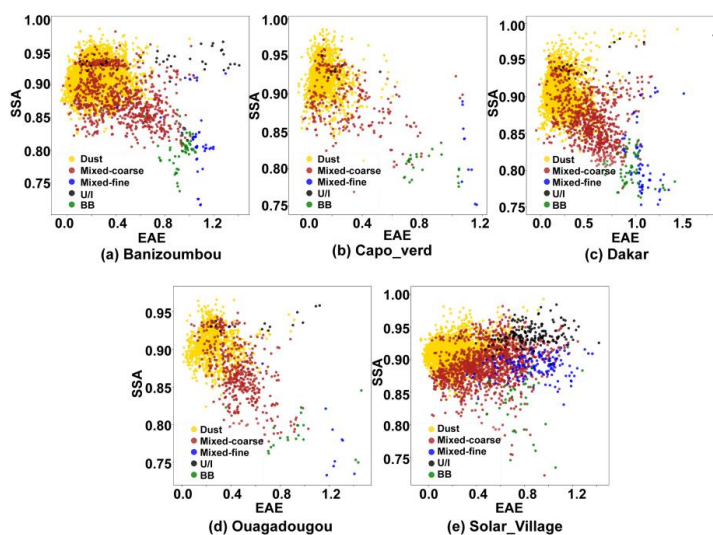
**Figure 8**. Identification of dust aerosol at dominant aerosol sites.

4.2.2 Mixed aerosol

442   Besides Ilorin in Africa, the mixed aerosol AERONET sites, including Kanpur,

443   Sede Boker, and XiangHe, are in Asia. The aerosol types at these four sites were

444   determined using the new hybrid algorithm (Figure 9). Mixed coarse aerosols

445   dominated the Kanpur, Ilorin, and Sede Boker sites, and mixed fine aerosols

446   dominated XiangHe. Part of the dust in Xianghe could be due to the Takla Desert in

447   spring and the westerly winds prevailing in western China, which transported dust

448   aerosols over long distances. Additionally, the U/I aerosol in Xianghe could be a result

449   of human activities, construction emissions, and fuel burning in winter. The BB

450   aerosol was traced to the burning of a small amount of biomass in Xianghe, located in

451   a suburban area.

452   Furthermore, excluding dust aerosols, we observed BB and U/I aerosols in the

453   Kanpur site in the Ganges Basin of India. A certain amount of U/I and dust aerosols

454   were also observed in Sede Boker, located in the industrial center of Israel, possibly

455   from the Arabian desert. Lastly, Ilorin had the most dust and least BB aerosols

456   because it is located in central Africa, often affected by the Saharan Desert and

457   African savannah.
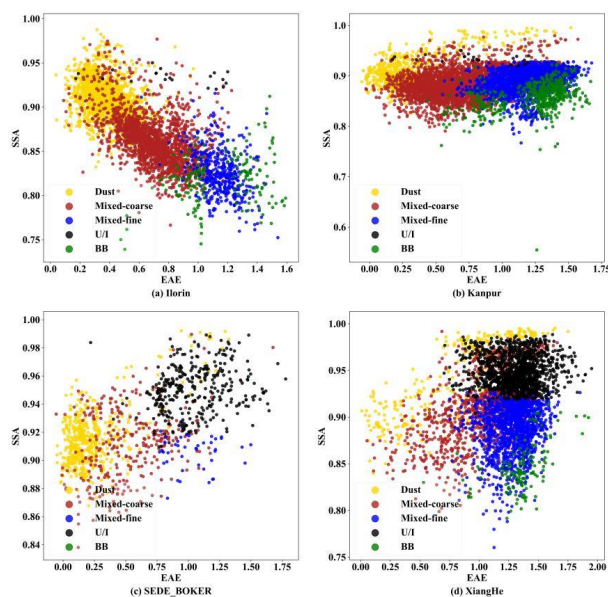
**Figure 9**. Same as Figure 8 but for Mixed aerosol.

4.2.3 Urban/industrial aerosol

461   All the selected AERONET sites for evaluating the performance of the new

462   hybrid algorithm in terms of U/I aerosol identification are in Europe or North America

463   (Figure 10). GSFC is located in the densely populated and industrially developed area

464   of Washington in the United States, explaining its complex aerosol type dominated by

465   the U/I aerosol followed by a few mixed and BB aerosols and a small amount of dust

466   aerosols.

467   Ispra is in Turin, one of Italy's largest industrial centers. However, dust-type

468   aerosols were identified, possibly transported from the Libyan desert when Italian

469   winters were controlled by southwesterly winds. Moreover, Mexico, where the

470   Mexico City site is located, is an industrialized country with modern industries and

471   agriculture, abundant oil production, and a dense population. Nevertheless, we

472   identified dust, mixed coarse, and BB aerosols in this site using the new hybrid

473   algorithm. These aerosol types could be from the Chihuahuan Desert, an inland desert

474   covering 12% of Mexico's area and a major source of coarse and dust aerosols.

