

Global aerosol mixtures and their multiyear and seasonal characteristics



M. Taylor ^{a,*}, S. Kazadzis ^{a,b}, V. Amiridis ^c, R.A. Kahn ^d

^a Institute for Environmental Research and Sustainable Development (IERSD), National Observatory of Athens (NOA), Metaxa & Vas Pavlou, Penteli, GR-15236 Athens, Greece

^b Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center (PMOD/WRC), Dorfstrasse 33, CH-7260 Davos Dorf, Switzerland

^c Institute for Astronomy, Astrophysics, Space Applications and Remote Sensing (IAASARS), National Observatory of Athens (NOA), Metaxa & Vas Pavlou, Penteli, 15236 Athens, Greece

^d NASA Goddard Space Flight Centre (GSFC), Greenbelt, MD-20771, USA

HIGHLIGHTS

- Cluster analysis of 7 years of global GOCART aerosol optical depth data.
- Global composition of ≥ 10 multiyear and seasonal aerosol mixture clusters.
- A taxonomy and visualization scheme for mapping aerosol mixtures.
- AERONET-derived optical and microphysical properties of aerosol mixtures.
- Spatiochemical assessment in the context of other classification studies.

ARTICLE INFO

Article history:

Received 3 February 2015

Received in revised form

12 June 2015

Accepted 16 June 2015

Available online 20 June 2015

ABSTRACT

The optical and microphysical characteristics of distinct aerosol types in the atmosphere are not yet specified at the level of detail required for climate forcing studies. What is even less well known are the characteristics of mixtures of aerosol and, in particular, their precise global spatial distribution. Here, cluster analysis is applied to seven years of 3-hourly, gridded $2.5^\circ \times 2^\circ$ aerosol optical depth data from the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model, one of the most-studied global simulations of aerosol type currently available, to construct a spatial partition of the globe into a finite number of aerosol mixtures. The optimal number of aerosol mixtures is obtained with a k-means algorithm with smart seeding in conjunction with a stopping condition based on applying the 'law of diminishing returns' to the norm of the Euclidean distance to provide upper and lower bounds on the number of clusters. Each cluster has a distinct composition in terms of the proportion of biomass burning, sulfate, dust and marine (sea salt) aerosol and this leads rather naturally to a taxonomy for labeling aerosol mixtures. In addition, the assignment of primary colors to constituent aerosol types enables true color-mixing and the production of easy-to-interpret maps of their distribution. The mean multiyear global partition as well as partitions deduced on the seasonal timescale are used to extract aerosol robotic network (AERONET) Level 2.0 Version 2 inversion products in each cluster for estimating the values of key optical and microphysical parameters to help characterize aerosol mixtures. On the multiyear timescale, the globe can be spatially partitioned into 10 distinct aerosol mixtures, with only marginally more variability on the seasonal timescale. In the context of the observational constraints and uncertainties associated with AERONET retrievals, bivariate analysis suggests that mixtures dominated by dust and marine aerosol can be detected with reference to their single scattering albedo and Angstrom exponent at visible wavelengths in conjunction with their fine mode fraction and sphericity. Existing multivariate approaches at classification appear to be more ambiguous. The approach presented here provides gridded ($1^\circ \times 1^\circ$) mean compositions of aerosol mixtures as well as tentative estimates of mean aerosol optical and microphysical parameters in planetary regions where AERONET sites do not yet exist. Spreadsheets of gridded cluster indices for multiyear and seasonal partitions are provided to

* Corresponding author.

E-mail address: mtaylor@noa.gr (M. Taylor).

facilitate further study of the global distribution of aerosol mixtures and possibly for the selection of new AERONET site locations.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

It is now known that aerosols contribute strongly to the change in radiative climate forcing and prediction uncertainty (Mishchenko et al., 2007a,b,c; Kahn, 2012; IPCC, 2013). However, much is still not known about the optical and physical properties of the aerosol components. For example, there is no suitable optical model for radiance retrievals of coarse-mode dust, even from the Sahara (e.g., Kalashnikova et al., 2005), and the time-varying, spectral single-scattering albedo of biomass burning particles is much too loosely constrained for many radiative forcing applications. Reduction of the aerosol uncertainty has been hindered by several things. Firstly, there is a gap in our detailed understanding of the global spatial and temporal variation of aerosol types (Hansen 2005), and secondly there is an absence of consensus on the optimal combination of optical and microphysical parameters that should be used for characterizing aerosol types (IPCC, 2013). Thirdly, satellite remote sensing data, while able to provide global spatiotemporal detail on aerosol amount and to some extent type, by itself it is insufficient for providing detailed aerosol microphysics.

Targeted suborbital measurements are expected to help to better constrain the uncertainty associated with direct aerosol radiative forcing (Kahn, 2012). Results from global circulation models (e.g. Mann et al., 2014) are helping to paint a picture of how aerosol is globally distributed (Stier et al., 2013; Kinne et al., 2013) and new studies of global decadal and/or multi-decadal trends in aerosol load are helping constrain its behavior in time (Zhang and Reid, 2010; Chin et al., 2014). It should be borne in mind that the convergence of results between IPCC/AR4 and AR5 in model intercomparison exercises suggests a reduction in model diversity but not necessarily in model uncertainty, due to the limitations of available observational constraints (Samset et al., 2014; Schwartz et al., 2014). Of note is the fact that there is a rapid growth in understanding 3D aerosol climatologies thanks to vertical profiles provided by LIDAR (e.g. Winker et al., 2013; Yu et al., 2010). These are likely to play a key role in comparative studies involving targeted suborbital measurements. With regard to the characterization of aerosol types, numerous studies have attempted to classify aerosol types from ground-based observations (e.g. Dubovik et al., 2002; Eck et al., 1999; Kalapureddy et al., 2009; Lee et al., 2010; Mielonen et al., 2009; Omar et al., 2005, 2009), and many satellite algorithms now adopt procedures for classifying aerosol types to improve the accuracy of their AOD retrieval (Higurashi and Nakajima, 2002; Jeong and Li, 2005; Kahn et al., 2010; Kaufman et al., 2005; Kim et al., 2007, 2013; Lee et al., 2009; Remer et al., 2005, 2008; Russell et al., 2014). Despite such studies at the local or regional level using bivariate or multivariate combinations of optical and/or microphysical parameters as proxies, there is no unambiguous protocol for the identifying aerosols of different types or mixtures. The latter is of particular importance and of increasing prominence due to cross-boundary and/or intercontinental transport of aerosol (e.g. Yu et al., 2013).

A proper characterization of aerosol requires knowledge of aerosol size, shape, and composition provided by optical and microphysical parameters (Dubovik and King, 2000). Inversion of radiation measurements made by ground-based CIMEL

sunphotometers (typically several per day) operated at sites contributing to AERONET (Holben et al., 1998) provide size, index-of-refraction, and shape information, but sites in this point-sampling network are unevenly distributed – particularly over ocean and uninhabited regions (e.g. deserts, ice caps oceans, and mountains). It should also be noted here that AERONET particle property retrievals are not as well-constrained as direct-sun spectral AOD measurements, and also that some of the assumptions in the sky scan retrieval are not physical (e.g. assuming that the indices of refraction are the same for both fine and coarse modes). The highest level of chemical detail is provided by surface or low altitude *in situ* measurements but laboratories and supersites tend to be located at urban centers and sample the atmosphere at altitudes well below the cloud layer (IPCC, 2013). Furthermore, there is a need for more studies comparing mass size distributions with columnar volume size distributions (e.g. Gerasopoulos et al., 2007) to test such assertions, especially considering that AERONET aerosol type has only been validated against *in situ* data in a few cases. From space, radiation measurements from spectroradiometer instruments onboard satellites (typically ≈ 1 per day) provide full-Earth viewing capacity but discerning aerosol types is still not unambiguous. Nevertheless, progress is being made in this direction. For example, inversion algorithms that retrieve aerosol properties from spectral multi-angle polarimetric satellite observations suggest that polarization information is expected to help with regard to determination of particle size and the real part of the refractive index with reference to volume size distributions (Dubovik et al., 2011), and some constraints on size, shape and also the SSA have been shown to be derivable from multi-angle, multispectral observations with MISR (Kalashnikova and Kahn, 2006; Chen et al., 2008; Kahn et al., 2010).

Given the current limitations in coverage and detail associated with using AERONET and satellite retrievals of aerosol type data, this paper presents a method for identifying, naming and visualizing global mixtures of aerosols from the output of global circulation or chemical transport models. The motivation for this is the work of Kahn et al. (2001) which analyzed the sensitivity of multi-angle imaging to AOD and aerosol type, based on global-scale model simulations of these quantities drawn from several sources. They produced a global climatology of likely aerosol mixtures by month from the model results, and showed the degree to which multiangle imaging is sensitive to natural mixtures of aerosols, particularly over the ocean. Here, we develop a methodology for spatially partitioning the globe into zones of aerosol mixtures of distinct characteristics. The partitioning is driven by the contribution of distinct tropospheric aerosol types to the AOD in each pixel of the global grid ($2.5^\circ \times 2^\circ$ longitude \times latitude) provided by the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model (Chin et al., 2000, 2002; 2014; Ginoux et al., 2001). The GOCART model provides global, continuous, gridded 3-hourly values of total AOD at 500 nm as well as the contribution to the total AOD of sulfate (SU), black carbon (BC), organic carbon (OC), desert (mineral) dust (DU) and sea salt (SS). With reference to the values of optical and microphysical parameters provided by AERONET sites in each cluster, we then assess the characteristics of the global partitioning obtained for a multiyear mean and for seasonal means of the GOCART chemical simulations in the context of

existing strategies for aerosol classification. In this way, a representation of the global spatial distribution, composition and other characteristics of aerosol mixtures are derived and assessed.

The paper is organized as follows. The model simulations and validation data used in this study are briefly presented in Section 2. Section 3 outlines the approach followed to cluster GOCART data to partition the globe spatially into aerosol mixtures as well as a new color-coding and a taxonomy for their labeling and visualization. In Section 4, the average composition of the clusters is presented for the whole (multiyear) period and for each season, and average values of optical and microphysical parameters extracted from available sunphotometer inversion products in each region are used to make an assessment of their characteristics and of several bivariate and multivariate classification schemes. The paper concludes in Section 5 with a summary of the main findings and suggestions for application of the results.

2. Data sources and uncertainties

GOCART (Collection G4P0) simulations spanning all days in the years 2000–2006 (inclusive) and gridded globally in pixels of size $2.5 \times 2^\circ$ were downloaded from the Goddard Earth Sciences, Data and Information Services Center (GES-DISC) interactive visualization and analysis portal at: <http://gdata1.sci.gsfc.nasa.gov>, and the percentage contribution of individual aerosol type to the total extinction AOD were derived for each pixel. The time-averaged global means were calculated, as were seasonal means based on extraction of all DJF (December, January and February), MAM (March, April and May), JJA (June, July and August) and SON (September, October and November) data. The methodology used to spatially partition the global multiyear and seasonal data is presented in Section 3.1. Sources of uncertainty in the GOCART model stem in part from uncertainty in the calculation of aerosol mass, associated with the emission inventories for individual aerosol types, chemical transport and precursors, and removal processes. In addition, uncertainties in the mass extinction efficiencies adopted are due to assumed particle hygroscopic properties, refractive indices, absorption, mixing state, injection heights (in particular for biomass burning smoke and dust), as well as assumptions of sphericity and size distribution of different aerosol types (Chin et al., 2002; Yu et al., 2010). In relation to biomass burning aerosol, the total dry mass burned in current source inventories is much too low (Chin et al., 2009), and the commonly applied assumption that smoke emissions are all uniformly distributed in the atmospheric boundary layer has also been called into question (Kahn et al., 2008; Val Martin et al., 2010). Another key issue is that the assumption in the Mie theory used to calculate particle optical properties is that particles are spheres; this assumption has been shown to lead to severe under-estimation of aerosol extinction and backscatter in the case of non-spherical particles (Mattis et al., 2002). The global partitions of aerosol mixtures based on cluster analysis of GOCART described in Section 3 should be seen in the context of these uncertainties.

With a view to characterizing the optical and microphysical properties and general characteristics of aerosol mixtures and distinct zones of the partitioning described in Section 3, the global record of AERONET (Holben et al., 1998) “All Points” Level 2.0 Version 2 inversion products was downloaded from <http://aeronet.gsfc.nasa.gov> and temporally aligned with the multiyear period spanned by the GOCART record (2000–2006 inclusive). These products are cloud-screened and quality assured. The accuracy of these products is reported by Dubovik et al. (2000, 2002, 2006). In addition, the global record of AERONET “All Points” Level 1.5 inversion products was obtained to address situations where Level 2.0 Version 2 records were missing (see Section 4.2). For both Level

1.5 and Level 2.0 records, the phase function at 180° for the central spectral wavelengths (440, 675, 870 and 1020 nm) was also appended (see Section 4.1 for details of its application to the estimation of LIDAR ratios). Finally, we note that, in order to facilitate future studies of co-located GOCART and/or AERONET data with satellite data gridded at 1×1 degrees, the GOCART data was downsampled from $2.5^\circ \times 2^\circ$ to $1^\circ \times 1^\circ$ pixel resolution using a simple cubic spline interpolation. We would like to reiterate here that, as the data quality for aerosol-type inversion products is very different from direct-sun AOD measurements, work still needs to be done to validate AERONET aerosol type products and further assess their uncertainties. We will return to this point in more detail in the results section where the optical and microphysical characteristics of global aerosol mixtures derived from AERONET inversion products are presented.

3. Methodology

Global time-averaged mean maps are first calculated for the total AOD and for the BC, OC, SU, DU and SS contribution on i) the multiyear timescale (01/01/2000–31/12/2006) and ii) the seasonal timescales DJF, MAM, JJA and SON. The ratio of the contribution of the aerosol type to the total AOD was calculated as a percentage in each pixel and the maps were then ‘unfolded’ and combined in serial form with data in each pixel occupying a row of the resultant $64,800 \times 8$ matrix (column 1 = longitude, column 2 = latitude and columns 3:6 = percentage contribution of BB = BC + OC, SU, DU and SS. Note that we have adopted a convention also used by GEOS-5 whereby combined aerosol maps produced from GOCART data are constructed by summing the percentages of BC and OC labeling the result as ‘BB’ (corresponding to ‘Biomass Burning’). Figs. 1 and 2 show the global spatial distribution of the mean total AOD (500 nm) on the multiyear and seasonal timescales, respectively. Fig. 1 shows that the regions of peak mean global aerosol load are broadly located over the Sahara and Nigeria in Northern Africa, in the Gobi desert to the north of the Himalayas, and over a large region centered on Beijing.

