

Aerosol classification by application of machine learning spectral clustering algorithm

Shantikumar S. Ningombam ^{a,*}, E.J.L. Larson ^b, G. Indira ^c, B.L. Madhavan ^d, Pradeep Khatri ^e

^a Indian Institute of Astrophysics, Bangalore, India

^b University of Colorado, Boulder, USA

^c 7, Middle Street, Keela Karisal Kulam, Sivagiri, Tenkasi, Tamil Nadu, India

^d National Atmospheric Research Laboratory (NARL), Gadank, India

^e Center for Atmospheric and Oceanic Studies (CAOS), Graduate School of Science, Tohoku University, Japan



ARTICLE INFO

Keywords:

Climate change
Earth's energy budget
Spectral clustering
Machine learning
AERONET

ABSTRACT

Precise understanding of aerosol classification is crucial for accurately quantifying the effects of aerosols on the Earth's energy budget, improving remote sensing retrieval algorithms, formulating climate change-related policies, and more. In this study, we used aerosol measurements from the quality assured AERosol Robotic NETwork (AERONET) and utilized a multivariate spectral clustering algorithm, a machine learning tool, to classify global aerosols. The spectral clustering algorithm is a variant of the clustering algorithm that employs eigenvalues and eigenvectors of the data matrix to project the data into a lower-dimensional space of a similar cluster. To accomplish this, we considered five aerosol optical parameters: fine-mode Aerosol Optical Depth, Extinction Angstrom Exponent, Absorption Angstrom Exponent, Single Scattering Albedo, and Refractive Index from 150 AERONET sites distributed in six continents (Africa, Asia, Australia, Europe, North and South America) during 1993 to 2022. Using the clustering analysis, we identified four primary aerosol types: dust, urban, biomass burning, and mixed aerosols. Among the continents, the African and Asian sites exhibited the highest contribution of dust aerosols, as the region has significant global dust sources. Conversely, Australia, Europe, North, and South America are predominantly influenced by fine-mode aerosols, given their considerable distance from major dust source regions.

1. Introduction

Aerosol radiative forcing is one of the key parameters in quantifying aerosol effects in Earth's energy budget perturbing, ground-atmosphere flux exchange, atmospheric thermodynamic stability, aerosol-cloud interaction, precipitation, and hydrological cycle (Cheng et al., 2008; Ding, 2013; Huang, 2016; Sun and Zhao, 2020, 2021). Classification of aerosol types (e.g., black carbon, urban/industrial, sea-salt, dust, etc.) is important for precise estimation of radiative forcing, since different aerosols have different radiative effects and atmospheric behavior (Bellouin et al., 2020). Further, due to their short residual time in the atmosphere, extensive information about the geographic and temporal aerosol distribution is one of the challenging tasks from the limited in situ aerosol measurement. It is also one of the most important and challenging tasks for satellite and sub-orbital remote sensing techniques to identify such different aerosol types (Russell et al., 2010; Lee et al., 2010; Levy et al., 2007). Additionally, fine and coarse mode aerosols

suspended in the atmosphere are difficult to be characterized both temporally and spatially due to their short residual time in the atmosphere and geographically diverse sources (Giles et al., 2012).

Many studies have been conducted to characterize aerosols on the threshold values of aerosol optical parameters, namely Extinction Angstrom Exponent (EAE), Single Scattering Albedo (SSA), and fine mode fraction (Ningombam et al., 2014a, 2015; Dumka et al., 2020; Lee et al., 2010). Since aerosols are highly heterogeneous in nature and influenced by local and synoptic meteorological conditions, such threshold classification schemes using two or three parameters may lead to improper results (Chen et al., 2019; Krishnaveni et al., 2023). Such analytical methods are sometimes unable to separate the different aerosol types when large data sets from multiple sites are combined together. For example, there is a similar threshold value of EAE for biomass burning and urban/industrial aerosols. Due to the close resemblance between urban/industrial and biomass burning aerosol types, threshold classification schemes do not work well. In

* Corresponding author.

E-mail address: shanti@iap.res.in (S.S. Ningombam).

such critical juncture, clustering techniques using multi-dimensional data are widely used when there is no prior information or supervised large data structure exist. Clustering method is an unsupervised machine learning tool to identify groups of similar dataset with two or more variable quantities. Such method relies on the underlying or hidden large unsupervised data structure to examine and extract a useful information from the unsupervised dataset. There are several clustering methods, such as, DBSCAN (Zhao et al., 2008), K-Means (Li et al., 2022; Jin and Han, 2011), spectral clustering (von Luxburg, 2007; Shi and Malik, 2000; Damle et al., 2019) and Mahalanobis distance (Hamill et al., 2016; Russell et al., 2014) clustering algorithms for aerosol classification. In recent years, spectral clustering method has become one of the most widely used algorithms due to better performance, less assumption, easier to implement and high processing speed than other traditional clustering algorithms, such as K-means algorithm. The main criteria of the algorithm is to cluster all spectrum of scattered data points into multiple groups based on their uniqueness without making any assumption. On the other hand, clustering techniques, like K-means, assume the data points assigned in spherical pattern. Such strong assumption of spherical shape may not always be relevant for large heterogeneity of data. The spectral clustering method attempts to connect the data of similar features, but may not necessarily have convex boundaries, unlike other conventional clustering algorithms. In such juncture, the spectral clustering algorithm helps to create more distinct clusters. The specific details of the spectral clustering algorithm and its application are well described by von Luxburg (2007). The more details of the algorithm are describing in Section 3.

In the current work, we used aerosol measurement from 150 AERONET sites across the globe in the North and South America, Europe, North and South Africa, Australia and Asia. The clustering algorithm used five variables of aerosol optical parameters from the inversion AERONET data product, such as AOD_Fine-mode, EAE, SSA, Absorption Angstrom Exponent (AAE), and Real Refractive Index (RRI) from the six continents. Since aerosols have spatial and temporal variations with short lifetime in the atmosphere, Land Used Land Cover (LULC) and aerosol optical parameters estimated from the inversion products are examined on temporal (yearly and seasonal) and spatial scales in the six continents before analyzing the clustering method. Such study will helps to understand the study region in details about the land classification and its aerosol climatology.

We organize this paper as follows: in Section 2, we describe about the study region and land classification for 150 AERONET sites. Section 3 describes about the AERONET data products, such as quality control, period of observations for each site along with the aerosol optical parameters used for aerosol classification. Further, methodology used for clustering techniques is also briefly described in this section. Section 4 describes about the spatial and temporal distribution of aerosol optical parameters along with the results obtained from the aerosol classification using spectral clustering and threshold methods. Finally, Section 5 describes about the summary, conclusions, and outlook of the current work.

2. Study region and land classification

The classification of aerosols performed using the 150 AERONET sites across the globe which spreads over North and South America, Europe, North and South Africa, Australia and Asia. The study also includes the LULC data (<https://www.esa-landcover-cci.org/>) produced by the European Space Agency (ESA) Climate Change Initiative (CCI), during 2015 full year data for a grid area of $2 \text{ km} \times 2 \text{ km}$ box for each AERONET site. LULC plays a vital role in the planning and managing of the urban environment, monitoring the changes in natural resources, helping in promoting and developing the sustainable growth (Feng et al., 2022). The methodology of LULC consists of four key process i.e., data acquisition, pre-processing, analysis/classification, product

generation and documentation. However, the extraction of LULC information from the remote sensing imagery is tricky for large urban regions (Tassi et al., 2021). LULC can be classified into homogeneous types, which consists of water/marine bodies, desert and vegetation from large areas and heterogeneous types, consisting of industrial and residential buildings, and roads. In the present work, classification of LULC is performed based on common land using associated types such as vegetation, marine, urban, desert and plantations regions. Vegetation land types include agricultural land, such as wetland, croplands, and forest lands. Marine land types are the regions consisting of streams, canals, sea, rivers, and lakes. Urban land types include the villages, towns, built up regions, residential, industrial, commercial and educational building structures. Plantation land types are comprises with groves, grasses, bushes, orchards, nurseries and horticulture plantations. Desert land types are categorized with different natural features in their geographic and climatic patterns, which are wrapped-up with little or no vegetation/plantation/water bodies. These desert regions are shaped with winds, unusual rainfall, bare exposed rock beds, salt marshes along with the desert pavement. Since plantation and vegetative land types have similar features, we combined the two land types as a single entity, i.e., vegetative land type in the current work.

Fig. 1 shows LULC types: vegetation, urban, marine, and desert land types for the 150 AERONET sites. Among the various land cover types, urban and vegetative land cover types are the two main types which contributed by 49% and 35%, respectively, while marine and desert contributed by 14% and 2%, respectively as seen in the **Table 1**. It is found that Europe, Asia, and South America have higher urban region than the vegetative land type. Europe has the highest urban sites with about 71% and followed by 16% and 13% by vegetation/crop/plantation and marine types, respectively. On the other hand, Australia has the highest contribution of marine type of land classification (46%) from the study of 11 selected sites. The Asian continent has the highest urban sites by 59% followed by 24%, 12%, and 5% for vegetation/crop/plantation, marine, and desert land types, respectively. Most of these selected sites in the Asian region are located in the south East Asian as seen in **Fig. 1**. Further, South America has about 48% of urban sites, which are closer with the vegetation/crop/plantation type (38%) from the total 21 AERONET sites. Further, this continent contributed 14% of marine type, which is a quite significant. The details of the LULC for 150 AERONET sites are shown in the Supplementary Table S1.

3. Data and methodology

The aerosol optical parameters used in the present study are taken from 39,411 daily observation at 150 AERONET sites in six continents (i.e., Africa, Australia, Asia, Europe, North America and South America) during July, 1993–March, 2022 as shown in **Table 1**. Further, the AERONET data used in the current work is Version 3.0 and Level 2.0 of inversion product, which is cloud-screened, quality controlled, and quality-assured products. The uncertainty in the AOD retrievals under clear sky condition is less than ± 0.01 for wavelengths greater than 440 nm and less than ± 0.02 at 440 nm wavelength. Further, the retrieval error in SSA is ± 0.03 with AOD greater than 0.4 (Dubovik and King, 2000; Holben et al., 1998). The aerosol optical products such as total, coarse and fine mode fractions of spectral AOD, AAOD, and SSA are retrieved through inversion scheme (Dubovik and King, 2000) using almucantar sky radiance measurements at 440, 675, 870, and 1020 nm. The EAE used in the current work is calculated from the spectral AOD values at wavelength range between 440 and 870 nm under the following relation:

$$EAE = -\frac{\ln(AOD_{870}) - \ln(AOD_{440})}{\ln(870) - \ln(440)} \quad (1)$$

While, AAOD(λ) at the respective wavelength is calculated using the following Equation:

$$AAOD(\lambda) = [1 - SSA(\lambda)] \times AOD(\lambda) \quad (2)$$

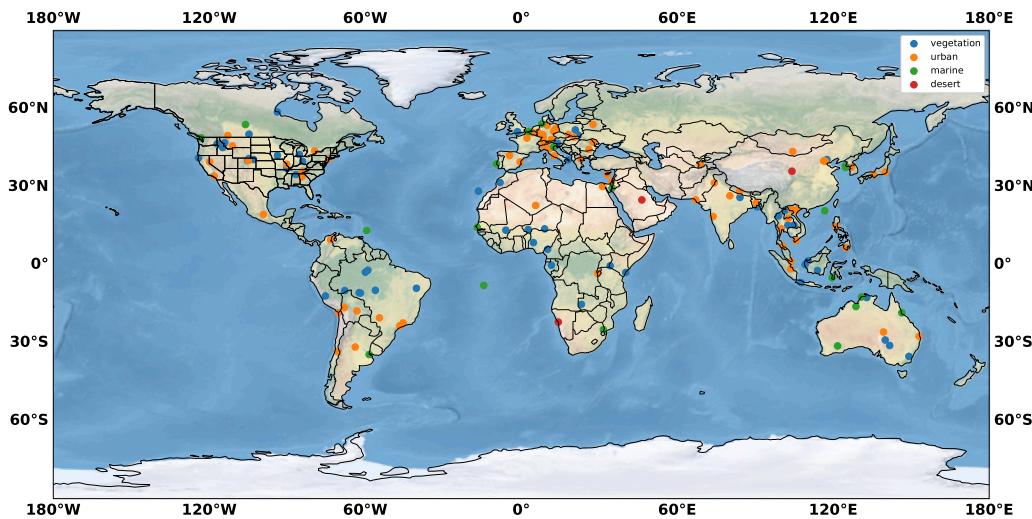


Fig. 1. Global map showing the 150 AERONET different sites, representing by solid circle in different colors for different land types, such as blue (vegetation), orange (urban), green (marine) and red (desert). For interpretation of color references, the reader is referred to the web version of this article.