475 Additionally, the literature shows that Mexico City is surrounded by forested
476 mountains, which experience many wildfires during the dry period between
477 November and May; this accounts for BB aerosols in Mexico City (Yokelson et al.
478 2007). Finally, the BB aerosols identified at the Moldova site could be attributed to its
479 rich vegetation cover.
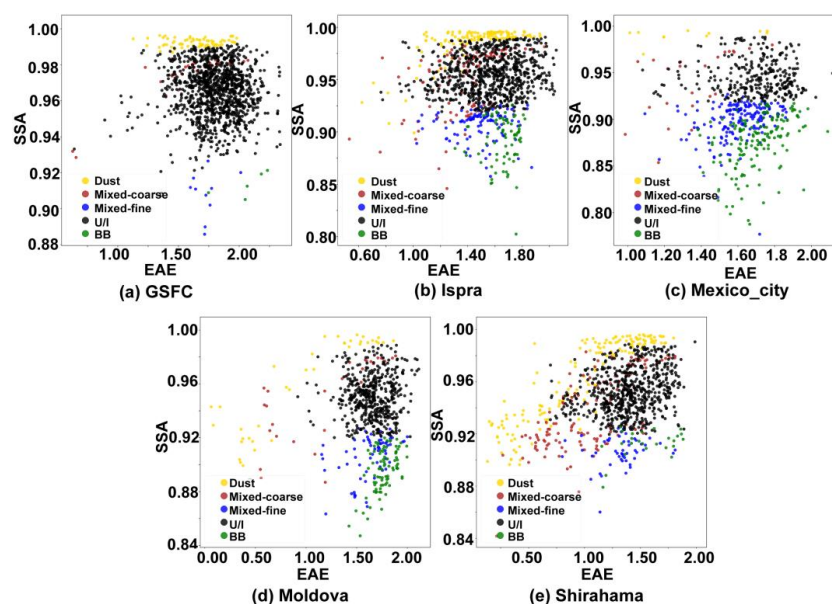
480
481 **Figure 10**. Same as Figure 9 but for urban/industrial aerosol.

4.2.4 Biomass burning aerosol

483   The selected sites were mainly located in the mountains and highlands. Figure 11
484 shows the aerosol types identified using the new hybrid algorithm. Large amounts of
485 BB aerosols were identified at all sites. Additionally, a small amount of dust and
486 mixed-coarse aerosols were identified at the Alta Floresta site, transported over a long
487 distance from the Patagonian Desert in Argentina, in southern South America.
488 Moreover, the city where the site is located is industrially developed and has a large
489 population; therefore, more U/I aerosols were identified using the new hybrid
490 algorithm. The geographically close Abracos Hill and Alta Floresta sites were
491 characterized by the same aerosol type and source. Furthermore, one data point in
492 Lake Argyle was classified as a dust aerosol. This means that, although the site is

23

493 located on the Kimberley Plateau, Australia has a large desert area, and coarse

494 aerosols still exist. Lastly, a few U/I and several dust-type aerosols were identified at

495 the Mongu site, possibly caused by aerosol emissions from nearby cities and dust

496 transport from the Saharan Desert.
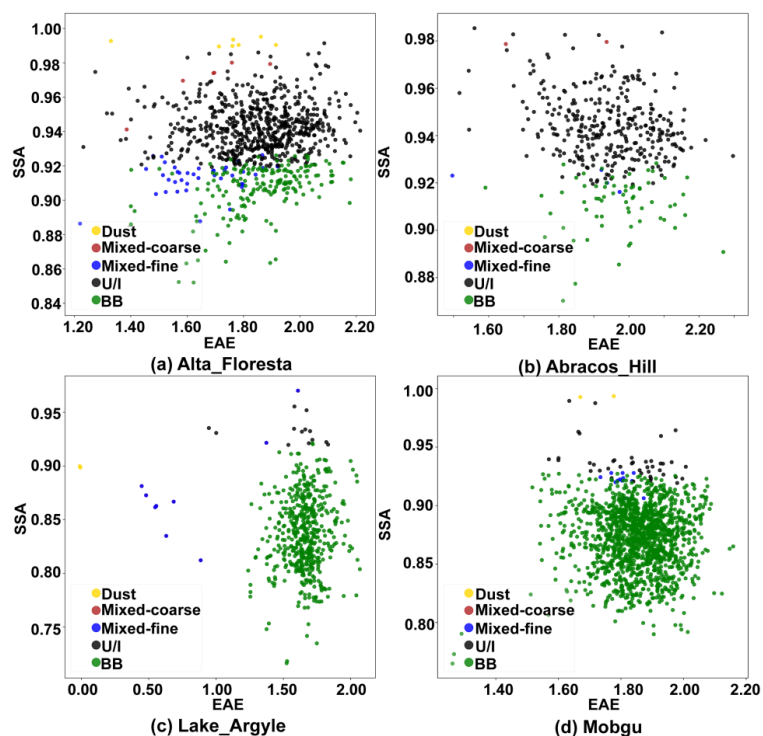
497

498 **Figure 11**. Same as Figure 10 but for BB aerosol.

4.3 Aerosol type distribution on a global scale

500   Given the advantages and accuracy of the new hybrid algorithm in identifying

501 aerosol types, we used it to divide the data of AERONET sites in different regions of

502 the world to obtain global aerosol type distribution information. The aerosol types of

503 each continent are shown in Figures 12-16. Additionally, Figure 17 shows the global

504 aerosol-type distribution. Notably, the pie chart was placed on each site in the study,

505 which is a "point source" assessment of the aerosol type and does not represent the

506 entire region (the size of the pie chart is independent of the optical properties).

507 Moreover, the sites were screened, and only those with valid data of > 100 aerosol

508 types were considered; however, offshore sites and sites classified as marine aerosol-

509 dominated by other literature were excluded.