The maps of Fig. 2 suggest that the spatial distribution of the mean total AOD does not vary strongly even on the seasonal timescale. Furthermore, the main sources of aerosol, while exhibiting some seasonal variation in intensity, do not show large displacement in the location of their peaks. Note that such findings should be seen within the context of the underlying assumptions and limitations of the GOCART model itself, that lead to uncertainty in the AOD retrievals for different aerosol types. Some of these have been described in Section 2. In particular, dust and especially biomass burning show large inter-annual AOD variations on the seasonal timescale at the regional level (Chin et al., 2009, 2014), and these effects are averaged-out in Figs. 1 and 2. A further limitation is that the GOCART model itself may not fully capture the magnitude of the regional, seasonal inter-annual variability.

In order to investigate how representative the 2000–2006 (inclusive) mean is of variability in the annual mean over the study period, we extracted gridded maps of the mean annual AOD and performed a quantitative comparison of each yearly map with the 2000–2006 mean. Since maps of AOD are effectively grey-scale images (i.e. color is used only to represent magnitude), the mean bias and root-mean squared error (RMSE) can be calculated directly by averaging differences over all 64,800 pixels. Note that the RMSE, although a good quantitative measure of magnitude, is often a poor measure of feature differences (e.g., Wang et al., 2004). Hence, we also calculated the value of the mean structural similarity index (SSIM) (Wang et al., 2004). The results of this quantitative comparison between the annual mean and the 2000–2006 mean for each year of the study period are collected in Table 1.

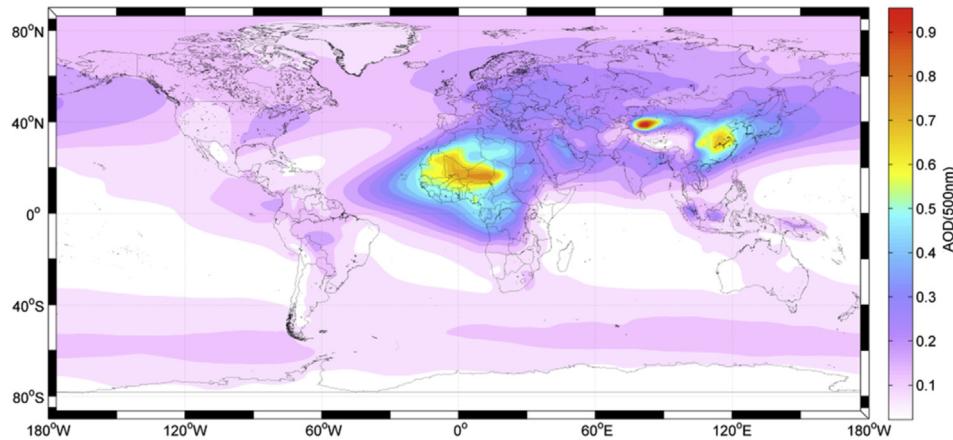


Fig. 1. The multiyear (2000–2006 inclusive) mean AOD (500 nm) from the GOCART model.

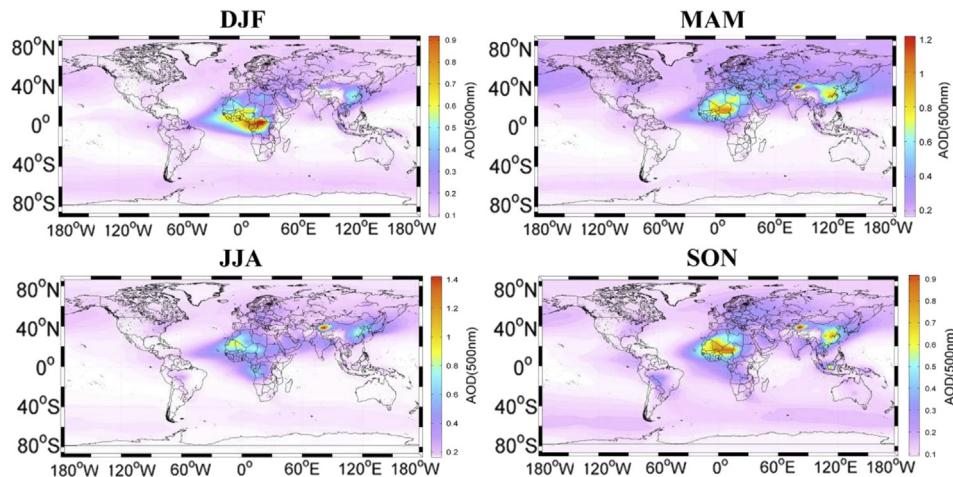


Fig. 2. The seasonal mean AOD (500 nm) from the GOCART model for the monthly triplets: DJF, MAM, JJA and SON.

Table 1

Statistics resulting from a comparison of the gridded ($1^\circ \times 1^\circ$) mean total global AOD (500 nm) map for the period 2000–2006 (inclusive) with the analogous annual mean maps. ‘Bias’ is the mean bias, ‘RMSE’ is the root mean squared error and ‘SSIM’ is the mean structural similarity index calculated over all 64800 gridded pixels.

Year	Bias	RMSE	SSIM
2000	0.0151	0.0520	0.999525
2001	-0.0028	0.0193	0.999949
2002	0.0022	0.0102	0.999985
2003	0.0037	0.0098	0.999985
2004	-0.0053	0.0102	0.999980
2005	0.0038	0.0121	0.999979
2006	0.0006	0.0159	0.999969

There is strong agreement in the structural similarity between maps of annual means and the 2000–2006 mean (perfect agreement is signified by a value of SSIM = 1). There is some variation in the mean bias and the RMSE, both of which are unitless, as is AOD. In particular, the year 2000 mean shows the strongest deviation from the 2000–2006 mean, with both a bias and a RMSE ≈ 3 times the maximum bias and RMSE obtained for other years in the study period. In the context of the accuracy of the AERONET retrieval of AOD, which is ± 0.02 (Holben et al., 2006) for the Level 1.5 inversion products and ± 0.01 (Eck et al., 1999) for the Level 2.0 inversion products, the calculated mean pixel bias and RMSE are well within

these accuracy limits for all annual means considered. In Figs. S1–S7 of the supplement accompanying this manuscript, maps of the mean total global AOD for the years 2000 through 2006 (inclusive), the 2000–2006 mean, calculated annual differences from the multiyear mean, and a log-density regression (over all pixels) of annual mean data versus the multiyear mean are presented. Based on these findings, we proceeded to use the multiyear 2000–2006 mean map as a proxy for the annual mean AOD maps. Note that implicit in this assumption is that the mean percentage contribution of aerosol types has similarly low interannual variability. In order to capture also the variation in individual aerosol components, results for seasonal means will also be presented in full. Fig. 3 presents the mean percentage contribution of these types to the total AOD for the global multiyear mean.

Fig. 3 shows that, based on the model simulations, the largest contributions to the mean global AOD are from dust and sulfate (whose color bar axes extend to nearly 100%). Dust is seen to occupy an extended (and slightly eastwardly-inclined) band of around 30 degrees of latitude, and 180° longitude in extent, situated between 10° and 40°N. Marine sea salt aerosol appears to be confined strongly to the Southern hemisphere and its peak occupies a band about 20°–30° of latitude, centered around 60°S, and extending across the globe in longitude. Peak contributions of biomass burning products (BB = BC + OC) occur as expected over the Amazon forest, in the African Savannah and the forests of

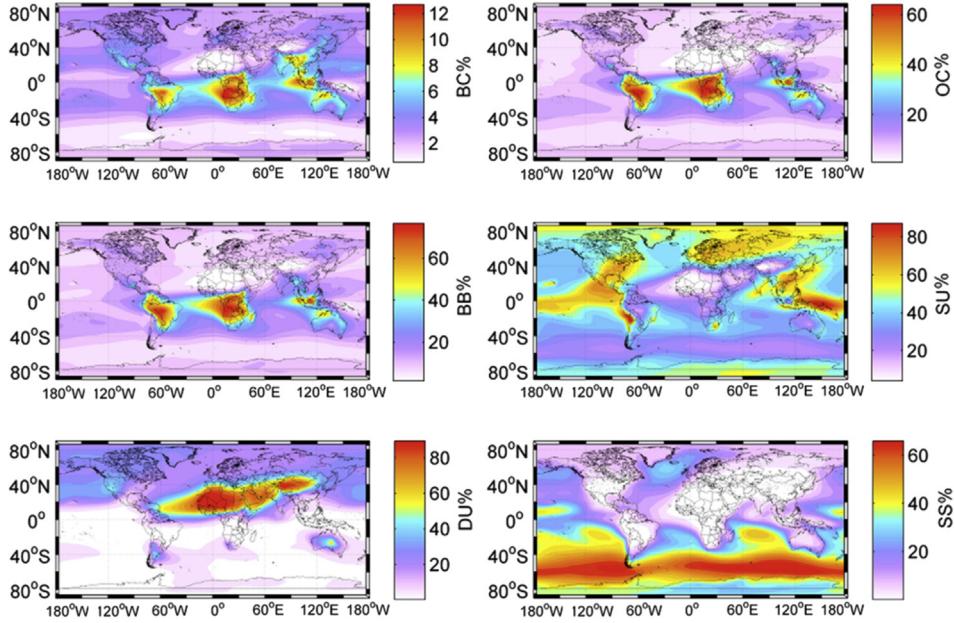


Fig. 3. The multiyear (2000–2006 inclusive) mean percentage contribution of BC, OC, BB = BC + OC, SU, DU and SS to the total AOD (500 nm) derived from GOCART model output data. Color axes are normalized to the maximum value in each case. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Borneo, Indonesia and North Western Australia. On this timescale, the Sahel does not appear as a major BB source. The contribution of sulfate is more dispersed. Strong peaks are observed over Indonesia's "Ring of Fire" and Santiago (Chile), but prominent and extensive contributions are clearly visible in the regions centered on Mexico City, the North East coast of the USA, Japan, the Eastern coast of China, as well as over the whole of the European continent (with the exception of the southern periphery).

3.1. Cluster analysis

The next task was to perform cluster analysis on the percentage contributions of each aerosol type for the global mean map calculated for the multiyear timescale and the seasonal timescale. A k-means clustering algorithm (MacQueen, 1967) calculated the norm of the Euclidean distances (which we will refer to as an 'energy' to be minimized) from cluster centers to every point in the 4D space of percentage contributions. To mitigate the effect of centers being confined to local minima, initial cluster centers were chosen randomly during multiple (we used 10 here) independent runs (for the same number of clusters) as per the prescription of David and Vassilvitskii (2007). The lowest energy case of these 'random seeds' was then retained as the best case. This entire procedure was repeated stepping through 1 to 24 cluster centers. The variation of the energy (E) with the number (N) of clusters is shown in Fig. 4. The global minima resulting from the random seeding process were found to be repeatable and robust. As an indication, the percentage difference in the Euclidean distance for each random seeding relative to that associated with the best seed never exceeded 5.6% for the case of $N = 10$ clusters. Furthermore, six out of ten of the seeds had a percentage difference of less than 1% from the optimal global minimum found.

Fig. 4 shows that this approach leads to a smoothly decaying variation of E with N . In order to identify the optimal number of clusters, we applied two methods. In the first method, as per Omar et al. (2005), the optimal number of clusters was detected with the condition that the change in energy ('delta E ' in the plot) fell below 10% of the initial energy value (for the case $N = 1$). For the multiyear

global mean data, this condition gives $N = 10$ clusters. This choice of stopping condition (10%), although subjective and on the conservative side, produces a global partition into spatial zones whose dominant aerosol types and/or mixtures are reflected by other studies in these regions. Actual retrieval sensitivity for MISR, to take that example, generally requires about 20% of AOD for a single component to be at all detectable, under good retrieval conditions (Kahn et al., 2001).

As an independent check on the consistency and potential validity of this result, we devised the following method for estimating upper and lower bounds on the optimal number of clusters obtained. When plotted on log–log (natural logarithm) axes, the curve of E versus N was found to be closely fit by a straight line (especially for $N > 4$) — signaling that the curve follows an exponential decay. The linear fit in log–log space was used to find the location of the integer values of N corresponding to consecutive halvings of E . In this space, points associated with the number of 'half-lives' (p) of E lie at equally-spaced distances along the best fit line as shown by the red triangles in the inlay in Fig. 4. In order to find lower and upper bounds for the optimal number of clusters, points having coordinates $\ln(N)$ and $\ln(E)$ closest to each value of p were used. For $p = 3$ half-lives, $N = 7$ (i.e. closest to the 7th black point along the best fit line in the inlay of Fig. 4) provides the lower bound and, for $p = 4$ half-lives, $N = 13$ (i.e. closest to the 13th black point along the best fit line in the inlay of Fig. 4) provides the upper bound. This method suggests that the optimal number of clusters N is between 7 and 13, in logical agreement with the result obtained using the first method where $N = 10$ when the 'delta $E < 10\%$ ' stopping condition was used. These findings are in general agreement with the study of the sensitivity of multiangle imaging to natural mixtures of aerosol in Kahn et al. (2001) where 13 clusters were found for up-to-four component mixtures and 5 clusters were obtained when differences in components but not their proportions were analyzed. With the stopping condition in mind, we adopted $N = 10$ as the "optimal" number of clusters. This approach was then repeated for the global mean maps unfolded for seasonal data. In Fig. S9 of the supplement accompanying this manuscript, plots of the process for each season (DJF, MAM, JJA and SON) are

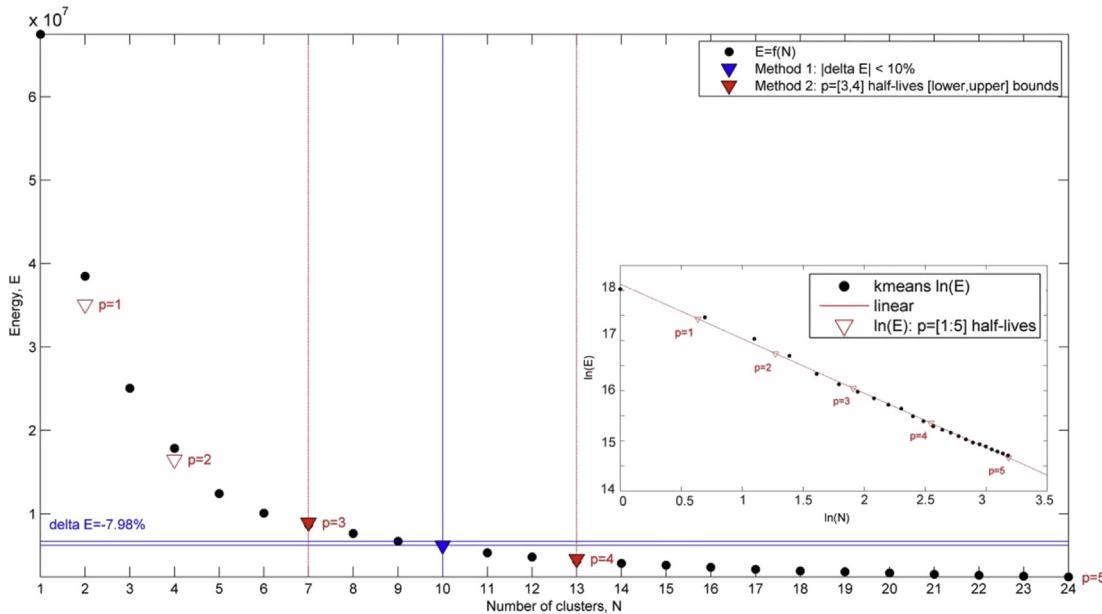


Fig. 4. Graph of the norm of the Euclidean distance ('energy', E) with the number (N) of clusters (1–24) for the multiyear mean percentage contribution of BB, SU, DU and SS to the total AOD (500 nm) with points shown as black dots. Red triangles indicate the values of (N, E) associated with 'half-lives' $p = [1:5]$ obtained by linear regression of the plot of E versus N in natural logarithm space (inlay). Blue horizontal lines indicate the point where ΔE drops below 10% (at $N = 10$) while red vertical lines indicate the lower and upper bounds on the optimal number of clusters. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

presented. During the season spanned by the months DJF, the optimal number of clusters is commensurate with that obtained for the multiyear mean ($N = 10$). For the other seasons MAM, JJA and SON, the "optimal" number of clusters is $N = 11$ suggesting perhaps an increase in the diversity of global aerosol mixtures during these months as compared to DJF. Once again, in the context of the stopping condition, it is unlikely that the difference between $N = 10$ and $N = 11$ is significant.