Table 1

Classification of various land types (percentage) such as, vegetative (Veg.), urban, marine, and desert for six continents used in the current work. The Table also shows total no. of AERONET sites (150) with available daily observation and duration (period).

Continents	Total sites	Days	Duration	Land cover types (%)
1. Africa	17	10 086	10/1995–1/2022	Veg. (64), Urban (18), Marine (12), Desert (6)
2. Asia	41	19 922	2/1999–2/2022	Urban (59), Veg. (24), Marine (12), Desert (5)
3. Australia	11	469	10/2001–1/2020	Marine (46), Veg. (36), Urban (18)
4. Europe	32	4964	7/1997–3/2022	Urban (71), Veg. (16), Marine (13)
5. North America	28	1672	6/1995–10/2021	Veg. (54), Urban (39), Marine (7)
6. South America	21	2298	7/1993–1/2022	Urban (48), Veg. (38), Marine (14)
Total	150	39 411	7/1993–5/2022	Urban (49), Veg. (35), Marine (14), Desert (2)

Further, AAE is the slope of the AAOD curves as a function of wavelength and is calculated similar to EAE at wavelength range between 440 and 870 nm using the following Equation:

$$AAE = -\frac{\ln(AAOD_{870}) - \ln(AAOD_{440})}{\ln(870) - \ln(440)} \quad (3)$$

The real refractive index (RRI) at 440 nm used in the current work has the uncertainty 0.04 as reported by Dubovik and King (2000). The separation of biomass and urban/industrial aerosols is complicated since both aerosols have significant fine mode contribution. Generally, relatively strong and relatively light absorbing aerosols are considered as biomass and urban/industrial aerosols, respectively (Lin et al., 2021). Further, the uncertainties of the inverted aerosol optical parameters such as SSA, sphericity and refractive index increases as the AOD becomes less than 0.4 (Dubovik and King, 2000). Due to these constraints, the current work considered only four main aerosol types such as dust, biomass burning, urban/industrial and mixed aerosols. Aerosol from any location may be contaminated by intrusions of other types of aerosols. However, it is difficult to name such type of aerosols. Due to such complexities, such types of aerosols are referred to as ‘Mixed’ aerosols (Giles et al., 2012). The mixed aerosol types may be overestimated or overlapping due to the mixing of different aerosol types such as urban/industrial, biomass burning or polluted dust aerosols due to prevailing meteorological conditions and geographic location of the site. For example, aerosols in large cities appear to be mixed with anthropogenic aerosols (BC) and/or polluted dust which contributed to the mixed aerosols.

Regarding to the selection for proper number of clusters, we examined from cluster 3 to 4 based on the information available on topography and aerosol climatology data for each continent (see references in Tables 1 to 3). However, the classification did not find a significant improvement for separation of different aerosol types in 4 clusters as compared to 3 clusters (not shown here). The clusters vary

in different regions because of the dominance of the aerosol type in the region. Due to such constraints, we used only 3 clusters in the current work with the four aerosol types such as dust, biomass burning, urban/industrial, and mixed aerosol types. The basic principle of the spectral clustering method involves computing a similarity in graph, projecting the data into a low-dimensional space, and creating clusters. In the clustering process, the data points (unsupervised) are treated as nodes of a graph which combines similar features/parameters from the unsupervised dataset. The nodes are then mapped or graphed using eigenvalues and eigenvectors of the data matrix such as Affinity matrix, Degree of matrix and Laplacian matrix. The nodes are then mapped to a low-dimensional space that can be easily segregated to form clusters. Projecting the data into a low-dimensional space is performed to account for the possibility that members of the same cluster may be far away in the given dimensional space. We ran the model using the nearest neighbor affinity at 12 initializations processes. Further, the affinity matrix, also called a similarity matrix, is an essential statistical technique used to arrange or organize mutual similarities between the dataset. In the process, the 5 variables (EAE, AAE, SSA, RRI, and fine mode fraction of AOD) which have similar features form a cluster, regardless of their original range. Each cluster has a centroid value of the five variables. The different centroids in different clusters are indicated by different types of aerosol presence in the region. We use literature values of known aerosol properties (e.g. Russell et al. (2010), Giles et al. (2012) and Bibi et al. (2016)) to determine the likely composition of each aerosol cluster. However, elaborating all the theoretical and technical details in the current work may be out of the current work as the objective of the current work is to explain the results of global aerosol classification using spectral clustering method. Further, more technical details of the algorithm are described by von Luxburg (2007), Shi and Malik (2000) and Damle et al. (2019).

For selecting the suitable parameters, aerosol size and its absorptivity are the two main parameters used in the aerosol classification.

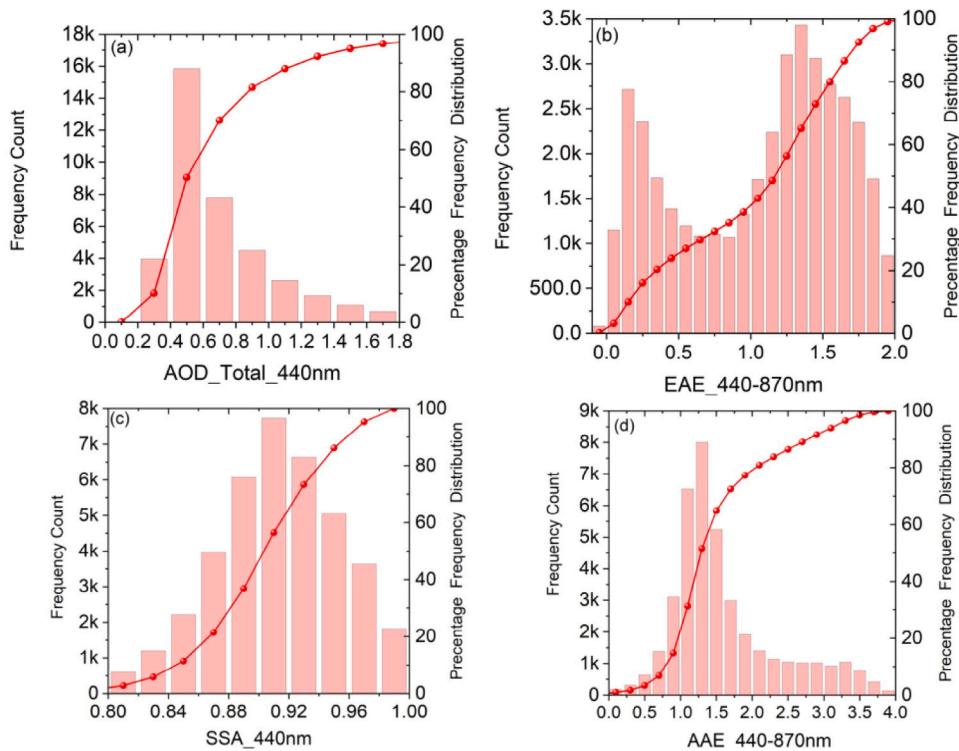


Fig. 2. Frequency distribution of AOD_Total at 440 nm, EAE_440-870, SSA at 440 nm, and AAE_440-870 obtained from the inversion scheme using 39,411 daily observation at 150 AERONET sites in six continent during 1993 to 2022 in the present work.

These two parameters are again associated with fine-mode AOD, EAE, AAE, SSA and RRI. We used the inversion products namely, fine-mode AOD, SSA and RRI at 440 nm and EAE_440-870 and AAE_440-870 nm to calculate the distance between points and identify the clusters. Since 440 nm (or 550/500 nm) is the most widely used wavelength in several aerosol classification (Hamill et al., 2016; Lee et al., 2010; Dumka et al., 2020), we have chosen this wavelength instead of other wavelengths. Further, it would be more suitable to consider this common wavelength to compare the current work with the rest of the work reported elsewhere. As a part of sensitivity analysis, we also carried out the clustering algorithm with the data at 675 nm in Africa and Europe as a few examples (not shown here). It is found that in most of the cases the centroids are quite closer with the clusters observed at 440 nm. However, using the classification at a longer wavelength, i.e., at 870 nm has fairly different centroids which suggests the spectral dependency on aerosol size parameters. Finally, changing of wavelength causes little changes in the aerosol classification.

We have also studied and compared the results of other clustering methods such as K-means and Dbscan. We observed that each method of aerosol classification has its limitation due to the nature of the methodologies adopted in it. For example, the Dbscan method is based on density-based spatial clustering of applications with more noise, unlike Spectral and K-means algorithms (as seen in the Supplementary Figure S3). On the other hand, the K-means data are more circular in shape and may not be able to capture more complex structures. The Spectral clustering algorithm is a variant of K-means clustering algorithm and that uses the eigenvectors of a similarity matrix to perform dimensionality reduction before applying K-means. Therefore, the K-means and Spectral methods have similar features in the aerosol classification (shown in the Supplementary Figure S3). However, the spectral clustering method is based on graph theory with a non-linear shape and can handle noisy data better than the K-means algorithm.

Classification of maritime aerosol is one of the most complex parts in the aerosol classification, because the AERONET inversions (Version 2) rarely yield inverted aerosol optical parameters of $AOD < 0.20$,

which we assumed for such aerosol types. Further, classification of marine aerosols are generally performed for low AOD, < 0.10 to the maximum limit of around 0.20 (Hamill et al., 2016; Yang et al., 2021). Due to such limitation, aerosol classification of marine origin can be performed in the current work as the inversion AERONET product used in the current work lies above the threshold value (~ 0.20) as shown in Fig. 2. The details of the spatial distribution of mean AOD_Total at 440 nm for 150 AERONET sites are shown in the Supplementary Figure S1.

4. Results and discussion

4.1. Global spatial distribution of aerosol optical parameters

Fig. 3 shows the spatial distribution of mean aerosol optical parameters such as coarse and fine mode AODs, AAOD, EAE, AAE, and SSA at 150 AERONET sites across the globe. The findings over different continents are listed as below:

The spatial distribution of aerosol optical properties in Africa is highly variable, due to the continent's diverse topography, climate, and land use. The African continent shows the lowest value of EAE (0.55 ± 0.51) and the highest value of AAE (2.23 ± 0.86), which could be due to the dominance of dust aerosols. Additionally, the continent shows a higher coarse mode AOD of 0.38 ± 0.29 at 440 nm. The study found that coarse mode aerosols, which are characterized by dust aerosols, are dominant in Northern Africa over the Sahara Desert and the Sahel region. Moreover, there are moderate contributions of fine mode fractions in central Africa, which could be due to biomass burning activities. Strongly absorbing aerosols over the Sahel region in West Africa reported in the earlier studies (Eck et al., 2001) are supported by the high AAE values (2.23 ± 0.86) in the current work. Among the continents, Northern Africa contributed the highest amount of dust aerosols, which are further transported across the eastern subtropical Atlantic to the Arabian Peninsula as well as southern Europe (Lin et al., 2021). The African continent has different types of aerosols due to the