510   Figure 12 shows pie charts of the aerosol types for each scanned AERONET site

511 in North America. The U/I aerosols, particularly in most mid-eastern regions,

512 contained mixed and small amounts of biomass aerosols. Additionally, the AERONET

513 sites in large cities, such as Chicago, New York, Toronto, Ottawa, and Boston, had U/I

514 aerosols. Many studies have shown that dust aerosols from the Saharan Desert can

515 cross the Atlantic Ocean to North America in summer. Moreover, there is an inland

516 desert in western North America, the Chihuahua Desert, responsible for a small

517 amount of dust and mixed aerosols at the AERONET sites in North America.

518 Additionally, wildfires in western North America and household wood burning

519 contribute to most BB aerosols yearly. The central region site is affected by the

520 environment, with an increased proportion of BB aerosols, and U/I aerosols are still

521 prevalent because the site is located in a large city and is densely populated.
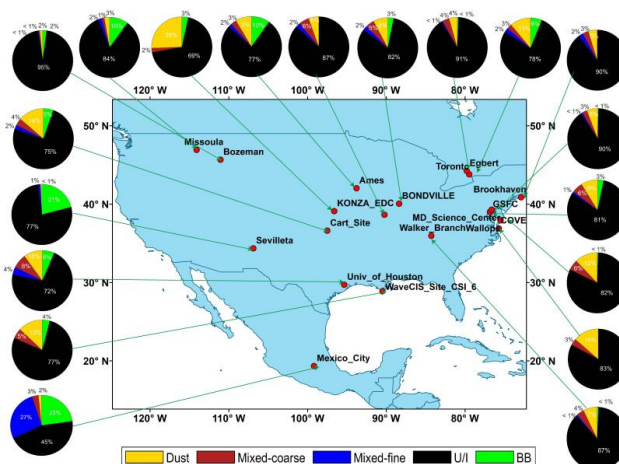
522

523 **Figure 12**. Pie charts of the aerosol types at the major sites of North American.

524   Figure 13 shows the aerosol types in Africa. Northern Africa has the largest desert

525 in the world, the Saharan Desert; therefore, dust aerosols dominate north of the

526 equator in Africa. However, some AERONET sites in the Sudanese steppe were

527 primarily BB, with some U/I aerosols in nearby urban sites. The Ilorin site is a typical

528   mixed aerosol site close to the equator with a small amount of BB aerosols. Most sites

529   close to the Atlantic coast were affected by dust aerosols, even those on the islands of

530   Cape Verde. The reliability of the new model in distinguishing U/I and BB aerosols is

531   demonstrated. Sites in Southern Africa, such as Namibia, Botswana, and Zambia, are

532   dominated by BB aerosols. Nevertheless, studies have shown the presence of U/I

533   aerosols at sites in the urban areas of South Africa. Although U/I and BB aerosols are

534   difficult to distinguish, the two can be identified in the context of a large urban

535   population and less biomass combustion, thus establishing the model's accuracy.
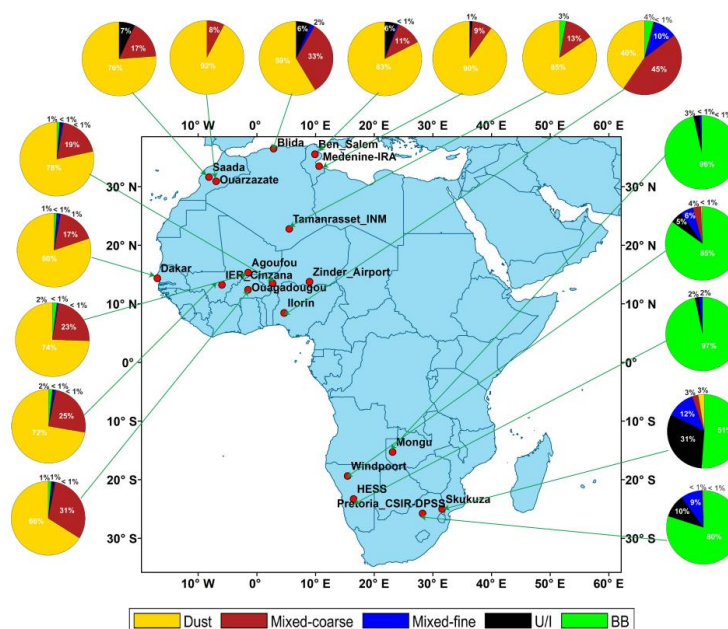
536

537   **Figure 13**. Same as Figure 12 but for Africa.

538   The aerosol types in South America are shown in Figure 14. Here, only eight sites

539   met the requirement for valid data >100 aerosol types. South America is mainly

540   dominated by mountainous plateaus, and under the influence of the Brazilian warm

541   current, many tropical rainforests are distributed in the south; therefore, the

542   background aerosols are mainly BB aerosols. As shown in Figure 14, large cities, such

543   as Rio Branco, Campo Grande, Manaus, Santa Cruz, and São Paulo, showed an

544   increased proportion of anthropogenic and mixed aerosols because of their large

545 population and developed industries. Due to the tropical rainforest climate in southern

546 South America, the proportion of BB aerosols increased, such as that at the Cuiaba

547 site near the Amazon River. Additionally, the Manaus site contained a small amount of

548 dust aerosols that were presumably transported across the Atlantic Ocean from