3.2. Taxonomy of aerosol mixtures

The result of applying the methodology outlined in Section 3.1 to the unfolded map of the multiyear mean and unfolded seasonal maps was a vector of cluster indices (ranging from 1 to N, the number of clusters obtained with the stopping condition) in each case. Each row of this vector assigns each pixel to a distinct cluster and has a direct 1:1 mapping with the matrix of unfolded aerosol type percentages (i.e. each pixel has a specific and distinct proportion of BB, SU, DU and SS). Isolating all instances of the same cluster index then allowed for descriptive statistics (central values and dispersion) derived from the percentage contributions of BB, SU, DU and SS to be produced for each cluster. Table 2 presents the results of this analysis for the multiyear mean global partition. Tables presenting full descriptive statistics (including also the median and the inter-quartile range) are presented in the supplement for the multiyear mean (Table S1) and for seasonal mean global partitions (Tables S3–S6). Note that the statistics reported represent the consistency of the model rather than its accuracy *per se*.

In this work then, each cluster corresponds to a mean aerosol mixture – containing a combination of BB, SU, DU and SS. When visualizing the clusters, we decided to follow as closely as possible the color scheme used for aerosol types by the CALIPSO LIDAR aerosol retrieval algorithm: black = smoke ('BB'), red = pollution ('SU'), yellow = dust ('DU') and blue = marine ('SS') with the closest parallel possible with the aerosol typing adopted by GOCART. This scheme has the desirable property that these are primary colors

and are visually easy to distinguish (i.e. they do not create ambiguities in perception due to color-blindness for example). With this color-coding, the cluster analysis gives rise to the following stacked bar chart of mean aerosol mixtures shown in Fig. 5 for the multiyear global means:

Note that cluster 4 is DU-dominated, cluster 5 is BB-dominated, cluster 8 is SU-dominated and cluster 9 is SS-dominated. All other clusters (1, 2, 3, 6, 7 and 10) are not clearly dominated by a single aerosol type but have at least one component in excess of ~20%. This leads fairly naturally to a naming convention whereby the dominant component can be thought of as the 'object' and the sub-dominant components as descriptive 'adjectives' (that can be logically ordered in increasing order of size) in the 'ordered adjective-object' taxonomy shown in Table 3. The result is five "Sulfate" clusters, two each of "Smoke" and "Dust," and one "Marine."

Stacked bar charts of mean aerosol mixtures obtained at the seasonal timescale as well as the resulting taxonomy are provided in Table S2 and Fig. S8 of the supplement. In the 4-component analysis performed by Kahn et al. (2001), clusters were associated with either dominant regions or seasons. Here, and without going into too much detail, we present a brief comparison of several of the largest clusters to assess the consistency in the outputs of the two models. Table 2 of Kahn et al. (2001) presents 5 climatological mixtures (as percentages of SU, SS, BC, OC and fine and coarse DU) classified into the following groups:

- 1) "Carbonaceous + Dusty Maritime"
- 2) "Dusty Maritime + Coarse Dust"
- 3) "Carbonaceous + Black Carbon Maritime"
- 4) "Carbonaceous + Dusty Continental"
- 5) "Carbonaceous + Black Carbon Continental"

with sub-divisions that depend on region and/or season. Group (1) has the highest global voxel occupy (34%). The composition of the southern mid-latitude ocean group (1c) is: 40%SU/32%SS/17%BB and 11%DU. The highest occupancy analog in our study is the "Marine SULFATE" (cluster 2) which has a somewhat lower

Table 2

Descriptive statistics of multiyear mean global aerosol mixtures extracted from the GOCART data record by application of cluster analysis.

GOCART V4	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Pixels	1187	16,020	15,486	2486	2019	2645	6958	4153	10,338	3508
AOD (500)	0.17	0.06	0.16	0.40	0.19	0.20	0.14	0.10	0.10	0.09
%BB (Mean)	35.9	14.0	12.5	5.7	61.2	11.2	10.4	15.0	9.0	34.2
%SU (Mean)	27.4	44.6	54.7	13.0	29.7	31.0	43.1	66.1	28.8	45.0
%DU (Mean)	30.7	4.7	25.6	80.2	6.0	53.1	27.0	4.7	3.8	6.8
%SS (Mean)	5.9	36.7	7.2	1.1	3.1	4.7	19.5	14.1	58.4	14.0
%BC (Mean)	5.9	3.0	3.4	1.7	9.3	3.2	3.2	3.4	1.7	6.7
%OC (Mean)	30.1	11.0	9.1	4.0	51.8	8.0	7.2	11.6	7.3	27.5
%BB (St.D.)	7.4	4.1	3.1	3.7	8.5	5.5	2.7	5.2	2.4	6.5
%SU (St.D.)	5.7	6.3	4.9	5.5	6.6	7.9	4.0	7.1	4.1	6.6
%DU (St.D.)	8.2	2.2	4.6	8.1	5.2	7.8	5.6	4.3	1.4	4.7
%SS (St.D.)	3.3	5.4	3.2	1.9	3.2	4.5	4.8	8.5	6.4	7.8
%BC (St.D.)	1.2	1.0	0.9	0.8	1.8	1.2	0.7	1.3	0.6	1.2
%OC (St.D.)	6.8	3.2	2.5	3.1	7.4	4.8	2.1	4.3	1.9	5.8

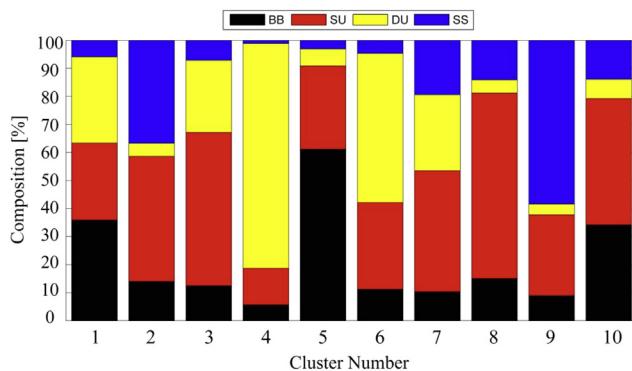


Fig. 5. A (primary) color-coded stacked bar chart displaying the aerosol composition of each cluster resulting from application of the k-means algorithm and the stopping condition to the global multiyear mean GOCART data. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 3

Taxonomy of aerosol mixtures for the multiyear global mean partition using the ordered adjective-object naming convention. The ‘object’ is highlighted in bold capital font and is the dominant aerosol type. The ‘adjectives’ preceding it are in increasing order of percentage contribution.

Cluster 1	Sulfurous dusty SMOKE
Cluster 2	Marine SULFATE
Cluster 3	Dusty SULFATE
Cluster 4	DUST
Cluster 5	Sulfurous SMOKE
Cluster 6	Sulfurous DUST
Cluster 7	Marine dusty SULFATE
Cluster 8	SULFATE
Cluster 9	Sulfurous MARINE
Cluster 10	Smokey SULFATE

occupancy (24.7%) but is located in the same region (see next Section) and has a very similar composition: 44.6%SU/36.7%SS/14.0%BB and 4.7%DU. By the same logic, (group 4a) occupies 21% of the voxels with a composition: 61%SU/0%SS/13%BB and 26%DU. The analogous cluster obtained here is “Dusty SULFATE” (cluster 3) which has a similar occupancy (23.9%) and composition: 54.7%SU/7.2%SS/12.5%BB and 25.6%DU to within 7–8%. These two examples are indicative of quite good quantitative agreement and suggest that there is some consistency between the two modeling approaches, giving credence to the actual clusters found in both modeling approaches. In Section 4, we perform an additional assessment of the clusters found here against classes obtained by

cluster analysis of AERONET sunphotometer data by Omar et al. (2005, 2009).

3.3. Visualization of aerosol mixtures

Having identified the composition of global mean aerosol mixtures on the multiyear and seasonal timescales, we now turn to visualizing the clusters as partitions of gridded ($1^\circ \times 1^\circ$) global maps to see how they are spatially distributed worldwide. A challenge here again is the use of color, and in particular, the creation of a color scheme for representing mixtures of aerosols. In order to retain a visual link with the colors assigned to pure aerosol types (BB = black, SU = red, DU = yellow and SS = blue) for each cluster, we simply mixed these primary colors in accordance with the percentage of each aerosol type assigned to each cluster (this is valid as the percentage contributions sum to 100%). As a further visual point of reference, we generated and added color keys to the resultant maps corresponding to i) pure types (e.g. 100% DU), ii) 50%:50% mixtures of any 2 pure types, and iii) the actual compositional color-mix of the 4 aerosol types BB, SU, DU and SS for each cluster. Fig. 6 illustrates this approach in the context of incrementally increasing the number (N) of clusters from N = 1 to N = 10 (see Section 3.1) for the case of the multiyear global mean.

4. Results

4.1. Global mean aerosol mixtures on the multiyear and seasonal timescales

The global distribution of aerosol mixtures for the multiyear global mean, where the optimal number of clusters is N = 10, can be seen in the bottom right panel of Fig. 6. Table 4 presents the number of pixels and the fraction of the global grid ($1^\circ \times 1^\circ$) occupied by each cluster. Cluster 2 ('Marine SULFATE') and Cluster 3 ('Dusty SULFATE') have the largest spatial extent and occupy the greatest number of pixels (24.7% and 23.9% of the global grid respectively). Furthermore, Cluster 2 and Cluster 9 ('Sulphurous MARINE') together span the southern oceans below about 15°S and occupy 40.7% of the grid, whereas Cluster 3 ('Dusty SULFATE') and Cluster 7 ('Marine Dusty SULFATE') together span nearly the entire northern hemispheric region (above about 15°N) with a combined occupancy of 34.6%.

With regard to aerosol mixtures obtained for seasonal means, it can be seen in Fig. 7 that the salient features of their spatial distribution are largely similar. Note that these maps represent the mixture of aerosol types and not amount (for this see Fig. 2), which varies seasonally to a much greater extent.

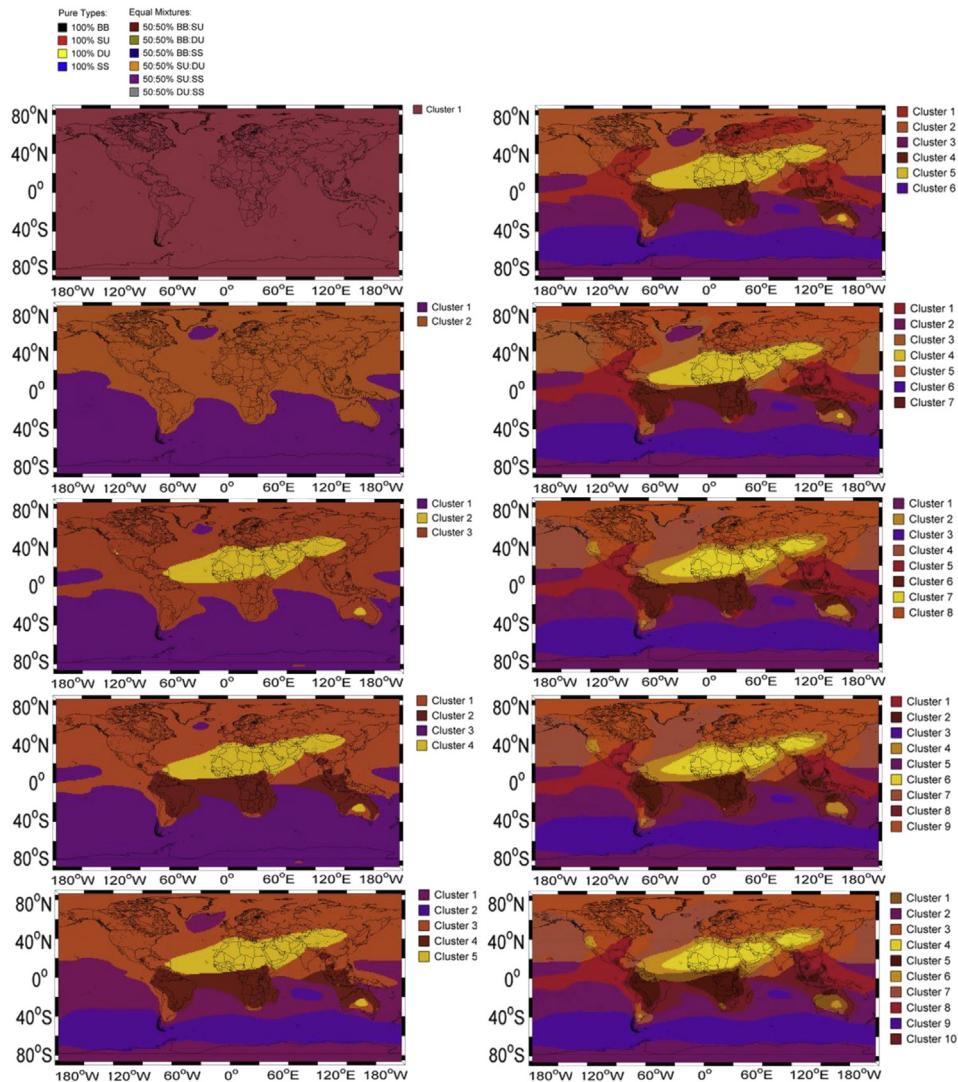


Fig. 6. The evolution of the global partition with the number of clusters for the case of the multiyear global mean. The spatial distribution of aerosol mixtures results from application of the k-means clustering algorithm to the multiyear mean of global GOCART chemical data spanning the period 2000–2006 (inclusive) for a different number (N) of clusters in each case. Note that colors are produced by mixing black, red, yellow and blue in direct proportion to the percentage contribution of each pure aerosol type (BB, SU, DU, and SS) in each cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 4

The number of pixels and the fraction of the globe occupied by each cluster for the partitioning of the multiyear mean and the seasonal means.