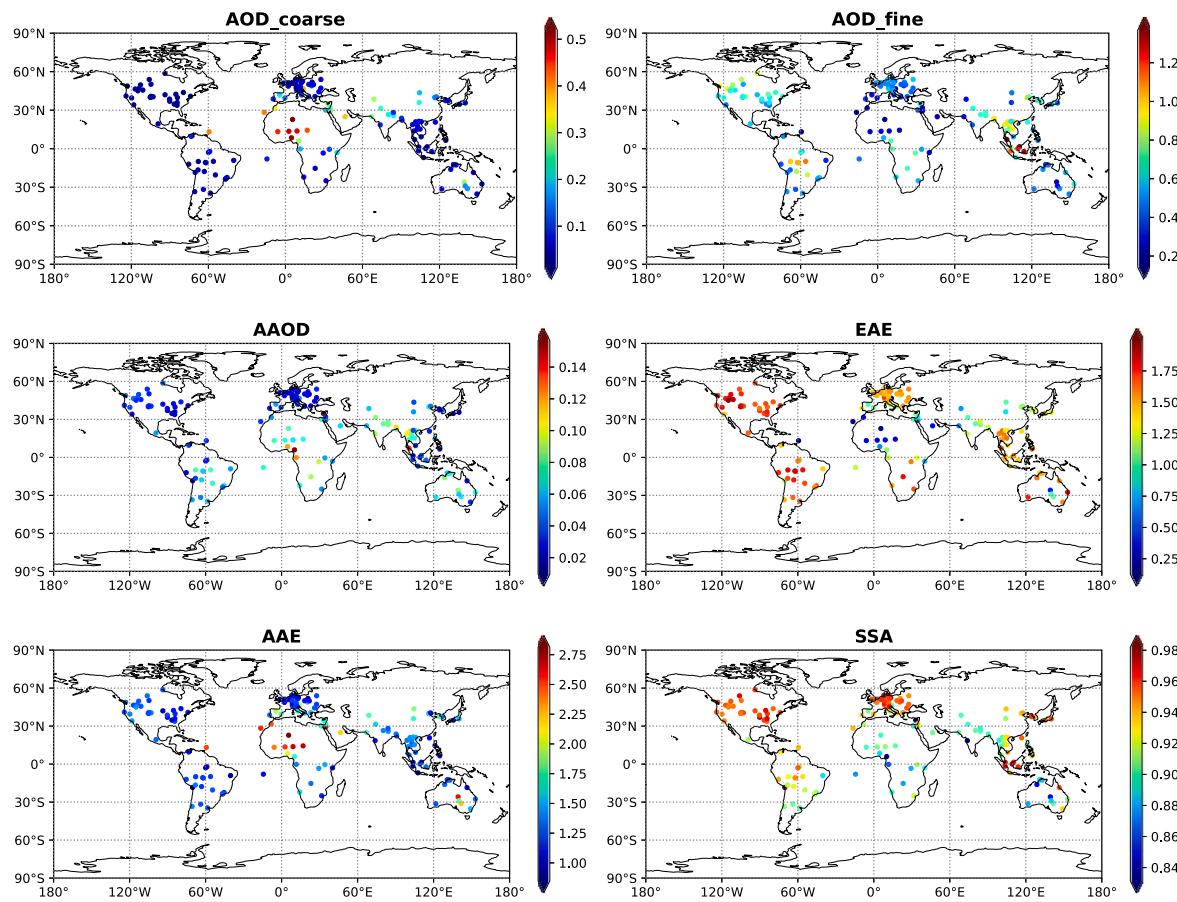


Fig. 3. Scatter plots of fine mode AODs, AAOD, EAE, AAE, SSA, and real refractive index. The color dot in each site implies the mean values in the color bar. For interpretation of color references, the reader is referred to the web version of this article.

wide range of climatological variability and topography (Peel et al., 2007).

The high EAE (>1.0) and moderate values of AAE (1.45 ± 0.58) may be attributed to the prominent biomass burning activities in the Asian region (Streets et al., 2003; Ningombam et al., 2020, 2023; Vadrevu et al., 2015). Further, the mixture of high EAE and low AAE is attributed to the presence of urban/industrial aerosols as there are significant number of sites located in the urban areas as seen in Table 1. The high quantity of absorptive aerosols is inversely related to SSA as seen in Fig. 3. The high contribution of coarse-mode aerosols over north Africa and Middle East (south Asia) are indicated by low EAE values as shown in Fig. 3. The most abundant aerosols in Asia are sulfates, which are produced by burning of fossil fuels. It is reported that anthropogenic emissions have exponentially increased due to the rapid industrialization and economic development over Asian polluted regions including the Indian subcontinent and eastern China during recent decades (Lu et al., 2011; Smith et al., 2011).

The global fire emission (fire count) data over Australia is about 8%–10% (Ningombam et al., 2023), demonstrating that Australia has a significant source of biomass burning aerosols. Due to different topography of the landmass and being surrounded by two oceans, different types of aerosols are observed in Australia, such as carbonaceous in the northern part, dust in the central part, and sulfate and sea salt over the densely populated north-western and south-eastern parts as well as coastal regions (Yang et al., 2021). Among the continents, Australia, Western Africa, and South Asia revealed minimum SSA value (<0.93) which indicated more absorbing aerosol that could be due to the origin of biomass burning aerosols.

The dominance of industrial and urban aerosol types in Europe is supported by high EAE (1.43 ± 0.40) and AAE (1.27 ± 0.64). The

contribution of aerosols emitted from industrial (fossil fuel burning) and urban pollution are falling under such threshold values of EAE and AAE (Russell et al., 2010; Bibi et al., 2016).

The average values of EAE and AAE in North America are 1.73 ± 0.23 and 1.22 ± 0.27 , respectively. The sampled data falls under the borderline between biomass burning and industrial/urban aerosols which may be attributed to the mixed types of aerosol from both cases. Such mixed type of aerosols is studied in detail under Section 4.2. The high SSA (>0.93) is observed mostly over western Europe, North America, South East Asia, and East Asia.

Similar to North America, land cover types in South America may suggest for mixing of two different types of aerosols such as industrial and biomass burning as observed from the estimated EAE (1.63 ± 0.38) and AAE (1.19 ± 0.43) values as shown in Fig. 3. On the other hand, South America has the highest contribution of fine-mode AOD (0.81 ± 0.63) among the continents as seen in the Table 2. Further, moderate values of fine-mode AOD are observed in Asia (0.65 ± 0.48), and North America (0.63 ± 0.39), followed by moderate values (<0.50) for the rest of the continents. The high values of fine-mode AOD are attributed from the biomass burning aerosols. The quantity of absorptive aerosols is determined from the AAOD values, while AAE measures the absorption capacity of aerosol in the atmosphere. South America, West Africa, South Asia, and Australia have high values of AAOD ($0.06 - 0.07$), indicating a large quantity of absorptive aerosols suspended in the air compared to other continents. Further, high values of AAE ($1.19 - 2.23$) are also noticed in these continents. On the other hand, high EAE values, which may be attributed to the dominance of fine-mode aerosols, are observed in North and South America, Western Europe, and South East Asia, and Australia. The high EAE values may be attributed either biomass burning or urban aerosols.

Table 2

Aerosol optical parameters, such as total, fine and coarse mode AODs, SSA, AAOD, and real refractive index at 440 nm, and EAE and AAE estimated at wavelength range between 440 and 870 nm using 440, 675, and 870 nm.

Aerosol Optical Parameters	Continents											
	Africa, n = 10 086		Asia, n = 19 922		Australia, n = 469		Europe, n = 4964		N.America, n = 1672		S. America, n = 2298	
	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
AOD_T	0.72 ± 0.35	0.59	0.80 ± 0.46	0.66	0.51 ± 0.15	0.47	0.51 ± 0.19	0.46	0.65 ± 0.39	0.53	0.86 ± 0.63	0.60
AOD_F	0.33 ± 0.23	0.25	0.65 ± 0.48	0.51	0.44 ± 0.16	0.41	0.45 ± 0.20	0.41	0.63 ± 0.39	0.51	0.81 ± 0.63	0.56
AOD_C	0.38 ± 0.29	0.32	0.15 ± 0.16	0.09	0.07 ± 0.06	0.05	0.06 ± 0.08	0.02	0.03 ± 0.02	0.02	0.05 ± 0.07	0.03
EAE	0.55 ± 0.51	0.33	1.11 ± 0.44	1.23	1.47 ± 0.35	1.56	1.43 ± 0.40	1.55	1.73 ± 0.23	1.75	1.63 ± 0.38	1.73
SSA	0.89 ± 0.04	0.89	0.90 ± 0.04	0.90	0.87 ± 0.05	0.87	0.95 ± 0.03	0.95	0.95 ± 0.03	0.95	0.93 ± 0.04	0.94
Ref_R	1.45 ± 0.06	1.45	1.47 ± 0.05	1.47	1.47 ± 0.07	1.47	1.44 ± 0.05	1.44	1.47 ± 0.06	1.48	1.47 ± 0.06	1.46
AAOD	0.07 ± 0.05	0.06	0.07 ± 0.04	0.06	0.07 ± 0.03	0.06	0.02 ± 0.01	0.02	0.04 ± 0.03	0.03	0.06 ± 0.04	0.05
AAE	2.23 ± 0.86	2.18	1.45 ± 0.58	1.36	1.31 ± 0.46	1.28	1.27 ± 0.64	1.15	1.22 ± 0.27	1.23	1.19 ± 0.43	1.23

4.2. Global temporal distribution of aerosol optical parameters

Table 3 show the temporal (seasonal) distribution of mean aerosol optical parameters, such as coarse and fine mode AODs, AAOD, EAE, AAE, and SSA from 39,411 days of observation at 150 AERONET sites across the globe. The details of the results in different six continents are listed as below:

Biomass burning in South Africa occur generally during August–October with a peak in September; maximum fire event (fire count) in North and central Africa occurs during December–November with a moderate value in June–July as revealed from the study of two decadal Moderate Resolution Imaging Spectroradiometer (MODIS) fire data (Ningombam et al., 2023). However, in Central and South Africa, burning starts in June, and the peak of the burning activity occurs in August with extension till November (Torres et al., 2009). The seasonal EAE is found to be the highest (0.80 ± 0.61) for the entire region during September–October–November (SON). Further, AAE is found to be the lowest (1.97 ± 0.85) during SON and highest (2.58 ± 0.79) in March–April–May (MAM). The high value of AAE and minimum value of EAE (0.31 ± 0.31) during MAM could be attributed to dust-laden atmosphere as seen in **Table 3**. Most of the major global dust source regions are located in the west coast of North Africa through the Middle East to Central Asia (Ginoux et al., 2010). For example, North Africa alone contributes about 55% of the global dust, while North Africa, the Middle East, and Asia together contribute to about 87% of the global dust emission (Ginoux et al., 2010).

The high dust-laden atmosphere in the Asian continent starts from MAM and extends till July–August as clearly observed from the seasonal highest coarse mode AOD (0.21 ± 0.22) with the least EAE (0.92 ± 0.50) during June–July–August (JJA). Such dust aerosols are found to be moderately absorbing aerosol in nature as SSA exhibits maximum (0.92 ± 0.04) during the season. The biomass burning activities in the continent varies as different types of crops cultivated in the region (Vadrevu et al., 2015). However, from the studies of two decadal fire emission data from MODIS, it reveals that maximum fire emission occurs during March–April for South Asia (Ningombam et al., 2023), although fire emission over the Indo-Gangetic plain the in India Peninsular, occurs at maximum peak in November–December (Ningombam et al., 2020). Seasonal fine-mode AOD varies from minimum 0.58 ± 0.50 in JJA to maximum 0.72 ± 0.41 in December–January–February (DJF) as seen in the **Table 3**. Since the continent has a maximum number of sites located in the Urban/Industrial region, there is also a contribution from urban aerosols which has similar features to biomass burning aerosols.

A relatively high portion of absorbing aerosols ($\text{SSA} \leq 0.86$) is observed in Australia during MAN, JJA, SON as seen in **Table 3**. Such absorbing aerosols (e.g., BC) are closely related to biomass burning emissions. Australia is also one of the main contributors of strongly absorbing aerosols in the Southern Hemisphere and the continent contributes about 8% of the global fire emission data (Ningombam et al., 2023). The continent shows a distinct seasonal cycle of absorbing aerosols as seen in the variation of SSA values on a seasonal time

scale. The highest seasonal fine-mode AOD is observed during SON as 0.43 ± 0.15 , which is probably due to the contribution from biomass burning aerosols, as the highest seasonal fire emission in Australia is observed during October (Ningombam et al., 2023). The continent does not reveal any strong dominance of dust aerosols from the observed values of low coarse mode AOD (0.02–0.10), and moderate values of EAE (1.27–1.76) and AAE (1.28–1.57) during the four seasons. Therefore, aerosol types present in Australia are mainly dominated by fine mode aerosols which may be attributed to biomass burning, urban or mixed types. However, Yang et al. (2021) reported for the presence of mixed aerosol types which are originated from the combination of biomass and dust aerosols in the region during all seasons. Further, they also observed dust-prone characteristics over Central Australia and biomass-burning-prone characteristics in Northern Australia. Such different aerosol types may be attributed to the different topographical regions used in the current work as aerosols are highly variable in nature (i.e., temporal and spatial) mainly due to different land types (topography) and different meteorological parameters (Ningombam et al., 2014b).

The present study found that the entire European region is dominated by weakly absorbing aerosols during the four seasons as SSA values varied from 0.94–0.95. On the other hand, EAE and AAE varied from 1.30–1.48 and 1.21–1.44, respectively during the different seasons. Further, the continent exhibits high fine mode (0.43–0.50) and low coarse mode (0.04–0.06) AODs during the seasons. These observed parameters reveal that the continent is dominated by fine-mode aerosols which may be attributed to biomass burning, urban or mixed types. Further, Logothetis et al. (2020) have rarely observed coarse particles in Central and East Europe with the exception of transported dust from the Saharan desert. Further, they also reported aerosol absorptivity in a narrow range of 0.94–0.96, while we observed 0.94–0.95 in the current work.