549 African dust at the same latitude.
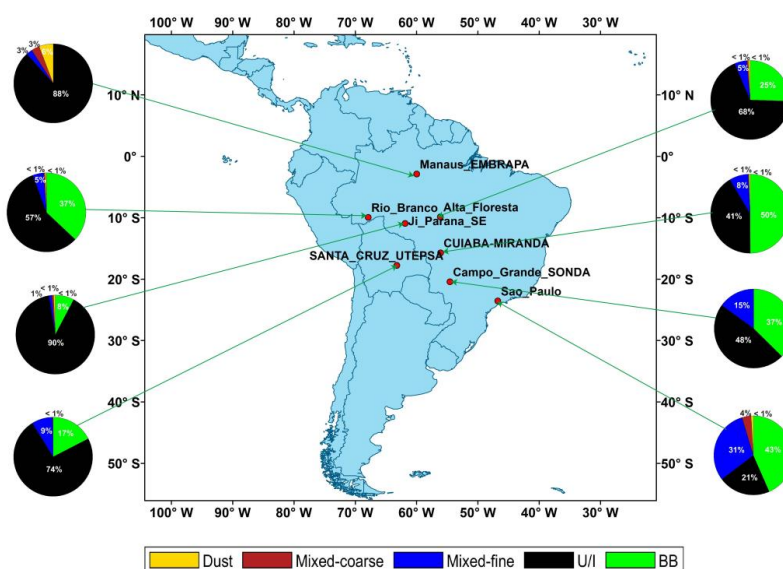
550

551 **Figure 14**. Same as Figure 12 but for South America.

552 The aerosol types in Asia are shown in Figure 15. In western Asia, influenced by

553 the Indian Desert, sites on the Indian Peninsula were dominated by coarse-particle

554 aerosols, including dust and mixed coarse aerosols. Kanpur and Pune are densely

555 populated cities in India, with more mixed-fine aerosols produced by human

556 activities. Additionally, in Southeast Asia, all sites contained BB aerosols, consistent

557 with Hamill (2014). This is because of the abundance of tropical rainforests in

558 Southeast Asia. Moreover, some urban sites, such as Singapore and Penang, had large

559 numbers of U/I and mixed-fine aerosols. The coastal areas of East Asia, which are

560 densely populated and industrially developed, were mainly dominated by U/I

561 aerosols. Moreover, dust aerosols appeared at these sites due to dust transported from

562 the Taklamakan Desert in East Asia.

27
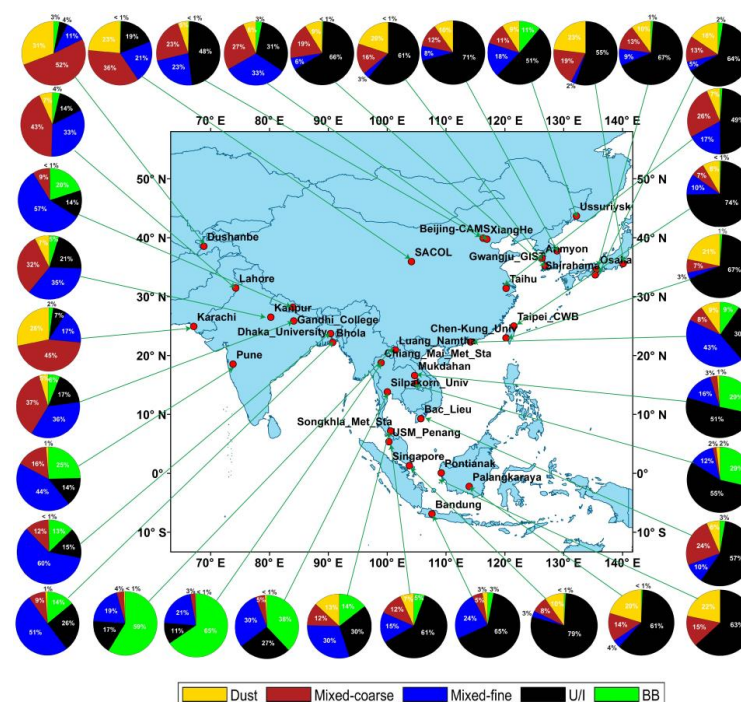
Figure 15. Same as Figure 12 but for Asia.

The inland areas of East Asia have a smaller population than the coastal areas; therefore, the proportion of U/I aerosols was small, and that of mixed aerosols was high. Generally, mixed aerosols are more easily overestimated than U/I aerosols; however, the new hybrid algorithm identified a larger proportion of U/I aerosols than mixed aerosols at Asian sites. Therefore, this new hybrid algorithm can be considered for improving the classification of mixed aerosols versus U/I aerosols.

Similarly, southern Europe, which is close to the Saharan and Arabian deserts, was dominated by dust aerosols, with small amounts of mixed and U/I aerosols. Northern European sites have many cities and a large population; therefore, the aerosol type was mainly U/I aerosols, identified using the new hybrid algorithm (Figure 16). Additionally, small amounts of BB aerosols were identified at most sites in Europe because of olive groves in agricultural lands in the EU, which produce 91% of the world's olive oil (Lopez-Pineiro et al., 2011). Papadakis et al. (2015) suggested that the biomass produced from olive oil is used for heating and industry, and its combustion produces carbonaceous aerosols, considered the major source of fine
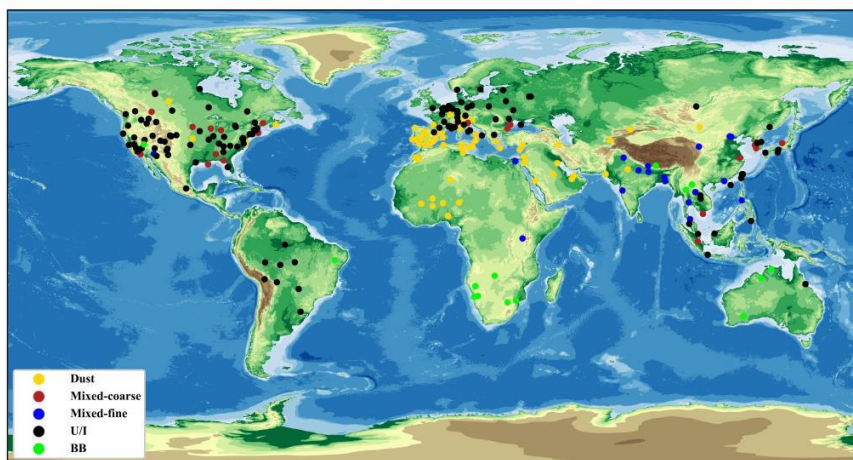
28