	2000–2006		DJF		MAM		JJA		SON	
	Pixels	Fraction	Pixels	Fraction	Pixels	Fraction	Pixels	Fraction	Pixels	Fraction
Cluster 1	1187	1.8%	2128	3.3%	5251	8.1%	8887	13.7%	8532	13.2%
Cluster 2	16,020	24.7%	10,842	16.7%	8592	13.3%	12,154	18.8%	6154	9.5%
Cluster 3	15,486	23.9%	12,092	18.7%	4296	6.6%	10,096	15.6%	8472	13.1%
Cluster 4	2486	3.8%	11,315	17.5%	2481	3.8%	1894	2.9%	15,186	23.4%
Cluster 5	2019	3.1%	2016	3.1%	9099	14.0%	3194	4.9%	6949	10.7%
Cluster 6	2645	4.1%	12,251	18.9%	7172	11.1%	3010	4.6%	2947	4.5%
Cluster 7	6958	10.7%	2970	4.6%	8513	13.1%	6280	9.7%	4785	7.4%
Cluster 8	4153	6.4%	5495	8.5%	3690	5.7%	2784	4.3%	3349	5.2%
Cluster 9	10,338	16.0%	4043	6.2%	2488	3.8%	6014	9.3%	2384	3.7%
Cluster 10	3508	5.4%	1648	2.5%	9294	14.3%	3063	4.7%	3665	5.7%
Cluster 11					3924	6.1%	7424	11.5%	2377	3.7%
Total	64,800	100.0%	64,800	100.0%	64,800	100.0%	64,800	100.0%	64,800	100.0%

These maps echo several key features (including the location and extent of dust-dominated aerosols and biomass burning products) of the cluster analysis of aggregate AOD data for an earlier

time period (1980–2000) drawn from 5 aerosol transport models in Kahn et al. (2001) for the months of January, April, July, and October (Plate 1 of Kahn et al., 2001).

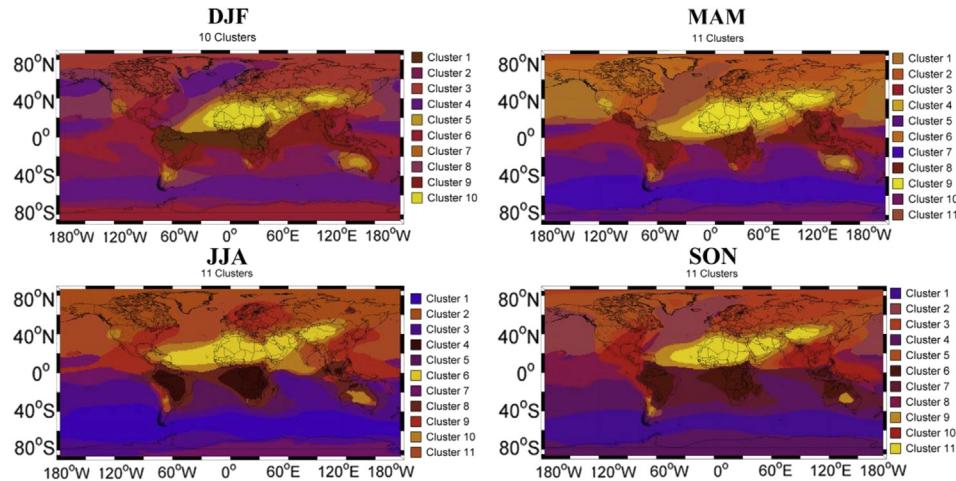


Fig. 7. The spatial distribution of aerosol mixtures resulting from application of the k-means clustering algorithm to the seasonal mean of global GOCART chemical data for the monthly triplets: DJF, MAM, JJA and SON.

The spatial extent (in pixels and as a percentage of the globe) of aerosol mixtures obtained for the seasonal means in shown in **Table 4**. Note that the composition of each aerosol mixture (cluster) for the multiyear mean are given by **Table 2**, whereas the composition of each aerosol mixture for the seasonal means are given by **Tables S3–S6** in the Supplement.

In **Table 5**, we have extracted the compositional data for clusters having the same taxonomical label for the multiyear mean and seasonal mean partitions. This has been done to provide an initial assessment of the variation of the composition of such clusters on the seasonal timescale in comparison with those captured by the multiyear mean partition.

Without going into too much analytical detail here, we draw the reader's attention to the 'DUST' and 'Sulfurous DUST' clusters as an illustration. For the multiyear mean partition, the 'DUST' cluster has DU = 80.2%. The corresponding cluster in the seasonal mean partitions has percentage contributions: DU = 80.7%, 85.6%, 76.7% and 80.3% for DJF, MAM, JJA and SON respectively – with a peak in MAM. The 'Sulfurous DUST' cluster of the multiyear mean partition has components: DU = 53.1% & SU = 31.0%. The corresponding cluster in the seasonal mean partitions has percentage components: (DU,SU)=(56.1%,24.5%), (62.5%,24.7%), (46.6%,32.9%) and (51.2%,30.5%) for DJF, MAM, JJA and SON respectively. These findings are qualitatively consistent with seasonal trends for dust and polluted dust aerosol in the literature.

In order to facilitate the uptake of the findings of this work, spreadsheets (EXCEL) of gridded ($1^\circ \times 1^\circ$) cluster indices for the multiyear mean global aerosol mixtures and seasonal mean (DJF, MAM, JJA and SON) global aerosol mixtures are freely available at: <http://apcg.meteo.noa.gr/aeromap/> together with tables presenting descriptive statistics (mean, standard deviation, median and interquartile range) of the 5-component (BC, OC, SU, DU and SS) percentage composition for each cluster of both the multiyear mean

global partition and the seasonal mean global partitions: DJF, MAM, JJA and SON. Interested readers wishing to perform a comparative analysis of the global partitions obtained here with those of Kahn et al. (2001) can download the "ClimLikely" dataset of the latter at the NASA Langley ASDC: https://eosweb.larc.nasa.gov/cgi-bin/misr_tools/clim_likely.cgi.

4.2. Optical and microphysical characteristics of mean global aerosol mixtures

In order to characterize aerosol mixtures, gridded aerosol mixture cluster indices were used to co-locate and extract inversion products (optical and microphysical aerosol parameters) derived from AERONET sunphotometer data for each partition. **Fig. 8** shows four individual clusters (the dust-dominated clusters 4 and 6 and the smoke-dominated clusters 5 and 10) for the case of the global multiyear mean overlaid with points indicating the locations of AERONET sites contributing Level 2.0 Version 2 inversion products.

In what follows we present mean values of optical and microphysical aerosol parameters extracted from the AERONET inversion data record for each cluster of the multiyear mean global aerosol partition. Spectral parameters retrieved include: the aerosol optical depth (AOD), the asymmetry factor (ASYM), the phase function at 180° ($P(180)$), the absorption aerosol optical depth (AAOD), the single scattering albedo (SSA), the real and imaginary parts of the complex refractive index (CRI-R and CRI-I respectively), and the LIDAR ratio (LR) retrieved at the central wavelengths 440, 675, 870 and 1020 nm. Although not provided directly by AERONET's inversion algorithm, spectral values of the LR were appended to the extracted dataset using the empirical relationship (Boyouk et al., 2010):

Table 5
The percentage contribution of BB/SU/DU/SS to the total AOD (500 nm) for the multiyear mean (shown in **Fig. 5**) and the seasonal means for clusters in each partition having the same label according to the taxonomy of Section 3.2.

Common mixture	Multiyear	DJF	MAM	JJA	SON
DUST	5.7/13.0/80.2/1.1	7.8/10.5/80.7/1.1	4.2/9.5/85.6/0.7	5.0/16.5/76.7/1.8	5.2/13.3/80.3/1.1
Sulfurous DUST	11.2/31.0/53.1/4.7	15.0/24.5/56.1/4.5	9.6/24.7/62.5/3.2	14.7/32.9/46.6/5.8	14.2/30.5/51.2/4.1
Dusty SULFATE	12.5/54.7/25.6/7.2	7.5/58.0/18.2/16.3	9.5/55.8/28.7/6.1	12.7/52.1/25.1/10.1	13.4/52.2/26.9/7.5
SULFATE	15.0/66.1/4.7/14.1	12.6/64.7/5.7/16.9	16.4/66.5/4.3/12.8	14.7/67.0/11.1/7.3	17.6/71.0/4.6/6.8
Marine SULFATE	14.0/44.6/4.7/36.7	10.9/50.9/5.9/32.3	9.2/49.8/4.1/36.9	8.9/49.0/2.8/39.2	13.2/57.2/2.8/26.7
Sulfurous MARINE	9.0/28.8/3.8/58.4	6.5/37.9/6.4/49.2	3.5/24.8/2.8/68.9	11.8/32.0/2.8/53.4	15.2/24.2/4.4/56.2

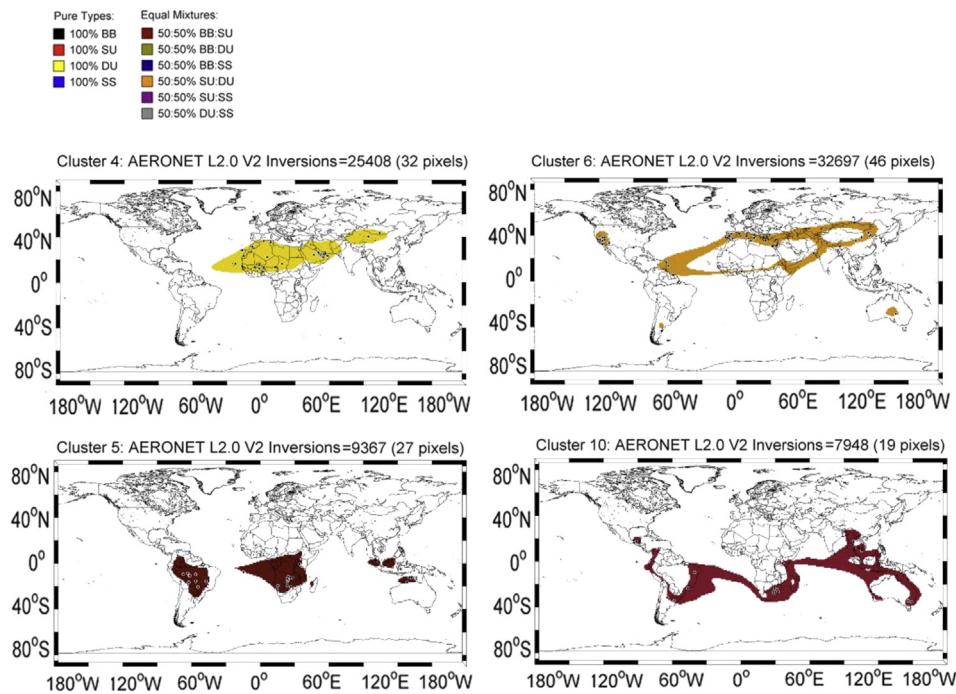


Fig. 8. Maps showing the spatial distribution of 4 aerosol mixtures associated with clusters 4, 5, 6 and 10 colored according to their compositional mixture, with AERONET sites contributing Level 2.0 Version 2 inversion products within each cluster shown as small circles. The total number of complete L2.0 V2 inversion records and the number of gridded $1^\circ \times 1^\circ$ pixels are also shown. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

$$LR(\lambda) = \frac{4\pi}{SSA(\lambda) \times P(180^\circ, \lambda)} \quad (1)$$

and the Angstrom Exponent AE(440/675) was calculated linearly from the AOD at these two wavelengths in log-space. Microphysical parameters retrieved include: the mean percentage of spherical particles (% Sphericity), the aerosol volume size distribution (AVSD) in 22 equally-spaced logarithmic radial bins from 0.05 to 15 μm , and AVSD-derived parameters related to fine ('f') and coarse ('c') modes: geometric radii (r_f and r_c), geometric standard deviations (σ_f and σ_c), volume concentrations (V_f and V_c), and the fine mode fraction $\eta = V_f/(V_f + V_c)$ which is a measurement of the proportion of fine mode aerosols to the total by volume (not number). For details of the calculation of AVSD-derived parameters, we refer the reader to Appendix A of Taylor et al. (2014). Regarding the accuracy of the AERONET sky scan retrievals, the overall uncertainty in AOD (under cloud-free conditions) is ± 0.01 for wavelengths longer than 440 nm (Dubovik et al., 2000) and the error in aerosol AVSD is estimated to be <10% for particle radii in the region of fine and coarse mode peaks in the radial interval between 0.1 μm and 7 μm (Dubovik et al., 2002). It should also be borne in mind that the retrieval (especially of the SSA based on fine and coarse modes) is implemented under the assumption that the CRI is the same for all particle sizes (Holben et al., 2006); which is not suitable for physical interpretations. Furthermore, the quality of all Level 2 inversion products is maximized at large solar zenith angles (>50°). Retrievals of the SSA and CRI in particular require large aerosol loads where AOD (440 nm) >0.4 and four coincident spectral measurements at 440, 675, 870 and 1020 nm. Aerosol loading is also important for accurate determination of % Sphericity which requires AOD (440 nm) >0.2 (Dubovik et al., 2006; Holben et al., 2006). Table 6 below presents the optical and microphysical parameter means for each of the 10 clusters of the multiyear partition.