Similar to Australia and Europe, North and South America are dominated by fine mode aerosols in all seasons as observed from the high values EAE (1.25–1.76) and fine mode AOD (0.44–0.93), and low values of coarse mode AOD (0.02–0.07) as shown in **Table 3**. The estimated AAE is moderate (0.95–1.30) at both continents during the four different seasons. Such moderate values of AAE and high AAE suggest the dominance of fine mode aerosols. Aerosols available in North America during JJA and SON are found to be less absorbing types like Europe (except, DJF) as $\text{SSA} \geq 0.95$ as shown in the **Table 3**. Such high values of SSA during the season could be due to the dominance of urban aerosols in the continents. On the other hand, the absorbing aerosols ($\text{SSA} < 0.95$) with high EAE could be either biomass burning or mixed aerosol types. Further, there are high chances for presence of such aerosols in North America as $\text{SSA} \leq 0.95$. Seasonal SSA in South America is <0.95 which may suggest the presence of biomass or mixed aerosol types in the continent.

4.3. Aerosol classification: Spectral clustering method

One of the most challenging tasks in aerosol classification is the separation between biomass and urban aerosols since both the parameters

Table 3

Aerosol optical parameters such as fine and coarse mode AODs, SSA and AAOD at 440 nm along with EAE and AAE estimated at wavelength range between 440 and 870 nm using 440, 675, and 870 nm.

Continent	Season	AOD_Fine	AOD_Corse	EAE	SSA	AAOD	AAE
Africa	DJF	0.46 ± 0.27	0.39 ± 0.29	0.66 ± 0.37	0.86 ± 0.04	0.10 ± 0.05	2.01 ± 0.73
	MAM	0.25 ± 0.13	0.47 ± 0.30	0.31 ± 0.31	0.89 ± 0.03	0.07 ± 0.03	2.58 ± 0.79
	JJA	0.28 ± 0.20	0.36 ± 0.27	0.53 ± 0.58	0.90 ± 0.03	0.06 ± 0.03	2.28 ± 0.92
	SON	0.36 ± 0.24	0.26 ± 0.22	0.80 ± 0.61	0.90 ± 0.03	0.05 ± 0.03	1.97 ± 0.85
Asia	DJF	0.72 ± 0.41	0.07 ± 0.07	1.27 ± 0.29	0.89 ± 0.04	0.07 ± 0.03	1.29 ± 0.44
	MAM	0.61 ± 0.46	0.19 ± 0.18	1.04 ± 0.50	0.90 ± 0.03	0.07 ± 0.04	1.57 ± 0.61
	JJA	0.58 ± 0.50	0.21 ± 0.22	0.92 ± 0.50	0.92 ± 0.04	0.05 ± 0.03	1.53 ± 0.69
	SON	0.71 ± 0.52	0.11 ± 0.11	1.19 ± 0.35	0.91 ± 0.03	0.06 ± 0.03	1.38 ± 0.50
Australia	DJF	0.41 ± 0.19	0.10 ± 0.09	1.27 ± 0.52	0.91 ± 0.04	0.04 ± 0.02	1.57 ± 0.85
	MAM	0.40 ± 0.16	0.03 ± 0.01	1.57 ± 0.34	0.86 ± 0.08	0.07 ± 0.05	1.35 ± 0.43
	JJA	0.40 ± 0.17	0.02 ± 0.01	1.76 ± 0.22	0.85 ± 0.03	0.07 ± 0.02	1.28 ± 0.15
	SON	0.43 ± 0.15	0.06 ± 0.05	1.46 ± 0.31	0.86 ± 0.04	0.07 ± 0.03	1.29 ± 0.40
Europe	DJF	0.50 ± 0.27	0.06 ± 0.14	1.30 ± 0.47	0.94 ± 0.03	0.03 ± 0.01	1.44 ± 0.77
	MAM	0.45 ± 0.19	0.06 ± 0.09	1.36 ± 0.41	0.95 ± 0.02	0.02 ± 0.01	1.28 ± 0.71
	JJA	0.43 ± 0.17	0.06 ± 0.08	1.48 ± 0.41	0.95 ± 0.02	0.02 ± 0.01	1.28 ± 0.62
	SON	0.48 ± 0.23	0.04 ± 0.07	1.43 ± 0.33	0.95 ± 0.03	0.02 ± 0.01	1.21 ± 0.56
N-America	DJF	0.50 ± 0.20	0.03 ± 0.01	1.60 ± 0.22	0.89 ± 0.04	0.06 ± 0.03	1.26 ± 0.32
	MAM	0.54 ± 0.21	0.03 ± 0.01	1.65 ± 0.16	0.92 ± 0.03	0.04 ± 0.02	1.25 ± 0.27
	JJA	0.66 ± 0.41	0.02 ± 0.01	1.76 ± 0.22	0.95 ± 0.02	0.03 ± 0.02	1.18 ± 0.26
	SON	0.62 ± 0.45	0.02 ± 0.01	1.70 ± 0.26	0.95 ± 0.03	0.03 ± 0.02	1.25 ± 0.28
S-America	DJF	0.49 ± 0.24	0.06 ± 0.04	1.37 ± 0.33	0.94 ± 0.04	0.03 ± 0.02	0.95 ± 0.49
	MAM	0.44 ± 0.19	0.07 ± 0.09	1.25 ± 0.42	0.94 ± 0.04	0.02 ± 0.02	1.15 ± 0.66
	JJA	0.73 ± 0.53	0.05 ± 0.11	1.66 ± 0.48	0.92 ± 0.04	0.05 ± 0.03	1.30 ± 0.47
	SON	0.93 ± 0.70	0.03 ± 0.02	1.69 ± 0.23	0.93 ± 0.03	0.06 ± 0.04	1.16 ± 0.32

Table 4

Centroids values of aerosol optical parameters obtained from three different clusters analysis for each continents with different aerosol types (D: dust, B: biomass burning, U: urban/industrial, M: mixed aerosols).

Sites	Cluster	Percentage	Aerosol	Aerosol optical parameters							
				No.	distribution	type	Fine_mode	EAE	AAE	SSA	RIR
Africa	1	31	B	0.57		1.17	1.46	0.88	1.44		
	2	40	D	0.24		0.25	3.01	0.92	1.44		
	3	29	M	0.22		0.33	2.02	0.88	1.50		
Asia	1	40	B	0.99		1.38	1.26	0.94	1.44		
	2	9	D	0.23		0.30	2.75	0.91	1.50		
	3	51	M	0.48		1.06	1.39	0.88	1.50		
Australia	1	46	B	0.50		1.63	1.28	0.84	1.49		
	2	50	U	0.41		1.42	1.22	0.90	1.45		
	3	4	D	0.20		0.41	2.84	0.87	1.53		
Europe	1	39	B	0.58		1.59	1.11	0.97	1.40		
	2	51	U	0.40		1.50	1.15	0.94	1.47		
	3	10	D	0.23		0.48	2.78	0.92	1.47		
N-America	1	65	U	0.57		1.82	1.21	0.96	1.48		
	2	29	M	0.51		1.52	1.23	0.91	1.46		
	3	6	B	1.82		1.74	1.39	0.95	1.50		
S-America	1	84	U	0.61		1.66	1.15	0.93	1.47		
	2	2	D	0.18		0.18	2.79	0.94	1.45		
	3	14	B	2.14		1.80	1.32	0.94	1.49		

have similar threshold values of EAE and AAE. However, the separation has improved significantly after using SSA and refractive index which reveal about absorbing and non-absorbing nature of different size ranges (Lee et al., 2010; Bibi et al., 2016; Russell et al., 2010). Generally, urban aerosols have higher SSA than the biomass burning aerosols. Aerosol classification from six different continents is performed using spectral clustering techniques on a wide range of land cover types which included coastal locations (marine), clean continental sites, remote high-altitude sites, locations that are heavily polluted regions due to urban/industrial activities, and biomass fire active region from 150 AERONET sites across the globe. Due to such large heterogeneity of land cover types as well as different meteorological parameters, it is likely to have different aerosol types. From the studies of LULC and aerosol classification, there are three main dominant aerosol types in the current work, which are dust, urban and biomass aerosols. Further,

mixed aerosol types of these three aerosols are also expected to present in the classification due to seasonal changes and transportation process. Similar with the current work, these three aerosol types are mainly reported in several studies for aerosol classification (Russell et al., 2010; Lee et al., 2010; Hamill et al., 2016; Bibi et al., 2016; Yang et al., 2021; Giles et al., 2012). The clustering techniques are performed for all the unorganized data into multiple groups based upon their uniqueness without making any assumption on the formation of clusters. Since the spectral clustering method needs unorganized data structures, each continent is grouped into one data set to apply the clustering algorithm using the five aerosol optical variables. Table 4 shows the percentage distribution of three main aerosol types for six different continents in the current work. The details of these cluster analysis for each continent are discussed in the following sections.

4.3.1. Africa

Although the African continent shows 64% of the total selected sites under vegetative/plantation land cover type in Table 1, most of the sites are closer to the desert regions, or in regions where desert dust is prominent as a result of long-range transport (Omar et al., 2005; Lee et al., 2010). Such dominance of desert dust aerosols in the continent is also supported by the low values of EAE (0.55 ± 51) and high values of AAE (2.23 ± 0.86) as shown in Table 2. The contribution due to urban/industrial aerosols is found to be weak from the selected AERONET sites in the continent. Table 4 shows the centroids values of five different aerosol optical parameters at three different clusters for reference aerosol types, such as dust (D), biomass (B), urban (U) and mixed (M). Based on the spectral cluster method, 40% of the total sites in Africa are dominated by dust aerosols, while 31% and 29% of the total sites are dominated by biomass and mixed aerosol types, respectively. Such different types of aerosols are illustrated from different reference clusters using the five aerosol optical parameters such as AAE, AAEE (at 870 nm/440 nm), AOD fine-mode, SSA and RRI (at 440 nm) shown in Fig. 4. The open circles colored by blue, orange and red are represented by centroid values of biomass, dust and mixed aerosols, respectively in the Figure.

Biomass burning aerosols in Africa are found to be one of the major sources of aerosols after dust aerosols. Dajuma et al. (2021) reported that there are increasing levels of local pollution from long-range transported sources of aerosols including biomass burning from

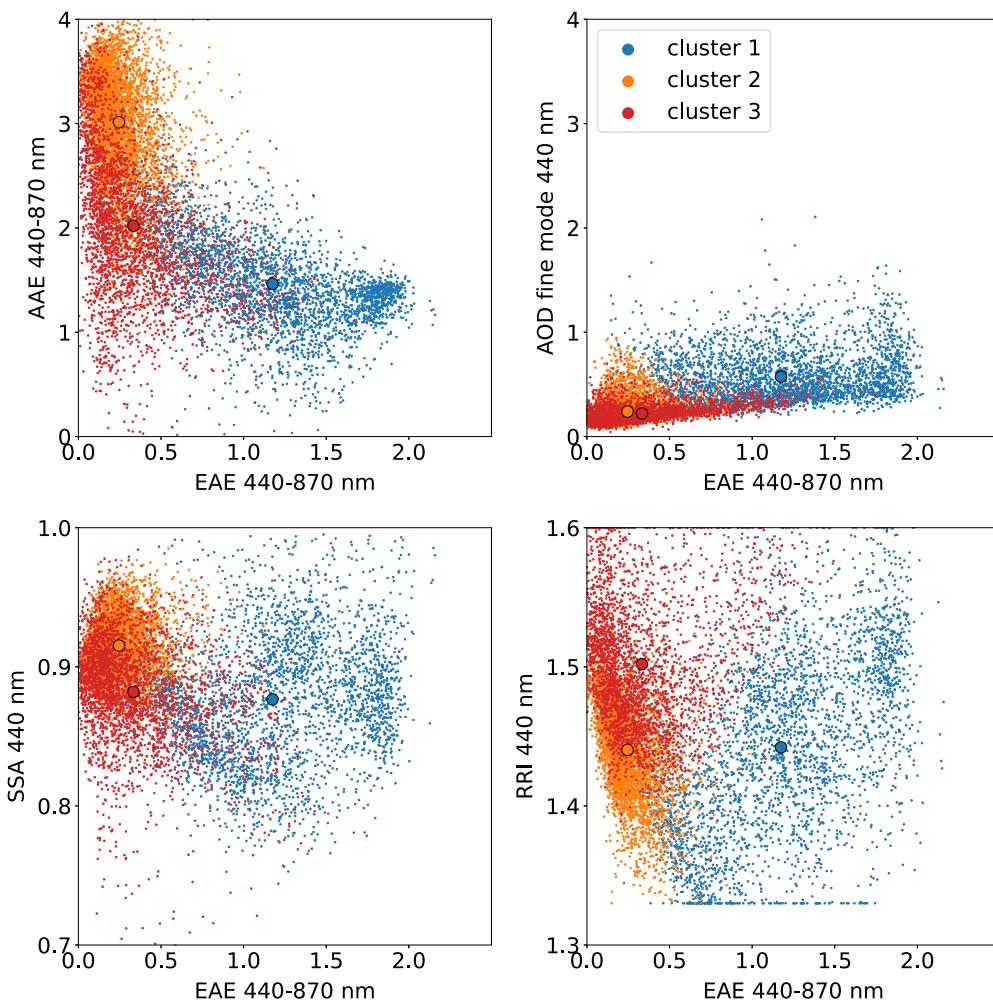


Fig. 4. Scatter plots between EAE_440-870 vs AAE_440-870, AOD_fine_mode, SSA_440, and RRI_440 using spectral clustering algorithm over Africa. The open circles colored by blue, orange and red are represented by centroid values of biomass, dust and mixed aerosols, respectively. For interpretation of color references, the reader is referred to the web version of this article.