580   particle aerosols in Europe during winter (Puxbaum et al., 2007).
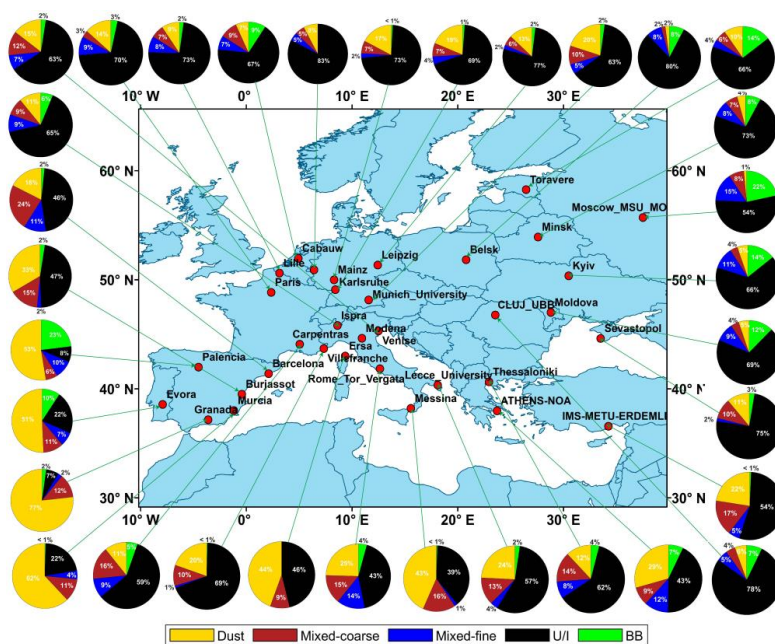
581

582   **Figure 16**. Same as Figure 12 but for Europe.

583

584   **Figure 17**. Global dominant aerosol types distribution based on AERONET sites.

585   The global distribution of dominant aerosols in the AERONET site is shown in

586   Figure 17. The graph does not include marine aerosols. There are more aerosol sites

587   on the global map than those on each continent because AERONET sites with > 5

588   years of data were selected for the global map; however, sites with > 100 valid data

589   points were required for each continent. The global distribution map shows that many

590   BB aerosols were distributed between 20°N and 20°S. This is because this region has

591   a predominantly tropical rainforest climate, with many tropical rainforests and more

592   carbon-containing aerosol emissions. This finding is consistent with those from

593   previous studies that found that global BB aerosols mainly originate from Africa

594   (approximately 52%), followed by South America (approximately 15%), equatorial

595   Asia (approximately 10%), boreal forests (approximately 9%), and Australia

596   (approximately 7%) (Van G. R. et al., 2010). Furthermore, the global distribution map

597   shows a clear distribution band of dust aerosols between 5°N and 35°N, originating

598   from the Saharan Desert in Africa and the Saudi Arabian Desert in Western Asia,

599   which are transported across the ocean to other regions.

5. Conclusion

601   We developed a new hybrid algorithm to support the rapid classification of

602   aerosol types by building an aerosol optical database for global AERONET sites. This

603   hybrid algorithm is a complex aerosol-type processing algorithm that effectively

604   integrates machine learning and density clustering algorithms. Additionally, this

605   algorithm is not limited by the amount of data and improves the accuracy of aerosol-

606   type classification. On investigating the aerosol types at specific sites with dominant

607   aerosols, we observed that different sites contained one or more aerosol types, with

608   the composition of some specific dominant aerosol sites being more complex than that

609   of others. The new algorithm showed a higher accuracy than that shown by algorithms

610   used in previous studies in identifying aerosol types at specific sites, particularly in

611   distinguishing between U/I and mixed-fine aerosols. Finally, the recognition results of

612   the new hybrid algorithm were closer to the baseline CRI, confirming that the hybrid

613   algorithm is better than the density-clustering algorithm. On investigating the aerosol

614   types at global sites across the continents using the new algorithm, we observed the

615   dominance of different types of aerosols at different sites, and the composition of

616 these could be logically and effectively attributed to the geographical location, energy
617 consumption structure, meteorological conditions and activities happening at the
618 respective sites.

619   In this study, the existing aerosol type identification algorithm was improved
620 using global ground-based AERONET optical property parameter data, and the spatial
621 distribution characteristics of global aerosol types were analyzed, which impacted
622 aerosol radiation research and optical thickness inversion accuracy. However, marine
623 aerosols were not considered in this study. This is a limitation of our study, and further
624 studies are required to include the optical properties of marine aerosols in model
625 building.

Author contributions

627 **Feng Zhang** designed the study. **Xiaoli Wei** analyzed the results, and wrote the
628 original draft. **Qian Cui** collected and processed the data. **Leiming Ma** revised the
629 paper and given constructive suggestions. **Wenwen Li** constructive comments on the
630 paper. **Peng Liu** revised the paper. All authors contributed to the study.