Although some 588 Level 2.0 records were extracted for cluster 2, there were no records containing data for AAOD, SSA, LR and the

CRI at this level of quality assurance. In order to estimate the values of these important parameters, we extracted a total of 5830 analogous records from the Level 1.5 dataset at the sites: Amsterdam Island, Dunedin, Guam, Rottnest Island and Tahiti. For cluster 9 only, 10 Level 2.0 records were available, so we extracted 378 complete records (all inversion products) from Level 1.5 data from Crozet Island to calculate mean values of all parameters. Tables S7–S10 of the supplement provide analogous data for the seasonal means (DJF, MAM, JJA and SON).

Below we present in graphical form, key trends in the full set of inversion products extracted parameters. Although a detailed comparison with the literature is beyond the scope of the present work, we highlight key results and refer the reader to relevant studies. Fig. 9 shows the variation of AOD with wavelength over the range 300–1100 nm for each cluster together with a least squares fit using cubic splines. The mean GOCART AOD (500 nm) is also plotted for comparison purposes.

Two features stand out. Firstly, with the exception of clusters 2, 5 and 10, the GOCART AOD (500 nm) is much higher than the mean AOD (500 nm) extracted from AERONET retrievals in the same cluster domains. This is likely due to the fact that the spatial sampling of the domains by AERONET sites is under-representing the mean AOD at 500 nm over each of these regions. In the case of cluster 2 ('Marine SULFATE'), there is excellent agreement even though AERONET sites occupy tiny fractions of only 5 pixels. High BB component Clusters 5 and 10 show a reverse trend with the GOCART AOD (500 nm) being much lower than the mean AOD (500 nm) extracted from AERONET retrievals. In addition to the effect of spatial representivity on AERONET versus GOCART mean AOD values, it should also be borne in mind that AERONET inversion products are obtained only under cloud-free conditions (Smirnov et al., 2000), whereas the GOCART model simulates values everywhere (Chin et al., 2002).

Fig. 10 presents the spectral behavior of several optical and microphysical parameters (ASYM, AAOD, SSA and CRI) commonly

Table 6

Descriptive statistics of key optical and microphysical parameters extracted from the global AERONET inversion record in each cluster for the multiyear global partition. Values are taken from the Level 2.0 Version 2 inversion product unless accompanied by a “*”, whereby the values are taken from the Level 1.5 Version 2 inversion product.

AERONET L2.0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Pixels	3	5	120	32	27	46	13	32	1	19
Records	2026	588	81,489	46,462	14,194	56,391	4901	27,319	378*	11,206
AOD (440)	0.07	0.06	0.18	0.37	0.24	0.16	0.09	0.23	0.04*	0.15
AE (440/675)	1.48	0.82	1.65	0.58	1.65	1.31	1.37	1.72	0.82*	1.53
H ₂ O	2.09	2.48	1.43	1.82	2.16	1.42	1.82	1.77	0.81*	1.71
ASYM (440)	0.71	0.72	0.70	0.74	0.67	0.71	0.71	0.71	0.73*	0.69
P180 (440)	0.27	0.31	0.18	0.19	0.19	0.19	0.24	0.18	0.50*	0.20
AAOD (440)	0.052	0.006*	0.047	0.060	0.086	0.078	0.052	0.038	0.003*	0.060
SSA (440)	0.93	0.92*	0.93	0.91	0.89	0.91	0.95	0.95	0.95*	0.91
LR (440)	78.5	50.7*	88.1	76.1	89.3	82.4	59.9	88.5	30.5*	87.9
CRI-R (440)	1.449	1.502*	1.445	1.478	1.490	1.477	1.510	1.422	1.516*	1.437
CRI-I (440)	0.010	0.007*	0.009	0.005	0.021	0.010	0.007	0.008	0.005*	0.013
r(f)	0.17	0.16	0.17	0.15	0.15	0.16	0.17	0.17	0.20*	0.15
σ(f)	0.46	0.49	0.44	0.52	0.42	0.45	0.46	0.44	0.48*	0.43
V(f)	0.01	0.01	0.02	0.03	0.03	0.02	0.01	0.03	0.00*	0.02
r(c)	2.54	2.68	2.84	2.27	3.04	2.68	2.43	2.96	2.05*	2.95
σ(c)	0.73	0.72	0.68	0.62	0.71	0.68	0.70	0.67	0.70*	0.69
V(c)	0.02	0.03	0.02	0.17	0.03	0.04	0.02	0.02	0.01*	0.03
η	0.28	0.18	0.54	0.15	0.48	0.34	0.36	0.60	0.21*	0.39
% Sphericity	96.9	90.0	81.8	1.8	98.9	43.9	69.9	98.1	54.7*	98.1

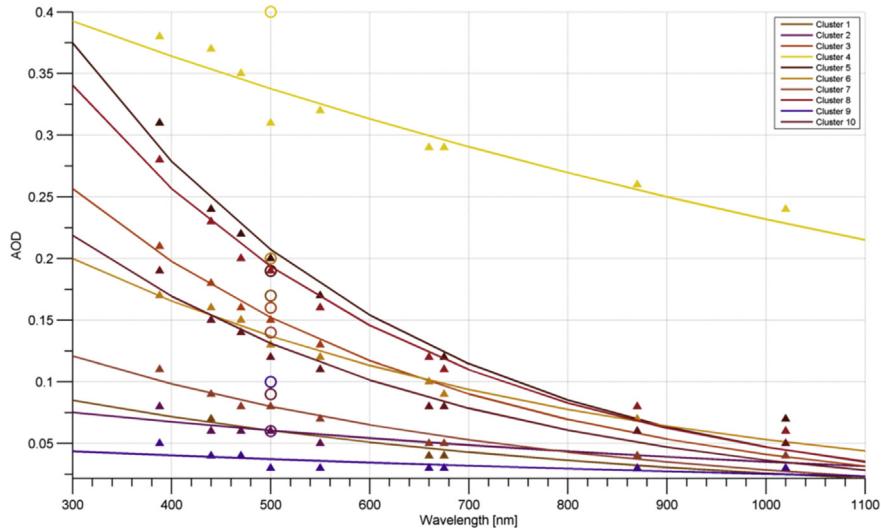


Fig. 9. Optimized least square regression fits to the spectral behavior of mean values of the AERONET-retrieved AOD for each cluster at the central wavelengths: 380, 440, 470 (extrapolated), 500, 550 (extrapolated), 660 (extrapolated), 675, 870 and 1020 nm (triangles). Note that the extrapolated values were calculated using the value of AE (440, 675). The circles at 500 nm are the mean cluster values of the total AOD obtained by GOCART for comparison.

used to characterize aerosol data by distinct types and their global climatology (see for example Dubovik et al., 2002; Eck et al., 2010; Chin et al., 2009; Russell et al., 2010, 2014).

With the exception of the AAOD (which presents a ‘logarithmic decay’ with wavelength for all clusters), other parameters show a vast difference in trends across aerosol mixtures. The AAOD suggests that sea-salt dominated aerosol can be directly distinguished from other mixtures in the visible part of the spectrum due to its significantly lower value of this parameter <0.01. Our inclusion of AAOD is based on initial findings reported by Russell et al. (2010, 2014). In the visible, the CRI-I also appears to discriminate between soot-laden aerosol mixtures and other mixtures when this parameter >0.01. ASYM, while unable to distinguish between dust-laden and marine-laden aerosol mixtures, does appear to separate these ‘natural’ types from mixtures containing varying significant proportions of BB and/or SU, particularly at longer wavelengths ≥870 nm. As described earlier, the AERONET inversion algorithm does not explicitly provide values of LR but via Eq. (1), they can be

estimated from sunphotometer retrievals. We note that the multiyear means for LR are strongly dependent on wavelength. Having said this, a glace at interpolated values at 550 nm reproduce the findings of other colleagues that the lowest mean values involve maritime aerosol, followed by dust-laden aerosol and mixtures containing BB or urban SU pollution (see for example Cattrall et al., 2005; Müller et al., 2007; Ansmann et al., 2013).

Finally, in this section, Fig. 11 presents mean volume size distributions and associated parameters for each cluster in the multiyear mean global partition.

Cluster 4 stands out. Referring to Table 5, this cluster has the largest coarse mode volume concentration ($V_c = 0.17$), the lowest fine mode fraction ($\eta = 0.15$) and proportion of spherical particles in the retrieval (% Sphericity = 1.8%). By comparison, all other clusters have coarse mode volume concentrations in the range $0.01 \leq V_c \leq 0.04$, fine mode fractions in the range $0.18 \leq \eta \leq 0.60$ and $43.9 \leq \% \text{Sphericity} \leq 98.9$. Another feature of Fig. 11 is that the coarse mode peaks appear to cluster around two distinct

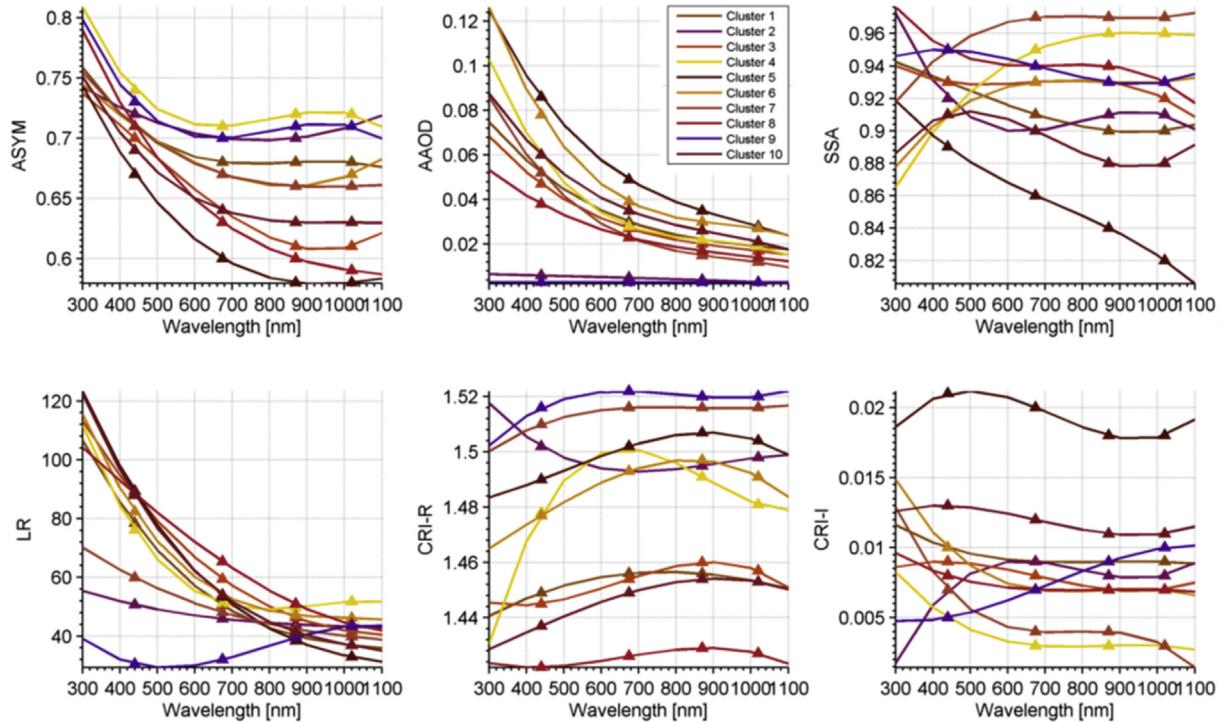


Fig. 10. Spectral behavior of the global mean values of key optical parameters (ASYM, AAOD, SSA, LR) and microphysical parameters (CRI-R and CRI-I) for each cluster from extracted AERONET data at 440, 675, 870 and 1020 nm. Interpolated lines are constructed using cubic spline fits to the extracted data.

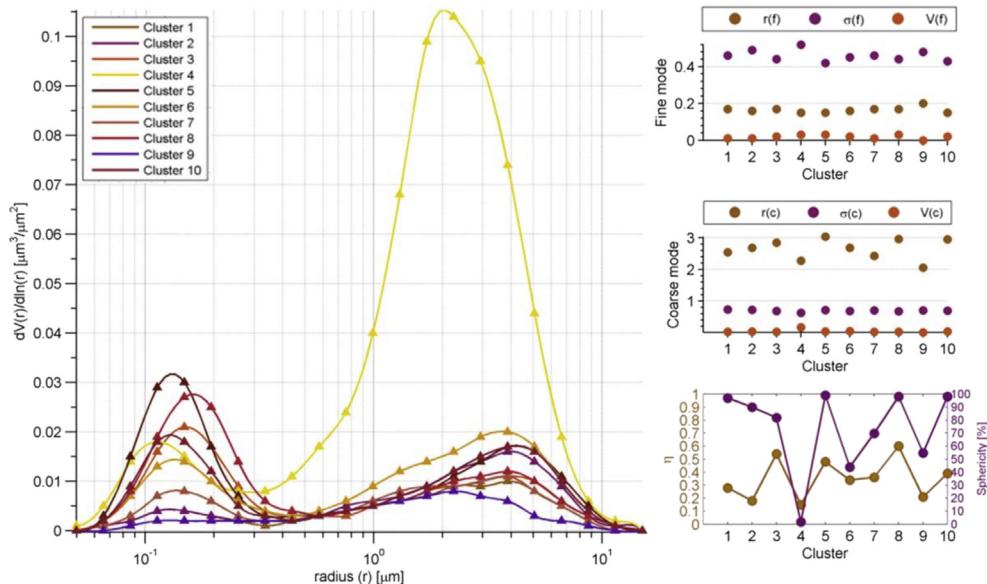


Fig. 11. Global mean size distributions for each cluster together with the derived fine and coarse mode parameters (r_f , σ_f , V_f and r_c , σ_c , V_c respectively). The percentage of spherical particles ('% Sphericity') is plotted with the fine mode fraction (η).

geometrical radii, the first ($r_c \approx 2.0\text{--}2.3 \mu\text{m}$) associated with the DU-dominated and SS-dominated clusters 4 and 9 and the second ($r_c \approx 2.4\text{--}3.1 \mu\text{m}$) associated with the other clusters. The location of the fine mode does not share this division and instead is confined to the narrow range: $0.15 \leq r_f \leq 0.20 \mu\text{m}$. Another detail here is that (with the exception of the SS-dominated cluster 9) the fine mode peaks of the cluster means AVSD do not exhibit skew or asymmetry. To the contrary, the behavior of the AVSD in the coarse mode region ($>0.6 \mu\text{m}$) either reveals a strong skew (clusters 2, 4 and 5)

or the existence of a 'shoulder' on the AVSD in the range $0.6\text{--}2 \mu\text{m}$. Such asymmetries and skews have been found to be associated with the presence of additional modes not fit by AERONET's inversion algorithm and constraint to bi-lognormality (Taylor et al., 2014). A particularly interesting example of a tri-modal size distribution attributed to fog-induced aerosol modification has been reported by Eck et al. (2012). New work in the direction of obtaining vertical profiles of fine and coarse mode size distributions with LIDAR/radiometer inversion codes (e.g. Granados-Muñoz et al.,

2014; Tsekeri et al., 2013) will be instrumental in validating such tendencies.