Central Africa and mineral dust from the Sahel and the Sahara over South West Africa. The biomass plumes originating from Central and Southern Africa as a result of agricultural activities, land management, livestock grazing, and crop production are further carried westward by jet towards the tropical Eastern Atlantic Ocean, the Gulf of Guinea in the mid-level troposphere as reported therein. The continent exhibits high contribution of dust aerosol in most of the seasons. Further, there are significant contribution of mixed aerosols which may associated with the combination of dust (transported) and biomass burning aerosols. Such mixed aerosols are represented in the overlapping data (see Fig. 4 in cluster 3), which indicate mixing of different aerosol types either due to influx or transported towards the observing sites. The reference centroid values of the mixed aerosol types are EAE (0.33), fine mode (0.22), AAE (2.02), SSA (0.88), and RIR (1.50) by open circles in red color in the Figure. Further, it is also possible that our technique may not be sufficient to distinguish them separately. Such overlapping cases are also observed in other clustering methods like K-mean and DBSCAN as shown in Supplementary Figure S3. The higher SSA values are expected when fine-mode (anthropogenic) and coarse mode (dust) aerosols, which are mostly transported from other major dust source region, are mixed together in the observing site (Che et al., 2015). During MAM the continent exhibits the highest contribution of dust aerosols with 75%, while the biomass burning aerosols peak in DJF with 33%. In addition to these aerosol types, the continent also exhibits urban aerosol type which only occurred in SON (47%) as seen in Table 5. The seasonal reference centroid values for these aerosol

types are shown in Supplementary Table S2. The dominance of biomass burning aerosols during DJF is also found to be coincident with the seasonal peak of fire events in the earlier studies (Ningombam et al., 2023).

4.3.2. Asia

The cluster analysis of aerosol classification in the Asian continent reveals about 51% dominated by mixed aerosols followed by biomass and dust aerosols by 40% and 9%, respectively as shown in Table 4. The open circles colored by blue, orange and red are represented by centroid values of biomass, dust and mixed aerosols, respectively in Fig. 5. The dominance of fine mode aerosols (urban/biomass) are also supported by the high values of EAE (1.11 ± 0.44) and AOD_Fine mode (0.65 ± 0.48) with moderate AAE (1.45 ± 58) as seen in Table 2. Although there is a significant number of urban sites, the biomass burning aerosols in the Asian region is an important contributor to air pollution in the Asian continent (Streets et al., 2003). The dominance of BC aerosols from urban/industrial regions in North and South Asia are 66.2% and 89.2%, respectively as reported by Lee et al. (2010) from the studies of total of 66 AERONET sites in the region. Further, the authors also reported that dust and mixed aerosols contributed to 8.4% and 11.2% in Northern Asia and 0.3% and 3.9% in South Asia, respectively.

The contribution of dust aerosol on the continent is found to be minimal as compared to biomass and urban/industrial aerosols, as seen in Fig. 5. Seasonal cluster analysis reveals that biomass burning

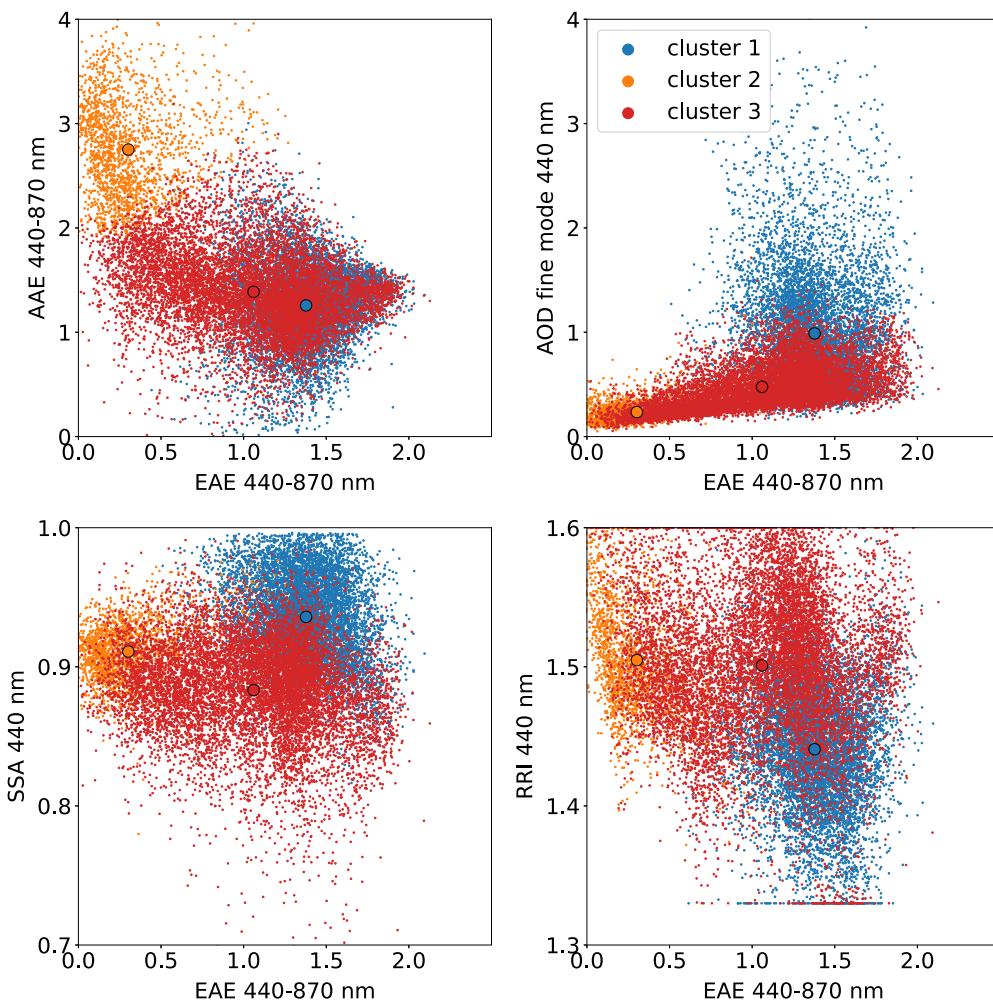


Fig. 5. Scatter plots between EAE₄₄₀₋₈₇₀ vs AAE₄₄₀₋₈₇₀, AOD_{fine_mode}, SSA₄₄₀, and RRI₄₄₀ using spectral clustering algorithm over Asia. The open circles colored by blue, orange and red are represented by centroid values of biomass, dust and mixed aerosols, respectively. For interpretation of color references, the reader is referred to the web version of this article.

Table 5

Seasonal distribution (%) of different aerosol types (D: dust, B: biomass burning, U: urban, M: mixed aerosols) based on centroid values for five aerosol optical parameters at six continents.

Sites	Seasons											
	DJF			MAM			JJA			SON		
	Cluster			Cluster			Cluster			Cluster		
	1	2	3	1	2	3	1	2	3	1	2	3
Africa	33%, M	34%, D	33%, B	75%, D	11%, B	14%, M	27%, B	43%, D	30%, M	47%, U	46%, D	7%, B
Asia	96%, B	1%, U	3%, D	48%, M	41%, B	11%, D	48%, D	47%, U	5%, B	78%, B	17%, D	5%, U
Australia	26%, B	38%, U	36%, D	NA	NA	NA	47%, B	23%, U	30%, M	56%, B	3%, D	41%, U
Europe	70%, U	16%, B	14%, D	42%, B	11%, D	47%, U	44%, B	46%, U	10%, D	85%, U	8%, B	7%, D
North America	45%, B	19%, M	36%, U	58%, B	31%, M	11%, U	33%, U	57%, B	10%, B	34%, M	17%, B	49%, U
South America	39%, U	45%, M	16%, B	31%, U	11%, D	58%, B	80%, M	13%, B	7%, D	52%, M	31%, U	17%, B

aerosols peak during DJF (96%) and SON (78%), while there are significant contribution of mixed aerosol types during MAM (48%). Further, significant contributions of urban and dust aerosols are also observed by 47%, and 48%, respectively of the total data set during JJA as seen in [Table 5](#).

4.3.3. Australia

The main aerosol types available in Australia are dominated by fine-mode aerosols, which may be attributed to biomass burning, urban, or mixed aerosol types as described in the earlier Section 4.2. Such aerosol climatological results are supported by the aerosols classification obtained using spectral clustering techniques from the 11 AERONET sites.

The percentage contributions of the three main aerosol types are 50%, 46%, and 4% for urban, biomass burning, and dust aerosols, respectively, as shown in [Table 4](#). The dominance of fine-mode aerosols (biomass/urban) is also supported from the high values of EAE (1.47 ± 0.35), AOD_{Fine-mode} (0.44 ± 0.16), and moderate high AAE (1.31 ± 0.46), as seen in [Table 2](#). The separation of fine-mode aerosols for biomass and urban aerosols can be clearly seen in the scatter plots of EAE₄₄₀₋₈₇₀ vs AAE₄₄₀₋₈₇₀, AOD_{fine-mode}, SSA₄₄₀, and RRI₄₄₀ using spectral clustering algorithm in [Fig. 6](#). The open circles colored by blue, green and orange are represented by centroid values of biomass, urban and dust aerosols, respectively in the Figure. Seasonal analysis of cluster analysis is performed at the 11 AERONET sites in

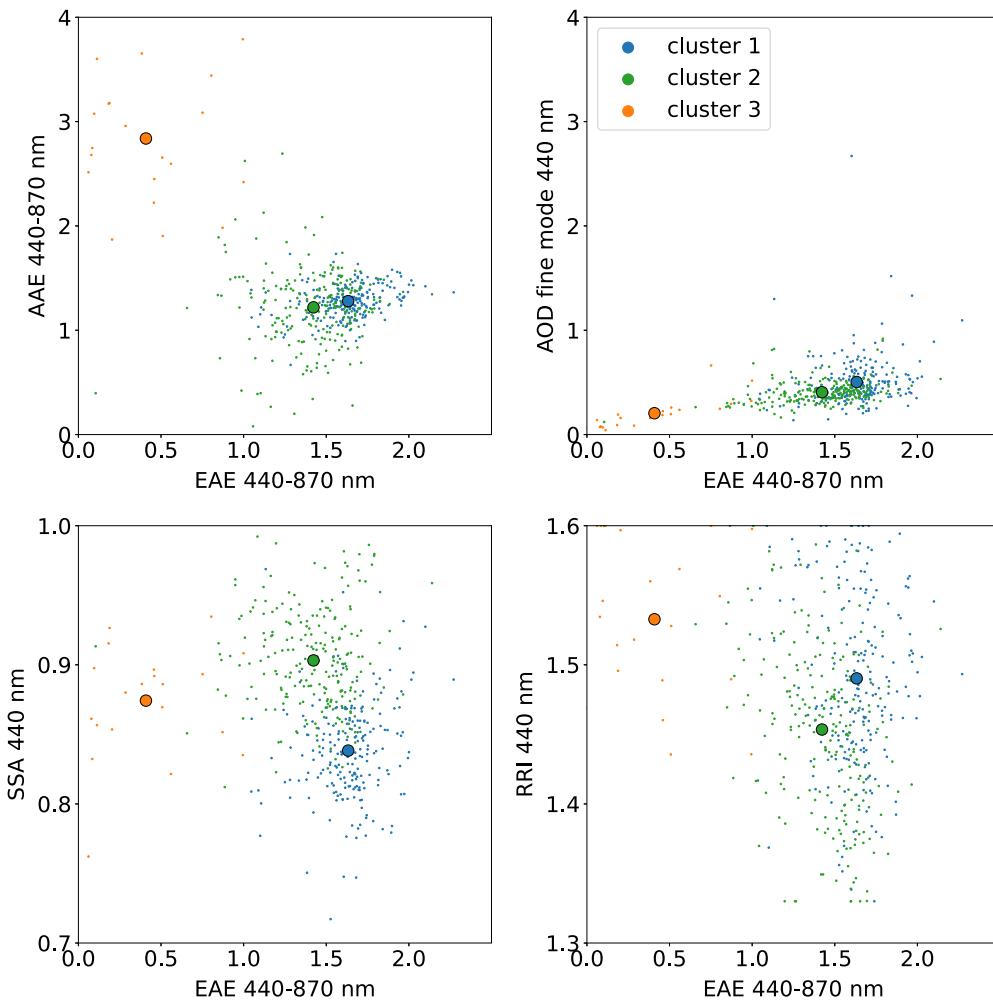


Fig. 6. Scatter plots between EAE₄₄₀₋₈₇₀ vs AAE₄₄₀₋₈₇₀, AOD_{fine mode}, SSA₄₄₀, and RRI₄₄₀ using spectral clustering algorithm over Australia. The open circles colored by blue, green and orange are represented by centroid values of biomass, urban and dust aerosols, respectively. For interpretation of color references, the reader is referred to the web version of this article.