Competing interests

632   The authors declare that they have no conflict of interest.

Acknowledgments

References

638 Van G. R., der W., Randerson, J. T., Giglio, L., Collatz, G. J., Mu, M., Kasibhatla, P. S., Morton, D.
639   C., Defries, R. S., Jin, Y., and Van Leeuwen, T. T.: Global fire emissions and the contribution of
640   deforestation, savanna, forest, agricultural, and peat fires (1997–2009), Atmos. Chem. Phys., 10,
641   11707–11735, https://doi.org/10.5194/acp-10-11707-2010, 2010.
642 Bahadur, R., Praveen, P. S., Xu, Y., and Ramanathan, V.: Solar absorption by elemental and brown
643   carbon determined from spectral observations, Proc. Natl. Acad. Sci. U. S. A., 109, 17366–17371,
644   https://doi.org/10.1073/pnas.1205910109, 2012.

645 Boselli, A., Caggiano, R., Cornacchia, C., Madonna, F., Mona, L., Macchiato, M., Pappalardo, G., and
646   Trippetta, S.: Multi year sun-photometer measurements for aerosol characterization in a Central
647   Mediterranean site, Atmos. Res., 104–105, 98–110, https://doi.org/10.1016/j.atmosres.2011.08.002,
648   2012.

649 Breiman: Random forests, Machine Learning, 45(1), 5–32, https://doi.org/10.1023/A:1010933404324,
650   2001.

651 Che, H., Bing, Q., Zhao, H., Xia, X., and Zhang, X.: Aerosol optical properties and direct radiative
652   forcing based on measurements from the China Aerosol Remote Sensing Network (CARSNET) in
653   eastern China, Atmos. Chem. Phys., 18, 405–425, https://doi.org/10.5194/acp-18-405-2018, 2018.

654 Choi, W., Lee, H., and Park, J.: A first approach to aerosol classification using space-borne
655   measurement data: Machine learning-based algorithm and evaluation, Remote Sens., 13, 1–21,
656   https://doi.org/10.3390/rs13040609, 2021a.

657 Choi, W., Lee, H., Kim, D., and Kim, S.: Improving spatial coverage of satellite aerosol classification
658   using a random forest model, Remote Sens., 13 (7):1268. https://doi.org/10.3390/rs13071268,2021b.

659 Dubovik, O. and King, M. D.: A flexible inversion algorithm for retrieval of aerosol optical properties
660   from Sun and sky radiance measurements, J. Geophys. Res. Atmos., 105, 20673–20696,
661   https://doi.org/10.1029/2000JD900282, 2000.

662 Dubovik, O., Holben, B., Eck, T. F., Smirnov, A., Kaufman, Y. J., King, M. D., Tanré, D., and Slutsker,
663   I.: Variability of absorption and optical properties of key aerosol types observed in worldwide
664   locations, J. Atmos. Sci., 59, 590–608, https://doi.org/10.1175/1520-0469, 2002.

665 Eck, T. F., Holben, B. N., Reid, J. S., Dubovik, O., Smirnov, A., O'Neill, N. T., Slutsker, I., and Kinne,
666   S.: Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols,
667   J. Geophys. Res. Atmos., 104, 31333–31349, https://doi.org/10.1029/1999JD900923, 1999.

668 Fernandez-Delgado, M., Cernadas, E., Barro, S., and Amorim, D.: Do we Need Hundreds of Classifiers
669   to Solve Real World Classification Problems?, J. Mach. Learn. Res., 15, 3133–3181,
670   https://dl.acm.org/doi/10.5555/2627435.2697065, 2014.

671 Giles, D. M., Holben, B. N., Eck, T. F., Sinyuk, A., Smirnov, A., Slutsker, I., Dickerson, R. R.,
672   Thompson, A. M., and Schafer, J. S.: An analysis of AERONET aerosol absorption properties and
673   classifications representative of aerosol source regions, J. Geophys. Res. Atmos., 117, 1–16,
674   https://doi.org/10.1029/2012JD018127, 2012.

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

678 Hess, M., Koepke, P., and Schult, I.: Optical properties of Aerosols and Clouds: The Software Package
679   OPAC, Bull. Am. Meteorol. Soc., 79, 831–844, https://doi.org/10.1175/1520-
680   0477(1998)079<0831:OPOAAC>2.0.CO;2, 1998.

681 Kalapureddy, M. C. R., Kaskaoutis, D. G., Ernest Raj, P., Devara, P. C. S., Kambezidis, H. D.,
682   Kosmopoulos, P. G., and Nastos, P. T.: Identification of aerosol type over the Arabian Sea in the
683   premonsoon season during the Integrated Campaign for Aerosols, Gases and Radiation Budget
684   (ICARB), J. Geophys. Res. Atmos., 114, 1–12, https://doi.org/10.1029/2009JD011826, 2009.

685 Kaskaoutis, D. G., Kharol, S. K., Sinha, P. R., Singh, R. P., Badarinath, K., Mehdi, W., and Sharma, M.:
686   Contrasting aerosol trends over South Asia during the last decade based on MODIS observations,
687   Atmos. Meas. Tech. Discuss., 4, 5275–5323, https://doi.org/10.5194/amtd-4-5275-2011, 2011.

688 Kiehl, J. T. and Briegleb, B. P.: The relative roles of sulfate aerosols and greenhouse gases in climate

689   forcing, Science (80-.)., 260, 311–314, http://dx.doi.org/10.1126/science.260.5106.311, 1993.