For plots analogous to Figs. 9, 10 and 11 resulting from the extraction of AERONET data for the seasonal mean partitions, we refer the reader to Figs. S10–S21 and Tables S7–S10 in the supplement accompanying this manuscript. Complete tables of extracted AERONET inversion products for multiyear mean and seasonal mean clusters are available at <http://apcg.meteo.noa.gr/aeromap/>.

4.3. Parameterization of global aerosol mixtures

In this section we briefly compare the taxonomy resulting from cluster analysis of GOCART data with the findings of both bivariate and multivariate classification models based on two or more aerosol parameters commonly used to perform aerosol typing.

4.3.1. Multivariate classification models

Omar et al. (2005) pioneered the first major quantitative study of AERONET-based aerosol mixtures and performed a cluster analysis of L1.5 V2 AERONET data with reference to 16 optical and microphysical parameters and obtained the categorization of aerosol into 6 types shown in Table 7.

For the multiyear mean, Table 8 presents values of this multivariate parameters set arising from our clustering of GOCART data with N = 10 clusters.

In order to ascertain whether there is a correspondence between the cluster characteristics obtained in this work and those obtained by Omar et al. (2005), we first normalized all cluster values to the maximum value for each parameter in Tables 7 and 8. We then calculated the norm of the Euclidean distance (from all 15 parameters) between each of the N = 10 clusters in our partition and the N = 6 clusters obtained by Omar et al. (2005) – i.e. 60 permutations. The results are collected in Table 9.

Association of each cluster identified by Omar et al. (2005) with the closest matching cluster obtained in this work is found by reading off the minimum value in each row of Table 9. This simple scheme allows for the results of cluster analyses from different studies to be compared. For the multivariate list of 15 parameters applied in Omar et al. (2005), the following associations are obtained from the minimum norm in each row:

- 'Dust' → cluster 4 ('DUST' in our taxonomy)
- 'Smoke' → cluster 10 ('Smokey SULFATE' in our taxonomy)
- 'Rural Background' → cluster 7 ('Marine Dusty SULFATE' in our taxonomy)
- 'Industrial Pollution' → cluster 8 ('SULFATE' in our taxonomy)

- 'Polluted Marine' → cluster 2 ('Marine SULFATE' in our taxonomy)
- 'Dirty Pollution' → cluster 10 ('Smokey SULFATE' in our taxonomy)

This comparison highlights the utility of the taxonomical system based on GOCART AOD data presented in this work. For example, although 'Polluted Marine' aerosol identified in Omar et al. (2005) is seen to comprise sea salt and sulfate, the taxonomy emphasizes that it is sulfate that dominates the composition of this particular aerosol mixture. Furthermore, labels like 'Rural Background' (which are quite common in the literature) are made more explicit being revealed as containing a dominant sulfate component infused with dust and sea salt in decreasing proportions. Note also that an ambiguity has been detected with this particular list of parameters; both 'Smoke' and 'Dirty Pollution' are associated with cluster 10, which is dominated by sulfate and a minor proportion comprising BC + OC.

A modified version of the 15 parameter multivariate classification scheme was introduced by Omar et al. (2009) for the operational algorithm of CALIOP that includes LIDAR ratios at 440 nm and 1020 nm as well as CRI values at both wavelengths but without the spectral optics parameters AOD, AE, ASYM and the important absorption parameter SSA. When we applied the cluster alignment method described above we found that the classification scheme of Omar et al. (2009) resulted in 4 indistinguishable 'DUST' clusters, suggesting that the new scheme underperforms the scheme of Omar et al. (2005) in the context of our GOCART-model deduced cluster analysis. The work of Omar et al. (2005, 2009) is progressive in that aerosol types are being studied in a multivariate manner using an array of AERONET inversion products. However, the sensitivity of the classifications to the choice of parameters used to perform the clustering shows that this is a challenging task. Instead, by performing cluster analysis on GOCART model AOD data for the composition of aerosol mixtures in each pixel, we avoid the problem of under- or over-specifying parameters in the classification process. More work is needed on the subject of multivariate analysis of aerosol mixtures and it is hoped that the results of the preceding sections will provide impetus for such studies.

4.3.2. Bivariate classification models

In Section 4.2 we saw how the spectral behavior of the AAOD is able to discriminate between sea-salt dominated aerosol and other mixtures in the visible part of the spectrum due to its significantly lower value of this parameter <0.01. With the advent in 1993 of AERONET and, in particular, with the production of the Version 2 inversion products from cloud-screened observations, access to a

Table 7

Results of the cluster analysis of AERONET Level 1.5 Version 2 data performed by Omar et al. (2005).

Omar et al., 2005	Dust	Smoke	Rural background	Industrial pollution	Polluted marine	Dirty pollution
AOD (673)	0.33	0.19	0.04	0.19	0.14	0.10
AE (441/673)	0.61	1.39	1.53	1.60	0.76	1.40
AE (673/873)	0.49	1.33	1.38	1.54	0.68	1.23
AE (873/1022)	0.28	1.04	0.95	1.29	0.53	0.85
SSA (673)	0.93	0.8	0.88	0.92	0.93	0.72
ASYM (673)	0.67	0.60	0.58	0.61	0.71	0.59
CRI-R (673)	1.452	1.520	1.449	1.410	1.394	1.410
CRI-I (673)	0.004	0.025	0.009	0.006	0.004	0.034
r (f)	0.12	0.14	0.13	0.16	0.17	0.14
σ (f)	0.39	0.45	0.41	0.42	0.48	0.43
V (f)	0.08	0.04	0.01	0.06	0.03	0.03
r (c)	2.83	3.73	3.59	3.55	3.27	3.56
σ (c)	0.65	0.76	0.74	0.73	0.69	0.76
V (c)	0.27	0.08	0.02	0.05	0.08	0.03
η	0.22	0.33	0.38	0.53	0.26	0.49

Table 8

Descriptive statistics of key optical and microphysical parameters extracted from the global AERONET inversion record in each cluster for the multiyear global partition of GOCART data (this work). Note that the quoted central wavelengths are marginally different (1–2 nm) to those reported by Omar et al. (2005). Values are taken from the Level 2.0 Version 2 inversion product unless accompanied by a “*”, whereby the values are taken from the Level 1.5 Version 2 inversion product.

AERONET L2	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
AOD (675)	0.04	0.05	0.09	0.29	0.12	0.09	0.05	0.11	0.03*	0.08
AE (440/675)	1.48	0.82	1.65	0.58	1.65	1.31	1.37	1.72	0.82*	1.53
AE (675/870)	0.44	0.48	1.32	0.46	1.48	0.94	0.73	1.50	-0.26*	1.18
AE (870/1020)	1.25	0.62	1.35	0.45	1.47	0.96	1.28	1.46	-0.30*	1.22
SSA (675)	0.91	0.90*	0.93	0.95	0.86	0.93	0.97	0.94	0.94*	0.90
ASYM (675)	0.68	0.70	0.64	0.71	0.60	0.67	0.67	0.63	0.70*	0.64
CRI-R (675)	1.456	1.493*	1.454	1.501	1.502	1.493	1.516	1.426	1.522*	1.449
CRI-I (675)	0.009	0.008*	0.008	0.003	0.020	0.007	0.004	0.007	0.007*	0.012
r (f)	0.17	0.16	0.17	0.15	0.15	0.16	0.17	0.17	0.20*	0.15
σ (f)	0.46	0.49	0.44	0.52	0.42	0.45	0.46	0.44	0.48*	0.43
V (f)	0.01	0.01	0.02	0.03	0.03	0.02	0.01	0.03	0.00*	0.02
r (c)	2.54	2.68	2.84	2.27	3.04	2.68	2.43	2.96	2.05*	2.95
σ (c)	0.73	0.72	0.68	0.62	0.71	0.68	0.70	0.67	0.70*	0.69
V (c)	0.02	0.03	0.02	0.17	0.03	0.04	0.02	0.02	0.01*	0.03
η	0.28	0.18	0.54	0.15	0.48	0.34	0.36	0.60	0.21*	0.39

Table 9

Norm of the Euclidean distance (from all 15 parameters) between each of the N = 10 clusters in our partition and the N = 6 clusters obtained by Omar et al. (2005).

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Dust	1.75	1.54	1.68	0.36	1.87	1.36	1.64	1.75	1.76	1.60
Smoke	0.91	1.07	0.62	1.48	0.64	0.59	0.86	0.82	1.76	0.43
Rural Background	0.73	0.96	0.67	1.82	1.14	0.61	0.54	0.97	1.60	0.62
Industrial Pollution	1.25	1.51	0.43	1.71	0.91	0.84	0.98	0.37	2.09	0.74
Polluted Marine	0.84	0.57	1.08	1.07	1.49	0.59	0.71	1.32	1.06	0.97
Dirty Pollution	0.96	1.13	0.77	1.84	0.70	0.81	1.00	0.99	1.71	0.60

larger array of aerosol optical and microphysical parameters encouraged a large number of studies attempting to classify ‘pure’ aerosol types and/or mixtures in 2D-plots based on pairs of parameters. The SSA has typically been used as a proxy for radiative absorption, and the AE or η has been used as proxy for the characteristic size of aerosol to classify aerosol with the pair combinations: {SSA, AE} (e.g. Mielonen et al., 2009) and {SSA, η } (e.g. Lee et al., 2010). Other studies have used the AOD as a proxy for aerosol volume, and the pair combination {AOD, AE} has been shown (Eck et al., 1999; Kalapureddy et al., 2009) to distinguish between dust (high AOD, low AE), marine (low AOD), and anthropogenic aerosols (high AOD, high AE), but is unable to sub-categorize anthropogenic aerosols into absorbing and non-absorbing without referring to geolocation information (Lee et al., 2010). In Fig. 12, the location of cluster centers overlaid on AERONET inversion products for three classification schemes: {AOD (440), AE (440/675)}, {SSA (440), AE (440/675)} and {Sphericity, η } are presented.

In the first two cases, although the points representing the AERONET data record suggest the existence of 3–4 discernable ‘shaded’ regions, the cluster centers do not lend themselves to an obvious demarcation in 2D. Furthermore, many clusters are very tightly bunched together making sets of inequalities corresponding to different aerosol types/mixtures difficult to construct. In MATLAB, we permuted through all parameter pair combinations and calculated the norm of the Euclidean distances in each 2D-space. The lowest norm (providing also the best visual separation) in the context of the global partition of multiyear and seasonal mean data was obtained with the microphysical parameters: {Sphericity, η } and is illustrated in the lower panel of Fig. 12. This choice of parameters, although new to the literature, better separates the clusters visually in 2D and is in line with a recent assessment of global aerosol type retrieved from MISR (Kahn and Gaitley, 2015) which also found that the distinctions between spherical vs. non-spherical

and fine vs. coarse mode are generally the most robust. Note also that the error bars are large and overlap in most cases, but should also be seen in the context of the density of points. The % Sphericity parameter should be considered an indication of dust only if the associated AE < 1.0 (Tom Eck, private communication). This is a significant limitation because, as we saw in the (SSA, AE) plot of Fig. 12, the tail of the dust distribution goes out to much larger values of AE (i.e., smaller particles) and transported dust can be dominated by particles having AE > 1.0. Despite ambiguities in the multivariate classification of Omar et al. (2005, 2009) and a lack of consensus on the most appropriate choice of parameters in bivariate classification schemes, it is hoped that new parameterizations such as the one presented in this work can help in this direction.

5. Discussion

The global spatial distribution of the multiyear and seasonal mean total AOD shows that the regions of peak aerosol load are dynamically stable and are broadly located over the Sahara and Nigeria in Northern Africa, in the Gobi desert to the north of the Himalayas, and over a large region centered on Beijing, based on the GOCART model analysis, but generally consistent with satellite observations. The main sources of aerosol, while exhibiting some seasonal variation in intensity, do not show large displacement in the location of their peaks. This is also partly due to the aggregation methodology adopted here, which means that major biomass burning regions for example, do not show up in comparison to dust.