Australia. However, due to a lack of sufficient data during MAM, we could not include any result for this season. The continent observed a strong dominance of dust aerosols only during DJF (36%), while biomass burning aerosols are observed during JJA (47%) and SON (56%) as shown in Table 5. The peak of biomass burning during SON is found to be coincided with the peak of fire emissions during the same season (Ningombam et al., 2023). Further, the Australian continent also exhibits mixed aerosol types during JJA (30%) as shown in Table 5 by cluster 3 with reference aerosol optical parameters of low SSA: 0.82, and high values EAE: 1.68 and RRI:1.57, as shown in the Supplementary Table S2.

4.3.4. Europe

Fig. 7 shows scatter plots for EAE₄₄₀₋₈₇₀ vs AAE₄₄₀₋₈₇₀, AOD_Fine-mode, SSA₄₄₀, and RRI₄₄₀ from 32 AERONET sites in Europe using spectral clustering algorithm and the open circles colored by blue, green and orange are represented by centroid values of biomass, urban and dust aerosols, respectively. These three clusters have three centroid values which corresponds to three reference aerosols, as shown in Table 5 and Supplementary Table S2. The European continent is mostly dominated by urban aerosol with 51% followed by biomass and dust aerosols with 39% and 10%, respectively, according to spectral aerosol types, as shown in Table 4. Seasonally, urban aerosols are dominant in all the seasons with 70%, 47%, 46%, and 85% during DJF, MAM, JJA, and SON, respectively, as seen in Table 5. The continent also shows

a strong presence of biomass burning aerosols during DJF, MAM, and JJA by 16%, 42%, and 44%, respectively. Among the aerosol types, dust aerosol exhibits the least, with less than 15% of the total aerosol contribution in the continent.

4.3.5. North America

According to spectral cluster analysis, North America contributed most significantly by urban aerosols (65%), followed by mixed (29%) and biomass (6%) aerosols, as shown in Table 4. Fig. 8 displays scatter plots between EAE₄₄₀₋₈₇₀ vs AAE₄₄₀₋₈₇₀, AOD_Fine-mode, SSA₄₄₀, and RRI₄₄₀ using spectral clustering algorithm over 28 AERONET sites in the continent. There is no significant contribution of dust aerosols in the continent as there are hardly any days for EAE values less than 0.5, which we assumed for dust/desert aerosol types. According to Liu et al. (2008), the main global dust source regions are located in North Africa, the Arabian Peninsula, the Indian sub-continent, and East Asia, based on cloud-free conditions and high-resolution global dust aerosols measurements taken from Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) data. The contribution of urban aerosol is found to be the least during MAM (11%), and moderate during SON (49%), DJF (36%), and JJA (33%). There are significant contributions of mixed aerosols during SON (34%), MAM (31%), and DJF (19%), despite significant contributions of biomass burning aerosols during JJA (57%), MAM (58%), and DJF (45%) as shown in Table 5.

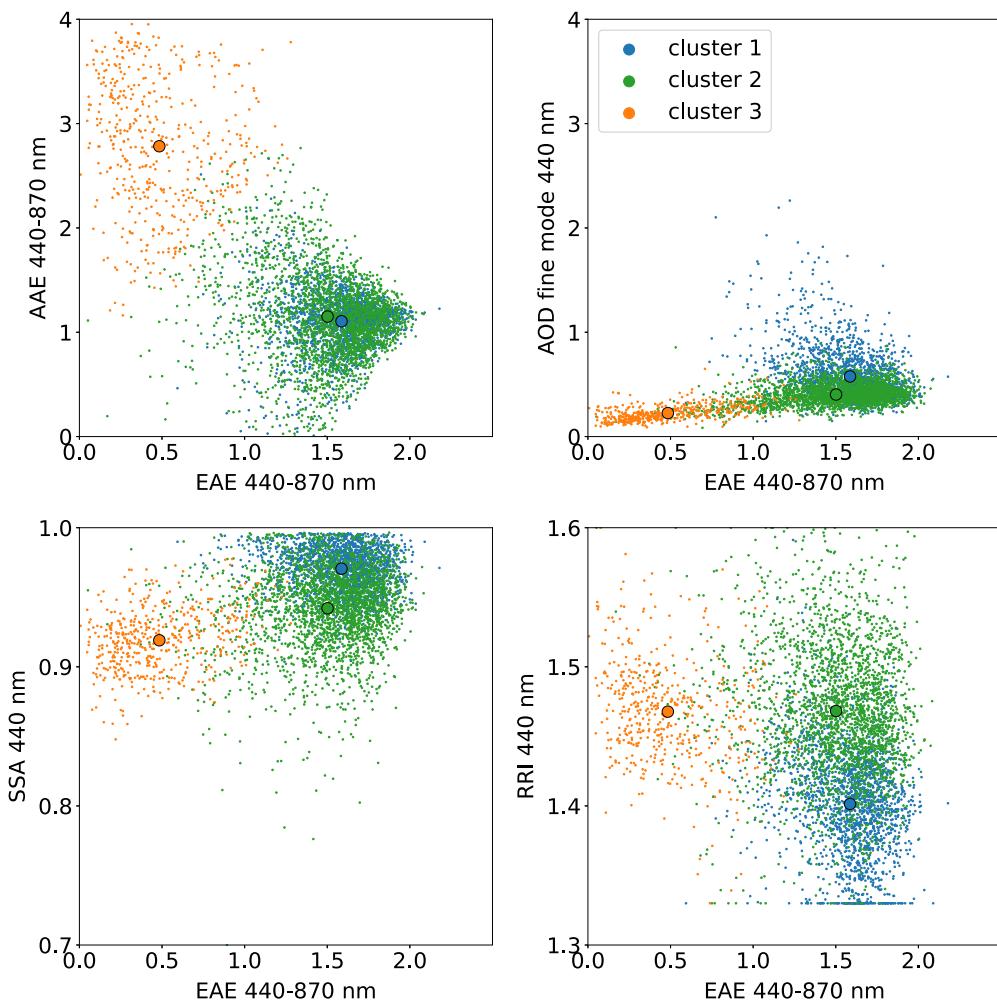


Fig. 7. Scatter plots between EAE_440-870 vs AAE_440-870, AOD_fine_mode, SSA_440, and RRI_440 using spectral clustering algorithm over Europe. The open circles colored by blue, green and orange are represented by centroid values of biomass, urban and dust aerosols, respectively. For interpretation of color references, the reader is referred to the web version of this article.

4.3.6. South America

The South American continent is dominated by urban aerosols with 84% followed by biomass burning and dust aerosols with 14% and 2%, respectively according to spectral cluster analysis shown in Table 4. Unlike North America, there are a considerable number of days for which $EAE < 0.5$ as shown in Fig. 9, indicating the presence of dust aerosols in the continent. However, the low contribution of dust aerosol from the yearly data did not show any significant distribution in the seasonal data. Further, it is found that the high contribution of urban aerosols (84%) from the total (mean) data has emerged into mixed aerosol types during the seasonal cluster analysis. The highest mixed aerosol types are observed during JJA (80%), SON (52%), and DJF (45%). The dominant mixed aerosol types are suggested from the low values of SSA (0.89–0.92) and high values of EAE (1.36–1.79) and RRI (1.47–1.49) as shown in Table 5 and Supplementary Table S2. Furthermore, the Supplementary Figure S2 shows the global RRI at 440 nm from 150 AERONET sites in the current study. South America is one of the largest contributors of biomass burning emissions globally, which originated mostly from the Northern South America (Ballesteros-González et al., 2020). Moreover, the continent contributed about 17% of the global fire data from the studies of more than two decades of MODIS fire count data (Ningombam et al., 2023). The contribution of biomass burning aerosols in the current study is dominated during MAM (58%), while the rest of the seasons show less than 20%. Furthermore, the contribution of urban aerosol is evident during DJF (39%), MAM (31%), and SON (31%), and there is no evidence of this aerosol type during JJA.

4.4. Aerosol classification: Threshold method

The classification of aerosols using spectral clustering method has examined with results obtained from threshold methods (Dumka et al., 2020; Bibi et al., 2016; Lee et al., 2010; Ningombam et al., 2014a, 2015). However, among the aerosol optical parameters, fine mode fraction (FMF) and SSA are found to be more suitable parameter when the different sites are considering in the same region (continent). The FMF denotes the contribution of dominant size to the total columnar AOD, while the SSA is an intrinsic property of columnar aerosol that provides the scattering fraction in total extinction. Therefore, we used FMF and SSA are used in the threshold value of aerosol classification in the current work. The aerosol classification using different threshold values of FMF (500 nm) and SSA (440 nm) is based on the methodology adopted by Lee et al. (2010) and more details of the aerosols classification are reported therein. Aerosol classification under the threshold method characterized dust, mixture, non-absorbing, BC and uncertainties which are found to be beyond the threshold values. Dust aerosols are classified as $FMF < 0.4$ and $SSA < 0.95$ as mentioned by Lee et al. (2010). Such aerosol types are found to be highest in Africa (55%) among the continents and these results are consistent with the spectral clustering method. Fine-mode aerosols, which may attribute to urban/industrial aerosols, are classified as $FMF > 0.6$ and $SSA > 0.95$ as mentioned in the threshold method. Such fine-mode aerosols are dominated in Europe, and North-America as seen in the Table 6. However, BC aerosols, which classified as $FMF > 0.6$ and $SSA < 0.95$ in the threshold method, are

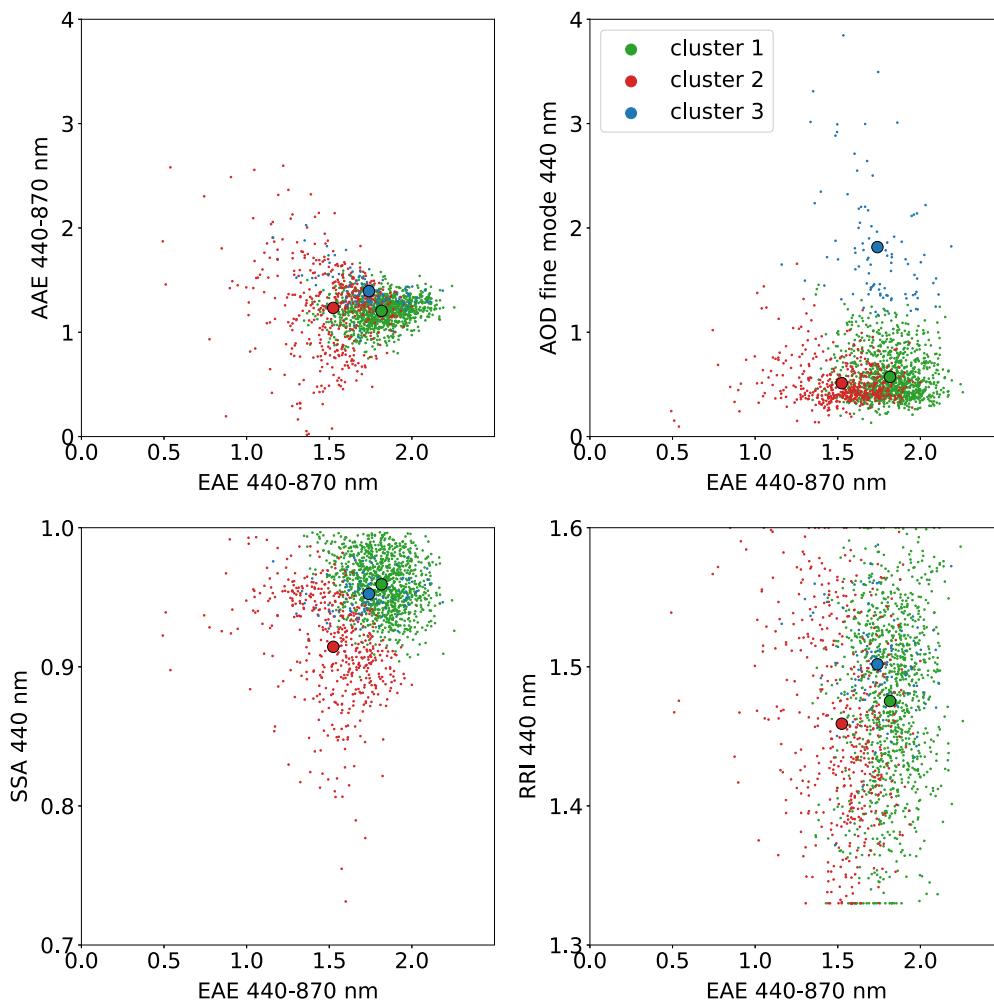


Fig. 8. Scatter plots between EAE_440-870 vs AAE_440-870, AOD_fine_mode, SSA_440, and RRI_440 using spectral clustering algorithm over North America. The open circles colored by green, red and blue are represented by centroid values of urban mixed and biomass aerosols, respectively. For interpretation of color references, the reader is referred to the web version of this article.