690   Kim, J., Lee, J., Lee, H. C., Higurashi, A., Takemura, T., and Song, C. H.: Consistency of the aerosol

691   type classification from satellite remote sensing during the Atmospheric Brown Cloud-East Asia

692   Regional Experiment campaign, J. Geophys. Res. Atmos., 112, 1–12,

693   https://doi.org/10.1029/2006JD008201, 2007.

694   Kumar, K. R., Kang, N., and Yin, Y.: Classification of key aerosol types and their frequency

695   distributions based on satellite remote sensing data at an industrially polluted city in the Yangtze

696   River Delta, China, Int. J. Climatol., 38, 320–336, https://doi.org/10.1002/joc.5178, 2018.

697   Lee, J., Kim, J., Song, C. H., Kim, S. B., Chun, Y., Sohn, B. J., and Holben, B. N.: Characteristics of

698   aerosol types from AERONET sunphotometer measurements, Atmos. Environ., 44, 3110–3117,

699   https://doi.org/10.1016/j.atmosenv.2010.05.035, 2010.

700   Levy, R. C., Remer, L. A., Mattoo, S., Vermote, E. F., and Kaufman, Y. J.: Second-generation

701   operational algorithm: Retrieval of aerosol properties over land from inversion of Moderate

702   Resolution Imaging Spectroradiometer spectral reflectance, J. Geophys. Res. Atmos., 112,

703   https://doi.org/10.1029/2006JD007811, 2007.

704   Li, K., Bai, K., Ma, M., Guo, J., Li, Z., Wang, G., and Chang, N. Bin: Spatially gap free analysis of

705   aerosol type grids in China: First retrieval via satellite remote sensing and big data analytics, ISPRS

706   J. Photogramm. Remote Sens., 193, 45–59, https://doi.org/10.1016/j.isprsjprs.2022.09.001, 2022.

707   Lin, J., Zheng, Y., Shen, X., Xing, L., and Che, H.: Global aerosol classification based on aerosol

708   robotic network (Aeronet) and satellite observation, Remote Sens., 13, 1–23,

709   https://doi.org/10.3390/rs13061114, 2021.

710   Lopez-Pineiro, A., Cabrera, D., Albarran, A., and Pefia, D.: Influence of two-phase olive mill waste

711   application to soil on terbuthylazine behaviour and persistence under controlled and field conditions,

712   J. Soils Sediments, 11, 771–782, https://doi.org/10.1007/s11368-011-0362-3, 2011.

713   Lu, F., Chen, S., Hu, Z., Han, Z., Alam, K., Luo, H., Bi, H., Chen, J., and Guo, X.: Sensitivity and

714   uncertainties assessment in radiative forcing due to aerosol optical properties in diverse locations in

715   China, Sci. Total Environ., 860, 160447, https://doi.org/10.1016/j.scitotenv.2022.160447, 2023.

716   Michael, I., Mishchenko, and, Larry, D., and Travis: Light scattering by polydisperse, rotationally

717   symmetric nonspherical particles: Linear polarization, J. Quant. Spectrosc. Radiat. Transf.,

718   https://doi.org/10.1016/0022-4073(94)90130-9, 1994.

719   Moraes, C. P. A., Fantinato, D. G., and Neves, A.: Epanechnikov kernel for PDF estimation applied to

720   equalization and blind source separation, Signal Processing, 189, 108251,

721   https://doi.org/10.1016/j.sigpro.2021.108251, 2021.

722   Nicolae, D., Vasilescu, J., Talianu, C., Binietoglou, I., Nicolae, V., Andrei, S., and Antonescu, B.: A

723   neural network aerosol-typing algorithm based on lidar data, Atmos. Chem. Phys., 18, 14511–14537,

724   https://doi.org/10.5194/acp-18-14511-2018, 2018.

725   Omar, A. H., Won, J. G., Winker, D. M., Yoon, S. C., Dubovik, O., and McCormick, M. P.:

726   Development of global aerosol models using cluster analysis of Aerosol Robotic Network

727   (AERONET) measurements, J. Geophys. Res. D Atmos., 110, 1–14,

728   https://doi.org/10.1029/2004JD004874, 2005.

729   Pace, G., di Sarra, A., Meloni, D., Piacentino, S., and Chamard, P.: Aerosol optical properties at

730   Lampedusa (Central Mediterranean). 1. Influence of transport and identification of different aerosol

731   types, Atmos. Chem. Phys., 6, 697–713, https://doi.org/10.5194/acp-6-697-2006, 2006.

732   Papadakis, G. Z., Megaritis, A. G., and Pandis, S. N.: Effects of olive tree branches burning emissions

733   on   PM2.5   concentrations,   Atmos.   Environ.,   112,   148–158,
734   https://doi.org/10.1016/j.atmosenv.2015.04.014, 2015.

735   Pathak, B., Bhuyan, P. K., Gogoi, M., and Bhuyan, K.: Seasonal heterogeneity in aerosol types over
736   Dibrugarh-North-Eastern   India,   Atmos.   Environ.,   47,   307–315,
737   https://doi.org/10.1016/j.atmosenv.2011.10.061, 2012.

738   Pawar, G. V., Devara, P. C. S., and Aher, G. R.: Identification of aerosol types over an urban site based
739   on   air-mass   trajectory   classification,   Atmos.   Res.,   164–165,   142–155,
740   https://doi.org/10.1016/j.atmosres.2015.04.022, 2015.