The mean 2000–2006 (inclusive) map was found to be strongly representative of the mean annual AOD (500 nm) with SSIM >0.9995 for all years (reflected also by a dominant linear trend in the log-density regression of yearly mean data on the multiyear mean). In the context of the accuracy of the AERONET retrieval of the AOD which is ± 0.02 for the Level 1.5 and ± 0.01 for the Level 2.0 inversion products (Dubovik et al., 2000, 2002, 2006; Eck et al.,

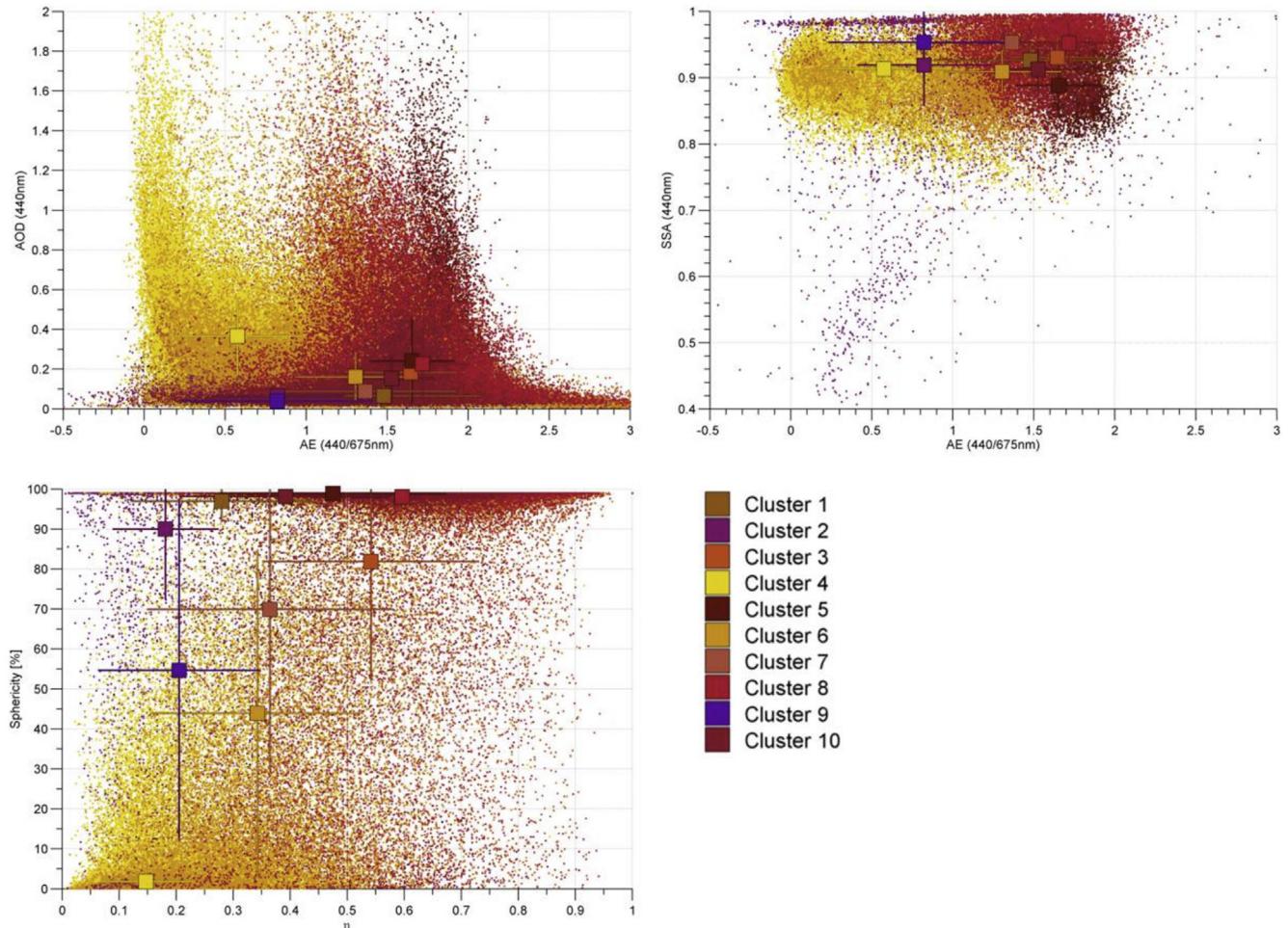


Fig. 12. The location of cluster centers overlaid on AERONET inversion products for two commonly used pair-combination spaces. {AOD (440), AE (440/675)} (top left panel), {SSA (440), AE (440/675)} (top right panel) and { $\%$ Sphericity, η } (lower panel). Error bars extend out to 1 standard deviation on both parameters and all wavelengths are given in nm.

1999), the mean pixel bias and RMSE were found to be within this accuracy limit supporting the assumption that the multiyear 2000–2006 mean is a good proxy for maps of the annual mean AOD (500 nm). For comparison with approaches based on direct-sun AERONET data, Eck et al. (1999) gives ± 0.01 as the AOD uncertainty. Overall, we found that the locations of peak BB, SU, DU and SS concentrations contributing to the multiyear mean AOD proxy map were found where expected. Although the patterns are generally supported by observations, severe fire or dust storm seasons for example, are underestimated since the model tends to under-represent variability on the seasonal and especially on interannual timescales.

Cluster analysis, based on the k-means algorithm with random seeding and a stopping condition, applied to the percentage contributions of each aerosol type in the global mean multiyear and seasonal maps, was found to mitigate the effect of centers being confined to local minima, and allowed for the specification of lower and upper bounds on the optimal number of clusters. For the multiyear mean, the optimal number of clusters is $N = 10$ and subsequent addition of clusters led to very minor modifications of the distribution of cluster centers, with the change in the norm of Euclidean distances being $<10\%$. In the case of seasonal mean maps, this approach identifies $N = 10$ clusters for the season spanned by the months DJF and $N = 11$ for the other seasons spanned by the months MAM, JJA and SON. In the context of the spatial extent of

the clusters, the lowest pixel count was 1187 for cluster 1 of the multiyear mean partition, and varied between 1648 and 2481 for the smallest cluster in the seasonal mean partitions. The mean and the median percentage contributions of BB, SU, DU and SS were found to be within 2% of each other for both cluster analysis applied to the multiyear mean and the seasonal mean maps. The same margin of error was found to be true of the difference between twice the standard deviation and 1.5 times the inter-quartile range suggesting that the underlying distributions are strongly symmetrical and near-Gaussian.

The taxonomical naming system based on constituent aerosol types comprising at least 20% of the overall compositional mixture allows for a simple ‘adjective’ – ‘object’ labeling of clusters that fits with expected trends. With this convention, the global partitioning of the multiyear mean map led to identification of two regions occupied by ‘DUST’ (3.8% of the globe) and ‘SULFATE’ (6.4% of the globe). The largest clusters comprised the 2-component mixtures: ‘Marine SULFATE’ and ‘Dusty SULFATE’ (occupying 24.7% and 23.9% of the globe respectively). As expected, clusters ('Sulfurous Dusty SMOKE' and 'Sulfurous SMOKE') containing a dominant smoke component, occupied only 1.8% and 3.1% of the globe respectively – providing a justification for combining biomass burning products (BC and OC) in a combined measure ('BB' = BC + OC). Table S2 and the brief discussion that follows Table 5 show how clusters having the same label in the taxonomy can be tracked and compared

across seasonal trends.

Section 4.2 compares the optical microphysical parameters of each cluster in the global partition of multiyear and seasonal means based on the mean value of extracted AERONET inversion products at co-located sites. There is a strong variation in the number of available AERONET Level 2.0 inversion records that can be used for calculating mean characteristic values for each cluster due to the unevenness of sites hosting operational CIMEL sunphotometers worldwide. Despite this, careful selection of co-located Level 1.5 data was able to provide an estimate in the case of data gaps. Although not ideal from the point of view of quality assurance homogeneity, the results presented in Section 4.2 are novel and provide a ‘first guess’ of optical and microphysical parameters at a resolution of $1^\circ \times 1^\circ$ for both mean multiyear mixtures and seasonal mixtures.

6. Conclusion

The initial results presented here suggest that aerosol mixtures are more diverse than typically provided by current satellite remote-sensing products (≈ 6 types). In the context of mean GOCART AOD for the period 2000–2006 in conjunction with a stopping condition of 10% in the change in cluster centers, the number of clusters in the global partition was found to be around $N = 10$. In addition to the somewhat subjective nature of the stopping condition itself, this finding should also be seen in the light of uncertainties in aerosol type retrievals provided by GOCART. Perhaps more important and significant is the finding that the number of clusters was found to be bounded by $N = 7$ and $N = 13$. Furthermore, the comparison of cluster spatial extent and composition in both the multiyear and seasonal partitions suggests that they are fairly robust and stable – i.e. their major characteristics do not change qualitatively on these timescales. This approach can be useful for future satellite missions that need aerosol typing or for climatological studies that deal with the impact of different aerosol types.

Global chemical transport models and/or circulation models have prognostic capacity but tend to be most reliable on the regional scale due to their dependence on chemical and meteorological boundary conditions. It would be interesting to assess whether source inventories currently used to specify model inputs can be supported by new parameterizations of aerosol mixtures (even of a ‘first guess’ nature like those presented here), particularly those inputs related to aerosol microphysics which can be a large source of uncertainty.

For forward-planning and placement of new sunphotometers in the growing AERONET array, understanding aerosol mixture characteristics can support assessments of priority locations. An important example of this is the prominent sea-salt dominated band in the southern oceans. Island sites including: Tahiti, Crozet_Island, Amsterdam_Island and Dunedin border it to the North, and coastal Antarctic sites including: Marambio, McMurdo, Utsteinen, and Vechernaya_Hill border it to the South. Ideal placement of CIMEL sunphotometers to measure this particular aerosol mixture associated with cluster 9 could include: South Georgia and the South Sandwich Islands, Montagu Island, the Heard and McDonald Islands and/or the French Southern & Antarctic Lands. A similar logic can be applied to other clusters to identify new locations for effective sampling of the atmospheric column with CIMEL sunphotometers. Global partitions presented in this work can supplement other approaches to help identify high priority locations for AERONET and/or *in situ* measurements. For example, a satellite-based study of coherent spatial features of the AOD bias between MODIS and MISR aerosol products (Shi et al., 2011) identified and documented a number of potential regions in

the American, African and Asian continents in particular.

Finally, the clusters identified in this work, despite coming from a single model (GOCART), provide a somewhat more objective partition of the globe into regions having the same mean aerosol composition that avoids *ad hoc* approaches based on imprecise divisions such as latitudinal bands, geographical borders or continents. Reciprocating the good will and provision of open data by co-workers in the field, we have made cluster index maps and relevant tables of cluster characteristics available online to facilitate new studies based on these partitions and to enable others to reproduce and extend these findings.

Acknowledgments

MT would like to thank Mian Chin for making high temporal resolution (3-hr) GOCART model output data available for the years 2000–2006, the principal investigators of AERONET for kindly maintaining CIMEL instruments and ensuring timely upload of high quality inversion products, and NASA/GES-DISC and the Giovanni team for their commitment and support for open data visualization, provision and technical support. This research has been financed by the “Development Proposals of Research Entities – KRPIS” framework which is funded by N.P. “Competitiveness and Entrepreneurship”, Action: “THESPIA – Development of synergistic and integrated methods and tools for monitoring, management and forecasting of Environmental parameters and pressures”. Part of the work was supported by the European Union Seventh Framework Programme (FP7-REGPOT-2012-2013-1), in the framework of the project BEYOND, under Grant Agreement No. 316210 (BEYOND – Building Capacity for a Centre of Excellence for EO-based monitoring of Natural Disasters). The work of R. Kahn is supported in part by NASA’s Climate and Radiation Research and Analysis Program, under H. Maring, NASA’s Atmospheric Composition Program under R. Eckman, and the EOS-MISR project. MT is grateful to the members of the Atmospheric Chemistry and Physics Group at the National Observatory of Athens for their hospitality and training in the field.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.atmosenv.2015.06.029>.

References

- Ansmann, A., Müller, D., Wandinger, U., Mamouri, R.E., 2013. Lidar profiling of aerosol optical and microphysical properties from space: overview, review, and outlook. In: First International Conference on Remote Sensing and Geo-information of Environment. International Society for Optics and Photonics, p. 879502.
- Boyouk, N., Léon, J.F., Delbarre, H., Podvin, T., Deroo, C., 2010. Impact of the mixing boundary layer on the relationship between PM_{2.5} and aerosol optical thickness. *Atmos. Environ.* 44 (2), 271–277.
- Cattrall, C., Reagan, J., Thome, K., Dubovik, O., 2005. Variability of aerosol and spectral lidar and backscatter and extinction ratios of key aerosol types derived from selected aerosol robotic network locations. *J. Geophys. Res.* <http://dx.doi.org/10.1029/2004JD005124>. *Atmospheres* (1984–2012):110(D10S11).
- Chen, W.T., Kahn, R.A., Nelson, D., Yau, K., Seinfeld, J.H., 2008. Sensitivity of multi-angle imaging to the optical and microphysical properties of biomass burning aerosols. *J. Geophys. Res.* 113, D10203.
- Chin, M., Rood, R.B., Lin, S.J., Müller, J.F., Thompson, A.M., 2000. Atmospheric sulfur cycle simulated in the global model GOCART: model description and global properties. *J. Geophys. Res.* 105 (D20), 24671–24724.
- Chin, M., Ginoux, P., Kinne, S., Torres, O., Holben, B.N., Duncan, B.N., Nakajima, T., 2002. Tropospheric aerosol optical thickness from the GOCART model and comparisons with satellite and Sun photometer measurements. *J. Atmos. Sci.* 59 (3), 461–483.
- Chin, M., Diehl, T., Dubovik, O., Eck, T.F., Holben, B.N., Sinyuk, A., Streets, D.G., 2009. Light absorption by pollution, dust, and biomass burning aerosols: a global model study and evaluation with AERONET measurements. *Ann. Geophys.* 27, 3439–3464.