Table 6

Classification of aerosols based on the threshold values of fine mode fraction (FMF) and SSA from total 150 AERONET sites in the six continents in the current work. These statistical analysis are made in % for each continent. The Table shows uncertainty (Uncer.), dust, mixture (Mix.), Non-absorbing (NA), and black carbon (BC) aerosol types.

Continents	Uncer.	Dust	Mix.	N.A.	BC
Africa	2	55	27	1	15
Asia	0	12	25	13	50
Australia	0	3	12	5	81
Europe	0	5	11	54	30
N-America	0	0	2	52	46
S-America	1	2	3	26	68

found to be maximum in Asia, Australia, and South America as seen in the Table 6. Classification of key aerosol types are identified from different values of SSA as reported by Dubovik et al. (2002). In their studies, the threshold values of SSAs are 0.90–0.98 for urban/industrial aerosol, 0.89–0.95 for biomass burning, 0.92–0.93 for desert dust, and 0.98 for oceanic aerosols.

The non-absorbing aerosols or weakly absorbing aerosols for $\text{SSA} > 0.95$ in the Table 6 can be characterized as urban/industrial aerosols. Weakly absorbing aerosols are usually emitted by anthropogenic activities, such as SO_2 produced by industries, injected into the atmosphere. Such weakly absorbing aerosols are confined in North America, Europe, East of Asia and Southern Africa (Lin et al., 2021).

We observed that each method of aerosol classification has its limitation as described earlier. For example, the threshold value of aerosol classification using fine mode fraction and SSA (Lee et al., 2010) has the capacity to identify dust, black carbon (slightly absorbing, moderately absorbing, highly absorbing), mixture and non-absorbing. However, in the case of spectral clustering method (machine learning algorithm), aerosols are classified based on the reference value (centroid) of cluster. Since aerosols are highly varied and its threshold value may not work well in all the continents. Therefore, spectral clustering algorithm has several advantages when we have multiple sites from different topography with different aerosol types.

5. Summary, conclusions, and outlook

In the current work, we classified four main aerosol types: dust, urban/industrial, biomass burning and mixed aerosols using the spectral clustering algorithm of machine learning tool from 39,411 daily observation of 150 AERONET sites in six continents (Africa, Asia, Australia, Europe, North, and South America) during 1993–2022. Further, classification of aerosols using threshold values of FMF and SSA are also performed to examine the results obtained from the spectral clustering method. We observed that each method of aerosol classification has its limitation. The spectral clustering method (machine learning algorithm) is based on the reference value (centroid) of cluster which have similar features among the unsupervised dataset (fine-mode AOD, EAE, AAE, SSA and RRI). Since aerosols are highly varied and its threshold

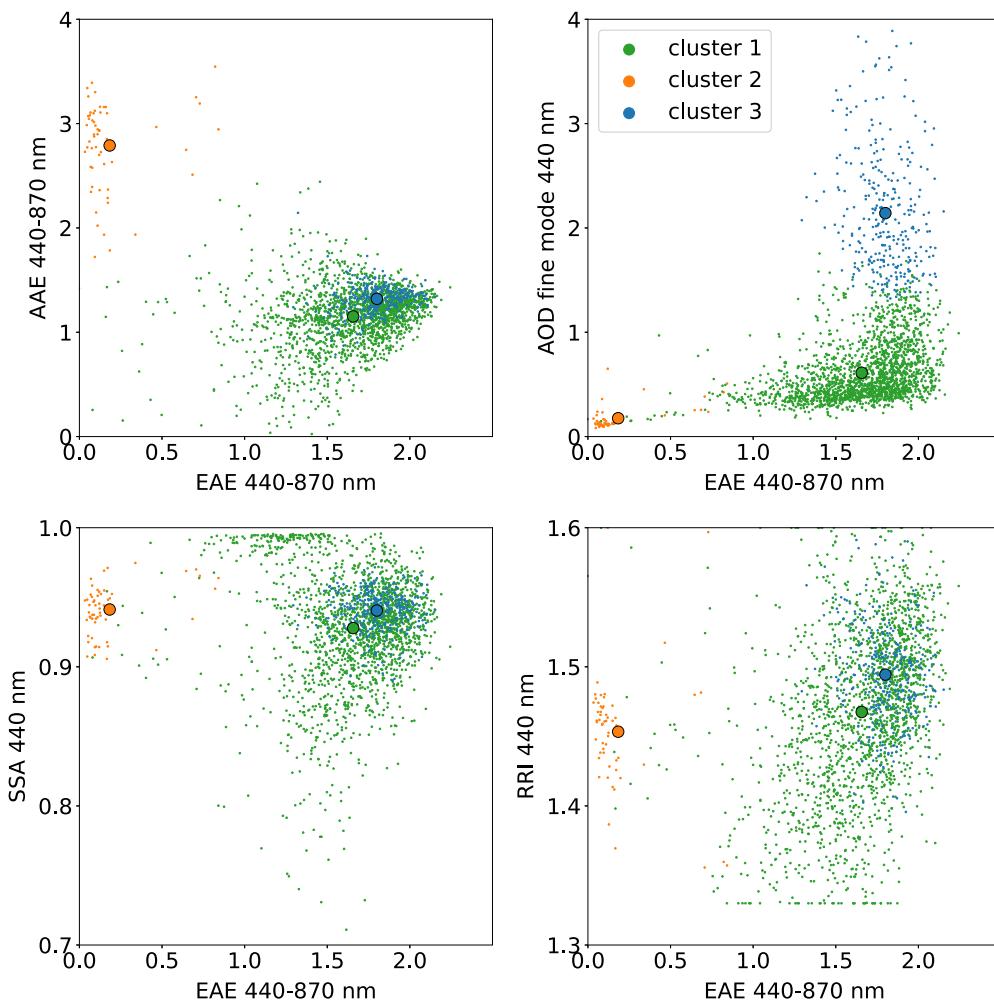


Fig. 9. Scatter plots between EAE₄₄₀₋₈₇₀ vs AAE₄₄₀₋₈₇₀, AOD_{fine_mode}, SSA₄₄₀, and RRI₄₄₀ using spectral clustering algorithm over South America. The open circles colored by green, orange and blue are represented by centroid values of urban, dust and biomass aerosols, respectively. For interpretation of color references, the reader is referred to the web version of this article.

value may not work well in all the continents. Therefore, spectral clustering algorithm has several advantages when we have multiple sites from different topography with different aerosol types. The main results of the aerosol classification in the current work are summarized as below:

Dominance of dust aerosol (40%) is observed in Africa, despite, there are significant contributions of biomass burning (31%) and mixed aerosol (29%) types. Seasonally, dust aerosols are dominated during MAM with 75% contribution, while biomass burning aerosols dominated during DJF with 33% contribution.

Asian continent exhibits significant contribution of mixed aerosol types (51%), despite there are significant contribution of biomass burning (40%) aerosols. Similar to Africa, mixed aerosols could be due to combination of transported dust/desert aerosols with urban/biomass burning aerosols. Seasonally, high biomass burning aerosols are observed during DJF (96%) and SON (78%). Further dust and urban aerosols are found to be dominant with 48% and 47%, respectively during JJA, while the mixed aerosols contribute by 48% during MAM.

Australia exhibits the highest biomass burning aerosols during SON with 56%, although the continent has different aerosol types dominated by urban aerosols with 50% followed by biomass burning aerosols with 46%. The continent shows mixed aerosol types during JJA with 30% contribution, which may be associated with the combination of urban and biomass burning aerosols as suggested from the values of low SSA (0.82) and high EAE (1.68) and RRI (1.57).

Europe has significant contribution of urban with 51% and biomass burning aerosols with 39% contribution. The high contribution of urban aerosols is also supported by the maximum number of urban sites (71%) located in the continent from the selected 32 AERONET sites in the current work.

North America has the highest contribution of urban aerosols with 65% contribution followed by mixed with 29% contribution and biomass burning with 6% contribution. Seasonally, the continent exhibits mixed aerosol types during DJF with 19% contribution, MAM with 31% contribution, and SON with 34% contribution, while dominance of biomass burning aerosols observed during MAM-JJA contributed by 57%–58%. Further, there are significant contribution of urban aerosols during all the seasons ranging from 11 to 49%.

The highest fine-mode AOD among the continents is observed in South America (0.81 ± 0.63) which could be due to the dominance of urban aerosols with 84% contribution and biomass burning aerosols with 14% contribution. Seasonally, this continent shows dominance of mixed aerosol types with 80% contribution during JJA, followed by 45% contribution in DJF and 52% contribution in SON, despite the seasonal peak of biomass burning aerosols in MAM with 58% contribution. It is found that the high contribution of urban aerosols with 84% from the multi-year data has merged with the biomass burning aerosols to form mixed aerosol types.

Finally, we demonstrated that the spectral clustering algorithm can separate dust, biomass burning, urban, and mixed aerosol types by comparing results with threshold values based method. Further, other

aerosol types such as marine or continental origin cannot be identified as the inversion AERONET product such as AOD are relatively low with value less than ~0.20, restricting to retrieve full aerosol optical parameters, including SSA, AAOD etc.

CRediT authorship contribution statement

Shantikumar S. Ningombam: Conceptualization, Methodology, Writing – original draft, Data reduction, Figures & analysis. **E.J.L. Larson:** Data reduction, Figures, Visualization, Editing. **G. Indira:** Data reduction, Visualization and plotting. **B.L. Madhavan:** Conceptualization & reviewing the manuscript. **Pradeep Khatri:** Conceptualization, Plotting and reviewing the manuscript.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors are thankful to the anonymous Reviewers for providing constructive comments which have improved the manuscript significantly. The AERONET data used in the present work are taken from 150 sites across the globe from the website: <https://aeronet.gsfc.nasa.gov> and the authors are thankful to the PI, Co-PIs and their staff of the respective AERONET sites who maintain the data for scientific use. We also acknowledge for using global Land Used and Land cover products in the current work from the European Space Agency (ESA) Climate Change Initiative (CCI) data. One of the authors, SSN thanks the Ministry of Earth Sciences, Government of India for providing a financial support under the project grant number: MoES/16/02/2021-RDESS.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at <https://doi.org/10.1016/j.apr.2023.102026>.