741   Puxbaum, H., Caseiro, A., Sánchez-Ochoa, A., Kasper-Giebl, A., Claeys, M., Gelencsér, A., Legrand,
742   M., Preunkert, S., and Pio, C.: Levoglucosan levels at background sites in Europe for assessing the
743   impact of biomass combustion on the European aerosol background, J. Geophys. Res., 112, D23S05,
744   https://doi.org/10.1029/2006JD008114, 2007.

745   Ramanathan, V., Crutzen, P. J., Lelieveld, J., Mitra, A. P., Althausen, D., Anderson, J., Andreae, M. O.,
746   Cantrell, W., Cass, G. R., and Chung, C. E.: Indian Ocean Experiment: An integrated analysis of the
747   climate forcing and effects of the great Indo-Asian haze, J. Geophys. Res. Atmos., 106,
748   https://doi.org/10.1029/2001JD900133, 2001.

749   Raut, J. C. and Chazette, P.: Radiative budget in the presence of multi-layered aerosol structures in the
750   framework of AMMA SOP-0, Atmos. Chem. Phys., 8, 6839–6864, https://doi.org/10.5194/acp-8-
751   6839-2008, 2008.

752   Redemann, J., Turco, R. P., Liou, K. N., Russell, P. B., Bergstrom, R. W., Schmid, B., Hobbs, P. V,
753   Hartley, W. S., Ismail, S., and Ferrare, R. A.: Retrieving the vertical structure of the effective aerosol
754   complex index of refraction from a combination of aerosol in situ and remote sensing measurements
755   during TARFOX, J. Geophys. Res., 105(D8), 9949– 9970, doi:10.1029/1999JD901044,2000.

756   Remer, L. A., Tanré, D., and Kaufman, Y. J.: Algorithm for remote sensing of tropospheric aerosol from
757   MODIS: Collection 005, 2009.

758   Rosenblatt, M.: Remarks on Some Nonparametric Estimates of a Density Function, Remarks on Some
759   Nonparametric Estimates of a Density Function. In: Davis, R., Lii, KS., Politis, D. (eds) Selected
760   Works of Murray Rosenblatt. Selected Works in Probability and Statistics. Springer, New York, NY.
761   https://doi.org/10.1007/978-1-4419-8339-8_13, 2011.

762   Sheridan, P. J., Delene, D. J., and Ogren, J. A.: Four Years of Continuous Surface Aerosol
763   Measurements   from   the   DOE / ARM   Southern   Great   Plains   CART   Site,   1–8,
764   https://doi.org/10.1029/2001JD000785, 2001.

765   Shin, S. K., Tesche, M., Noh, Y., and Müller, D.: Aerosol-type classification based on AERONET
766   version 3 inversion products, Atmos. Meas. Tech., 12, 3789–3803, https://doi.org/10.5194/amt-12-
767   3789-2019, 2019.

768   Siomos, N., Fountoulakis, I., Natsis, A., Drosoglou, T., and Bais, A.: Automated aerosol classification
769   from spectral UV measurements using machine learning clustering, Remote Sens., 12, 1–18,
770   https://doi.org/10.3390/rs12060965, 2020.

771   Tanré, D., Kaufman, Y. J., Holben, B. N., Chatenet, B., Karnieli, A., Lavenu, F., Blarel, L., Dubovik,
772   O., Remer, L. A., and Smirnov, A.: Climatology of dust aerosol size distribution and optical
773   properties derived from remotely sensed data in the solar spectrum, J. Geophys. Res. Atmos., 106,
774   18205–18217, https://doi.org/10.1029/2000JD900663, 2001.

775   Tong, H., Lakey, P. S. J., Arangio, A. M., Socorro, J., Kampf, C. J., Berkemeier, T., Brune, W. H.,
776   Pöschl, U., and Shiraiwa, M.: Reactive oxygen species formed in aqueous mixtures of secondary

777   organic aerosols and mineral dust influencing cloud chemistry and public health in the
778   Anthropocene, Faraday Discuss., 200, 251–270, https://doi.org/10.1039/c7fd00023e, 2017.

779   Wu, Y., Li, J., Xia, Y., Deng, Z., Tao, J., Tian, P., Gao, Z., Xia, X., and Zhang, R.: Size-resolved
780   refractive index of scattering aerosols in urban Beijing: A seasonal comparison, Aerosol Sci.
781   Technol., 55, 1070–1083, https://doi.org/10.1080/02786826.2021.1924357, 2021.

782   Yang, M., Howell, S. G., Zhuang, J., and Huebert, B. J.: Attribution of aerosol light absorption to black
783   carbon, brown carbon, and dust in China - Interpretations of atmospheric measurements during
784   EAST-AIRE, Atmos. Chem. Phys., 9, 2035–2050, https://doi.org/10.5194/acp-9-2035-2009, 2009.

785   Yokelson, R. J., Urbanski, S. P., Atlas, E. L., Toohey, D. W., Alvarado, E. C., Crounse, J. D., Wennberg,
786   P. O., Fisher, M. E., Wold, C. E., and Campos, T. L.: Emissions from forest fires near Mexico City ,
787   Atmos. Chem. Phys., 7, 5569–5584, https://doi.org/10.5194/acp-7-5569-2007, 2007.

788   Yousefi, R., Wang, F., Ge, Q., and Shaheen, A.: Long-term aerosol optical depth trend over Iran and
789   identification of dominant aerosol types, Sci. Total Environ., 722,
790   https://doi.org/10.1016/j.scitotenv.2020.137906, 2020.

791   Zhang, L. and Li, J.: Variability of major aerosol types in China classified using AERONET
792   measurements, Remote Sens., 11, https://doi.org/10.3390/rs11202334, 2019.

793