- Chin, M., Diehl, T., Tan, Q., Prospero, J.M., Kahn, R.A., Remer, L.A., Yu, H., Sayer, A.M., Bian, H., Geogdzhayev, I.V., Holben, B.N., Howell, S.G., Huebert, B.J., Hsu, N.C., Kim, D., Kucsera, T.L., Levy, R.C., Mishchenko, M.I., Pan, X., Quinn, P.K., Schuster, G.L., Streets, D.G., Strode, S.A., Torres, O., Zhao, X.-P., 2014. Multi-decadal aerosol variations from 1980 to 2009: a perspective from observations and a global model. *Atmos. Chem. Phys.* 14, 3657–3690.
- David, A., Vassilytskii, S., 2007. The Advantages of careful Seeding: SODA '07. In: Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 1027–1035.
- 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 (D16), 20673–20696.
- Dubovik, O., Smirnov, A., Holben, B.N., King, M.D., Kaufman, Y.J., Eck, T.F., Slutsker, I., 2000. Accuracy assessment of aerosol optical properties retrieval from AERONET sun and sky radiance measurements. *J. Geophys. Res.* 105, 9791–9806.
- 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.
- Dubovik, O., Sinyuk, A., Lapyonok, T., Sinyuk, A., Mishchenko, M.I., Yang, P., Eck, T.F., Volten, H., Munoz, O., Veihelmann, B., van der Zander, W.J., Sorokin, M., Slutsker, I., 2006. Application of light scattering by spheroids for accounting for particle non-sphericity in remote sensing of desert dust. *J. Geophys. Res.* 111 <http://dx.doi.org/10.1029/2005JD006619>, D11208.
- Dubovik, O., Herman, M., Holdak, A., Lapyonok, T., Tanre, D., Deuze, J.L., Ducos, F., Sinyuk, A., Lopatin, A., 2011. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. *Atmos. Meas. Tech.* 4, 975–1018.
- Eck, T.F., Holben, B.N., Reid, J.S., Dubovik, O., Smirnov, A., O'Neill, N.T., Kinne, S., 1999. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. *J. Geophys. Res.* 104 (D24), 31333–31349. *Atmospheres* (1984–2012).
- Eck, T.F., Holben, B.N., Sinyuk, A., Pinker, R.T., Goloub, P., Chen, H., Xia, X., 2010. Climatological aspects of the optical properties of fine/coarse mode aerosol mixtures. *J. Geophys. Res. Atmos.* 115, D19 (1984–2012).
- Eck, T.F., Holben, B.N., Reid, J.S., Giles, D.M., Rivas, M.A., Singh, R.P., Goloub, P., 2012. Fog-and-cloud-induced aerosol modification observed by the Aerosol Robotic Network (AERONET). *J. Geophys. Res.* 117 (D7). *Atmospheres* (1984–2012).
- Gerasopoulos, E., Kouliou, E., Kalivitis, N., Kouvarakis, G., Saarikoski, S., Mäkelä, T., Mihalopoulos, N., 2007. Size-segregated mass distributions of aerosols over Eastern Mediterranean: seasonal variability and comparison with AERONET columnar size-distributions. *Atmos. Chem. Phys.* 7 (10), 2551–2561.
- Ginoux, P., Chin, M., Tegen, I., Prospero, J.M., Holben, B.N., Dubovik, O., Lin, S.J., 2001. Sources and distributions of dust aerosols simulated with the GOCART model. *J. Geophys. Res.* 106 (D17), 20255–20320.
- Granados-Muñoz, M.J., Guerrero-Rascado, J.L., Bravo-Aranda, J.A., Navas-Guzmán, F., Valenzuela, A., Lyamani, H., Alados-Arboledas, L., 2014. Retrieving aerosol microphysical properties by Lidar-Radiometer Inversion Code (LIRIC) for different aerosol types. *J. Geophys. Res.* 119 (8), 4836–4858. *Atmospheres*.
- Hansen, J., 2005. Earth's energy imbalance: confirmation and implications. *Science* 308, 1431–1435.
- Higurashi, A., Nakajima, T., 2002. Detection of aerosol types over the East China Sea near Japan from four-channel satellite data. *Geophys. Res. Lett.* 29 (17), 17–21.
- Holben, B.N., Eck, T.F., Slutsker, I., Tanré, D., Buis, J.P., et al., 1998. AERONET – a federated instrument network and data archive for aerosol characterization. *J. Rem. Sens. Environ.* 66, 1–16.
- Holben, B.N., Eck, T.F., Slutsker, I., Smirnov, A., Sinyuk, A., Schafer, J., Dubovik, O., 2006. AERONET's version 2.0 quality assurance criteria. In: Asia-Pacific Remote Sensing Symposium. International Society for Optics and Photonics, 64080Q.
- Intergovernmental Panel on Climate Change (IPCC), 2013. Climate Change 2013 The Physical Science Basis: Contribution of the Working Group I to the Fifth Assessment Report of the IPCC. Cambridge University Press, New York.
- Jeong, M.J., Li, Z., 2005. Quality, compatibility, and synergy analyses of global aerosol products derived from the advanced very high resolution radiometer and total ozone mapping spectrometer. *J. Geophys. Res.* 110 (D10). *Atmospheres* (1984–2012).
- Kahn, R.A., 2012. Reducing the uncertainties in direct aerosol radiative forcing. *Surv. Geophys.* 33 (3–4), 701–721.
- Kahn, R.A., Gaitley, B.J., 2015. An analysis of global aerosol type as retrieved by MISR. *J. Geophys. Res. Atmos.* 120, 4248–4281. <http://dx.doi.org/10.1002/2015JD023322>.
- Kahn, R.A., Banerjee, P., McDonald, D., 2001. Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean. *J. Geophys. Res.* 106 (D16), 18219–18238. *Atmospheres* (1984–2012).
- Kahn, R.A., Chen, Y., Nelson, D., Leung, F.-Y., Li, Q., Diner, D., Logan, J., 2008. Wildfire smoke injection heights: two perspectives from space. *Geophys. Res. Lett.* 35, L04809. <http://dx.doi.org/10.1029/2007GL032>.
- Kahn, R.A., Gaitley, B.J., Garay, M.J., Diner, D.J., Eck, T.F., Smirnov, A., Holben, B.N., 2010. Multiangle Imaging Spectroradiometer global aerosol product assessment by comparison with the aerosol robotic network. *J. Geophys. Res.* 115, D23209. <http://dx.doi.org/10.1029/2010JD014601>.
- Kalapureddy, M.C.R., Kaskaoutis, D.G., Ernest Raj, P., Devara, P.C.S., Kambezidis, H.D., Kosmopoulos, P.G., Nastas, P.T., 2009. Identification of aerosol type over the Arabian Sea in the premonsoon season during the Integrated Campaign for Aerosols, Gases and Radiation Budget (ICARB). *J. Geophys. Res. Atmos.* 114, D17 (1984–2012).
- Kalashnikova, O.V., Kahn, R.A., 2006. Ability of multiangle remote sensing observations to identify and distinguish mineral dust types: 2. Sensitivity over dark water. *J. Geophys. Res.* 111 (D11). *Atmospheres* (1984–2012).
- Kalashnikova, O.V., Kahn, R.A., Sokolik, I.N., Li, W.-H., 2005. The ability of multi-angle remote sensing observations to identify and distinguish mineral dust types: part 1. Optical models and retrievals of optically thick plumes. *J. Geophys. Res.* 110 <http://dx.doi.org/10.1029/2004JD004550>.
- Kaufman, Y.J., Boucher, O., Tanré, D., Chin, M., Remer, L.A., Takemura, T., 2005. Aerosol anthropogenic component estimated from satellite data. *Geophys. Res. Lett.* 32 (17).
- Kim, S.W., Yoon, S.C., Kim, J., Kim, S.Y., 2007. Seasonal and monthly variations of columnar aerosol optical properties over east Asia determined from multi-year MODIS, LIDAR, and AERONET Sun/sky radiometer measurements. *Atmos. Environ.* 41 (8), 1634–1651.
- Kim, M.H., Kim, S.W., Yoon, S.C., Omar, A.H., 2013. Comparison of aerosol optical depth between CALIOP and MODIS-aqua for CALIOP aerosol subtypes over the ocean. *J. Geophys. Res.* 118 (23), 13–241. *Atmospheres*.
- Kinne, S., O'Donnell, D., Stier, P., Kloster, S., Zhang, K., Schmidt, H., Stevens, B., 2013. MAC-v1: a new global aerosol climatology for climate studies. *J. Adv. Mod. Earth Syst.* 5 (4), 704–740.
- Lee, K.H., Li, Z., Kim, Y.J., Kokhanovsky, A., 2009. Atmospheric aerosol monitoring from satellite observations: a history of three decades. In: Atmospheric and Biological Environmental Monitoring. Springer, Netherlands, pp. 13–38.
- Lee, J., Kim, J., Song, C.H., Kim, S.B., Chun, Y., Sohn, B.J., Holben, B.N., 2010. Characteristics of aerosol types from AERONET sunphotometer measurements. *J. Atmos. Environ.* 44, 3110–3117.
- MacQueen, J.B., 1967. Some methods for classification and analysis of multivariate observations. In: Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability, 1. University of California Press, pp. 281–297.
- Mann, G.W., Carslaw, K.S., Reddington, C.L., Pringle, K.J., Schulz, M., Asmi, A., Hansson, H.C., 2014. Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity. *Atmos. Chem. Phys.* 14 (9), 4679–4713.
- Mattis, I., Ansmann, A., Müller, D., Wandinger, U., Althausen, D., 2002. Dual-wavelength Raman lidar observations of the extinction-to-backscatter ratio of Saharan dust. *Geophys. Res. Lett.* 29 (9), 1306. <http://dx.doi.org/10.1029/2002GL0147>.
- Mielonen, T., Arola, A., Komppula, M., Kukkonen, J., Koskinen, J., de Leeuw, G., Lehtinen, K.E.J., 2009. Comparison of CALIOP level 2 aerosol subtypes to aerosol types derived from AERONET inversion data. *Geophys. Res. Lett.* 36, L18804. <http://dx.doi.org/10.1029/2009GL039609>.
- Mishchenko, M.I., Cairns, B., Kopp, G., Schueler, C.F., Fafaul, B.A., Hansen, J.E., Hooker, R.J., Itchikawich, T., Maring, H.B., Travis, L.D., 2007a. Accurate monitoring of terrestrial aerosols and total solar irradiance: introducing the glory mission. *Bull. Amer. Meteorol. Soc.* 88, 677–691.
- Mishchenko, M.I., Geogdzhayev, I.V., Cairns, B., Rossow, W.B., Carlson, B.E., Lacis, A.A., Liu, L., Travis, L.D., 2007b. Long-term satellite record reveals likely recent aerosol trend. *Science* 315, 1543. <http://dx.doi.org/10.1126/science.1136709>.
- Mishchenko, M.I., Geogdzhayev, I.V., Cairns, B., Carlson, B.E., Chowdhary, J., Lacis, A.A., Liub, L., Rossow, W.B., Larry, D., Travis, L.D., 2007c. Past, present, and future of global aerosol climatologies derived from satellite observations: a perspective. *J. Quant. Spectrosc. Rad. Trans.* 106, 325–347.
- Müller, D., Ansmann, A., Mattis, I., Tesche, M., Wandinger, U., Althausen, D., Pisani, G., 2007. Aerosol-type-dependent lidar ratios observed with Raman lidar. *J. Geophys. Res.* 112 (D16). *Atmospheres* (1984–2012).
- Omar, A.H., Winker, D.M., Kittaka, C., Vaughan, M.A., Liu, Z., Hu, Y., Trepte, C.R., Rogers, R.R., Ferrare, R.A., Lee, K.P., Kuehn, R.E., Hostetler, C.A., 2009. The CALIPSO automated aerosol classification and Lidar ratio selection algorithm. *J. Atmos. Oceanic Technol.* 26, 1994–2014.
- Remer, L.A., Kaufman, Y.J., Tanré, D., Mattoo, S., Chu, D.A., Martins, J.V., Holben, B.N., 2005. The MODIS aerosol algorithm, products, and validation. *J. Atmos. Sci.* 62 (4), 947–973.
- Remer, L.A., Kleidman, R.G., Levy, R.C., Kaufman, Y.J., Tanré, D., Mattoo, S., Martins, J.V., Ichoku, C., Koren, I., Yu, H., Holben, B.N., 2008. Global aerosol climatology from the MODIS satellite sensors. *J. Geophys. Res.* 113, D14S07. <http://dx.doi.org/10.1029/2007JD009661>. *Atmospheres*.
- Russell, P.B., Bergstrom, R.W., Shinozuka, Y., Clarke, A.D., DeCarlo, P.F., Jimenez, J.L., Strawa, A., 2010. Absorption Angstrom exponent in AERONET and related data as an indicator of aerosol composition. *Atmos. Chem. Phys.* 10 (3), 1155–1169.
- Russell, P.B., et al., 2014. A multiparameter aerosol classification method and its application to retrievals from space-borne polarimetry. *J. Geophys. Res.* 119 <http://dx.doi.org/10.1002/2013JD021411>. *Atmospheres*.
- Samset, B.H., Myhre, G., Schulz, M., 2014. Upward adjustment needed for aerosol radiative forcing uncertainty. *Nat. Clim. Change* 4 (4), 230–232. <http://dx.doi.org/10.1038/nclimate2170>.
- Schwartz, S.E., Charlson, R.J., Kahn, R.A., Rodhe, H., 2014. Earth's climate sensitivity: apparent inconsistencies in recent analyses. *Earth's Future*. <http://dx.doi.org/10.1002/2014EF000273>.
- Shi, Y., Zhang, J., Reid, J.S., Hyer, E.J., Eck, T.F., Holben, B.N., Kahn, R.A., 2011. A critical examination of spatial biases between MODIS and MISR aerosol products –

- application for potential AERONET deployment. *Atmos. Meas. Tech.* 4, 2823–2836. <http://dx.doi.org/10.5194/amt-4-2823-2011>.
- Smirnov, A., Holben, B.N., Eck, T.F., Dubovik, O., Slutsker, I., 2000. Cloud screening and quality control algorithms for the AERONET database. *Remote Sens. Environ.* 73, 337–349.
- Stier, P., Schutgens, N.A.J., Bellouin, N., Bian, H., Boucher, O., Chin, M., Zhou, C., 2013. Host model uncertainties in aerosol radiative forcing estimates: results from the AeroCom prescribed intercomparison study. *Atmos. Chem. Phys.* 13 (6), 3245–3270.
- Taylor, M., Kazadzis, S., Gerosopoulos, E., 2014. Multi-modal analysis of aerosol robotic network size distributions for remote sensing applications: dominant aerosol type cases. *Atmos. Meas. Tech.* 7, 839–858.
- Tsekeri, A., Amiridis, V., Kokkalis, P., Basart, S., Chaikovsky, A., Dubovik, O., Gross, B., 2013. Application of a synergetic lidar and sunphotometer algorithm for the characterization of a dust event over Athens, Greece. *Br. J. Environ. Clim. Change* 3 (4), 531–546.
- Val Martin, M., Logan, J.A., Kahn, R.A., Leung, F.-Y., Nelson, D., Diner, D., 2010. Fire smoke injection heights over North America constrained from the terra multi-angle imaging spectroradiometer. *Atmos. Chem. Phys.* 10, 1491–1510.
- Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P., 2004. Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Imag. Process* 13 (4), 600–612.
- Winker, D.M., Tackett, J.L., Getzewich, B.J., Liu, Z., Vaughan, M.A., Rogers, R.R., 2013. The global 3-D distribution of tropospheric aerosols as characterized by CALIOP. *Atmos. Chem. Phys.* 13 (6), 3345–3361.
- Yu, H., Chin, M., Winker, D.M., Omar, A.H., Liu, Z., Kittaka, C., Diehl, T., 2010. Global view of aerosol vertical distributions from CALIPSO lidar measurements and GOCART simulations: regional and seasonal variations. *J. Geophys. Res.* 115 (D4). *Atmospheres* (1984–2012).
- Yu, H., Chin, M., West, J.J., Atherton, C.S., Bellouin, N., Bergmann, D., Bey, I., Bian, H., Diehl, T., Forberth, G., Hess, P., Schulz, M., Shindell, D., Takemura, T., Tan, Q., 2013. A multimodel assessment of the influence of regional anthropogenic emission reductions on aerosol direct radiative forcing and the role of intercontinental transport. *J. Geophys. Res.* 2013 (118), 700–720. <http://dx.doi.org/10.1029/2012JD018148>.
- Zhang, J., Reid, J.S., 2010. A decadal regional and global trend analysis of the aerosol optical depth using a data-assimilation grade over-water MODIS and level 2 MISR aerosol products. *Atmos. Chem. Phys.* 10, 10949–10963. <http://dx.doi.org/10.5194/acp-10-10949-2010>.