References

- Ballesteros-González, K., Sullivan, A.P., Morales-Betancourt, R., 2020. Estimating the air quality and health impacts of biomass burning in northern South America using a chemical transport model. *Sci. Total Environ.* (ISSN: 0048-9697) 739, 139755. <http://dx.doi.org/10.1016/j.scitotenv.2020.139755>.
- Bellouin, N., Quaas, J., Gryspenert, E., Kinne, S., Stier, P., Watson-Parris, D., et al., 2020. Bounding global aerosol radiative forcing of climate change. *Rev. Geophys.* 58, e2019RG000660. <http://dx.doi.org/10.1029/2019RG000660>.
- Bibi, H., Alam, K., Bibi, S., 2016. In-depth discrimination of aerosol types using multiple clustering techniques over four locations in Indo-Gangetic plains. *Atmos. Res.* 181, 106–114.
- Che, H., et al., 2015. Application of aerosol optical properties to estimate aerosol type from ground-based remote sensing observation at urban area of northeastern China. *J. Atmos. Sol.-Terr. Phys.* 132, 37–47. <http://dx.doi.org/10.1016/j.jastp.2015.06.015>.
- Chen, Q.X., Shen, W.X., Yuan, Y., Tan, H.P., 2019. Verification of aerosol classification methods through satellite and ground-based measurements over Harbin, Northeast China. *Atmos. Res.* 216, 167–175. <http://dx.doi.org/10.1016/j.atmosres.2018.09.022>.
- Cheng, Y.F., Wiedensohler, A., Eichler, H., Su, H., Gnauck, T., Brueggemann, E., Herrmann, H., Heintzenberg, J., Slanina, J., Tuch, T., Hu, M., Zhang, Y.H., 2008. Aerosol optical properties and related chemical apportionment at xinken in Pearl River Delta of China. *Atmos. Environ.* 42, 6351–6372. <http://dx.doi.org/10.1016/j.atmosenv.2008.02.034>.
- Dajuma, A., et al., 2021. Biomass burning effects on the climate over Southern West Africa during the summer monsoon. In: Oguge, N., Ayal, D., Adeleke, L., da Silva, I. (Eds.), *African Handbook of Climate Change Adaptation*. Springer, Cham, http://dx.doi.org/10.1007/978-3-030-45106-6_86.
- Damle, A., Minden, V., Ying, L., 2019. Simple, direct and efficient multi-way spectral clustering. *Inf. Inference: J. IMA* 8, 181–203. <http://dx.doi.org/10.1093/imaiai/ayy008>.
- Ding, A.J., 2013. Intense atmospheric pollution modifies weather: a case of mixed biomass burning with fossil fuel combustion pollution in eastern China. *Atmos. Chem. Phys.* 13, 10545–10554. <http://dx.doi.org/10.5194/acp-13-10545-2013>.
- Dubovik, O., Holben, B., Eck, T.F., Smirnov, A., Kaufman, Y.J., King, M.D., Tanré, D., Slutsker, I., 2002. Variability of absorption and optical properties of key aerosol types observed in worldwide locations. *J. Atmos. Sci.* 59, 590–608. [http://dx.doi.org/10.1175/1520-0469\(2002\)059<0590:VOAAOP>2.0.CO;2](http://dx.doi.org/10.1175/1520-0469(2002)059<0590:VOAAOP>2.0.CO;2).
- Dubovik, O., King, M.D., 2000. A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. *J. Geophys. Res.* 105, 20673–20696.
- Dumka, U.C., Ningombam, Shantikumar S., Kaskaoutis, D.G., Madhavan, B.L., Song, H.-J., Angchuk, Dorje, Sonam, Jorphail, 2020. Long-term (2008–2018) aerosol characteristics at high-altitude location over western trans-himalayan region: role of coarse and fine mode aerosols in aerosol radiative forcing estimation. *Sci. Total Environ.* 734, 139354.
- Eck, T.F., et al., 2001. Characterization of the optical properties of biomass burning aerosols in Zambia during the 1997 ZIBBEE field campaign. *J. Geophys. Res. Space Phys.* 106, 3425–3448, 2001.
- Feng, S., Li, W., Xu, J., Liang, T., Ma, X., Wang, W., Yu, H., 2022. Land use/land cover mapping based on GEE for the monitoring of changes in ecosystem types in the upper Yellow River basin over the Tibetan Plateau. *Remote Sens.* 14 (5361), <http://dx.doi.org/10.3390/rs14215361>.
- Giles, D.M., Holben, B.N., Eck, T.F., Sinyuk, A., Smirnov, A., Slutsker, I., Dickerson, R.R., Thompson, A.M., Schafer, J.S., 2012. An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source regions. *J. Geophys. Res.* 117 (D17203), <http://dx.doi.org/10.1029/2012JD018127>.
- Ginoux, P., Prospero, J.M., Gill, T.E., Hsu, N.C., Zhao, M., 2010. Global-scale attribution of anthropogenic and natural dust sources and their emission rates based on MODIS Deep Blue aerosol products. *Rev. Geophys.* 50, 10875–10893.
- Hamill, P., et al., 2016. An AERONET-based aerosol classification using the Mahalanobis distance. *Atmos. Environ.* 140, 213–233. <http://dx.doi.org/10.1016/j.atmosenv.2016.06.002>.
- Holben, B.N., et al., 1998. AERONET-A federated instrument network and data archive for aerosol characterization. *Remote Sens. Environ.* 66, 1–16.
- Huang, X., 2016. Effects of aerosol–radiation interaction on precipitation during biomassburning season in East China. *Atmos. Chem. Phys.* 16, 10063–10082. <http://dx.doi.org/10.5194/acp-16-10063-2016>.
- Jin, X., Han, J., 2011. K-means clustering. In: Sammut, C., Webb, G.I. (Eds.), *Encyclopedia of Machine Learning*. Springer, Boston, MA, http://dx.doi.org/10.1007/978-0-387-30164-8_425.
- Krishnaveni, A.S., Madhavan, B.L., Ratnam, M.V., 2023. Aerosol classification using fuzzy clustering over a tropical rural site. *Atmos. Res.* 282, 106518. <http://dx.doi.org/10.1016/j.atmosres.2022.106518>, 0169-8095.
- Lee, J., Kim, J., Song, C.H., Kim, S.B., Chun, Y., Sohn, B.J., Holben, B.N., 2010. Characteristics of aerosol types from AERONET Sun photometer measurements. *Atmos. Environ.* 44, 3110–3117. <http://dx.doi.org/10.1016/j.atmosenv.2010.05.035>.
- Levy, R.C., Remer, L.A., Dubovik, O., 2007. Global aerosol optical properties and application to moderate resolution imaging spectroradiometer aerosol retrieval over land. *J. Geophys. Res.* 112 (D13210), <http://dx.doi.org/10.1029/2006JD007815>.
- Li, J., Hendricks, J., Righi, M., Beer, C.G., 2022. An aerosol classification scheme for global simulations using the K-means machine learning method. *Geosci. Model Dev.* 15, 509–533. <http://dx.doi.org/10.5194/gmd-15-509-2022>.
- Lin, J., Zheng, Y., Shen, X., Xing, L., Che, H., 2021. Global aerosol classification based on aerosol robotic network (AERONET) and satellite observation. *Remote Sens.* 13 (1114), <http://dx.doi.org/10.3390/rs13061114>.
- Liu, D., Wang, Z., Liu, Z., Winker, D., Trepte, C., 2008. A height resolved global view of dust aerosols from the first year CALIPSO lidar measurements. *J. Geophys. Res.* 113 (D16214), <http://dx.doi.org/10.1029/2007JD009776>.
- Logothetis, S.-A., et al., 2020. Aerosol classification in Europe, Middle East, North Africa and Arabian Peninsula based on AERONET version 3. *Atmos. Res.* 239, 104893. <http://dx.doi.org/10.1016/j.atmosres.2020.104893>.
- Lu, Z., Zhang, Q., Streets, D.G., 2011. Sulfur dioxide and primary carbonaceous aerosol emissions in China and India 1996–2010. *Atmos. Chem. Phys.* 11, 9839–9864.
- Ningombam, S.S., Bagare, S.P., Khatri, P., Sohn, B.J., Song, H.-J., 2015. Estimation of aerosol radiative forcing over an aged-background aerosol feature during advection and non-advection events using a ground-based data obtained from a Prede Skyradiometer observation. *Atmos. Res.* 164–165 (2015), 76–83. <http://dx.doi.org/10.1016/j.atmosres.2015.05.001>.
- Ningombam, S.S., Bagare, S.P., Srivastava, A.K., Kanawade, V.P., Singh, R.B., Padhy, S.K., 2014b. Temporal asymmetry in aerosol optical characteristics: A case study at high-altitude station, Hanle, in Ladakh region. *J. Atmos. Sol.-Terr. Phys.* 121, 123–131. <http://dx.doi.org/10.1016/j.jastp.2014.10.012>.
- Ningombam, S.S., Bagare, S.P., Srivastava, A.K., Sohn, B.J., Song, H.-J., Larson, E., 2014a. Aerosol radiative forcing over a high-altitude station Merak, in the trans-himalayan region during advection of anthropogenic events from the Indo-Gangetic Plain. *Atmos. Environ.* 98 (2014), 253–259. <http://dx.doi.org/10.1016/j.atmosenv.2014.08.061>.

- Ningombam, S.S., Dumka, U.C., Srivastava, A.K., Song, H.-J., 2020. Optical and physical properties of aerosols during active fire events occurring in the Indo-Gangetic Plains: Implications for aerosol radiative forcing. *Atmos. Environ.* 223, 117225. <http://dx.doi.org/10.1016/j.atmosenv.2019.117225>.
- Ningombam, S.S., Khatri, P., Larson, E.J.L., Dumka, U.C., Sarangi, C., Vineeth, R., 2023. Classification of MODIS fire emission data based on aerosol absorption Angstrom exponent retrieved from AERONET data. *Sci. Total Environ.* 858, 159898. <http://dx.doi.org/10.1016/j.scitotenv.2022.159898>.
- Omar, A.H., Won, J.-G., Winker, D.M., Yoon, S.-C., Dubovik, O., McCormick, M.P., 2005. Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements. *J. Geophys. Res.* 110, D10S14. <http://dx.doi.org/10.1029/2004JD004874>.
- Peel, M.C., Finlayson, B.L., McMahon, T.A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrol. Earth Syst. Sci.* 11, 1633–1644. <http://dx.doi.org/10.5194/hess-11-1633-2007>.
- Russell, P.B., Bergstrom, R.W., Shinozuka, Y., Clarke, A.D., DeCarlo, P.F., Jimenez, J.L., Livingston, J.M., Redemann, J., Dubovik, O., Strawbridge, A., 2010. Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition. *Atmos. Chem. Phys.* 10, 1155–1169. <http://dx.doi.org/10.5194/acp-10-1155-2010>.
- Russell, P.B., Kacenelenbogen, M., Livingston, J.M., et al., 2014. A multiparameter aerosol classification method and its application to retrievals from spaceborne polarimetry. *J. Geophys. Res.: Atmos.* 119, 9838–9863. <http://dx.doi.org/10.1002/2013JD021411>.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (8), 888–905.
- Smith, S.J., et al., 2011. Anthropogenic sulfur dioxide emissions: 1850–2005. *Atmos. Chem. Phys.* 11, 1101–1116.
- Streets, D.G., Yarber, K.F., Woo, J.-H., Carmichael, G.R., 2003. Biomass burning in Asia: Annual and seasonal estimates and atmospheric emissions. *Glob. Biogeochem. Cycles* 17 (4), 1099. <http://dx.doi.org/10.1029/2003GB002040>.
- Sun, Y., Zhao, C., 2020. Influence of saharan dust on the large-scale meteorological environment for development of tropical cyclone over North Atlantic Ocean Basin. *J. Geophys. Res.: Atmos.* 125, <http://dx.doi.org/10.1029/2020JD033454>.
- Sun, Y., Zhao, C., 2021. Distinct impacts on precipitation by aerosol radiative effect over three different megacity regions of eastern China. *Atmos. Chem. Phys.* 21, 16555–16574. <http://dx.doi.org/10.5194/acp-21-16555-2021>.
- Tassi, A., Gigante, D., Modica, G., Martino, L.Di., Vizzari, M., Pixel-vs, 2021. Object-based landsat 8 data classification in google earth engine using random forest: The case study of Maiella National Park. *Remote Sens.* 13 (2299), <http://dx.doi.org/10.3390/rs13122299>.
- Torres, O., Chen, Z., Jethva, H., et al., 2009. OMI and MODIS observations of the anomalous 2008–2009 Southern Hemisphere biomass burning seasons. *Atmos. Chem. Phys.* 10, 3505–3513, www.atmos-chem-phys.net/10/3505/2010/.
- Vadrevu, K.P., Lasko, K., Giglio, L., Justice, C., 2015. Vegetation fires, absorbing aerosols and smoke plume characteristics in diverse biomass burning regions of Asia. *Environ. Res. Lett.* 10, 105003. <http://dx.doi.org/10.1088/1748-9326/10/10/105003>.
- von Luxburg, U., 2007. A tutorial on spectral clustering. *Stat. Comput.* 17, 395–416. <http://dx.doi.org/10.1007/s11222-007-9033-z>.
- Yang, X., Zhao, C., Yang, Y., Fan, H., 2021. Long-term multi-source data analysis about the characteristics of aerosol optical properties and types over Australia. *Atmos. Chem. Phys.* 21, 3803–3825. <http://dx.doi.org/10.5194/acp-21-3803-2021>.
- Zhao, W., Hopke, P.K., Prather, K.A., 2008. Comparison of two cluster analysis methods using single particle mass spectra. *Atmos. Environ.* 42, 881–892.