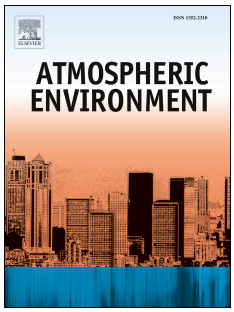
Accepted Manuscript



Global aerosol mixtures and their multiyear and seasonal characteristics

M. Taylor, S. Kazadzis, V. Amiridis, R.A. Kahn

PII: S1352-2310(15)30170-9

DOI: 10.1016/j.atmosenv.2015.06.029

Reference: AEA 13907

To appear in: *Atmospheric Environment*

Received Date: 3 February 2015

Revised Date: 12 June 2015

Accepted Date: 16 June 2015

Please cite this article as: Taylor, M., Kazadzis, S., Amiridis, V., Kahn, R.A., Global aerosol mixtures and their multiyear and seasonal characteristics, *Atmospheric Environment* (2015), doi: 10.1016/j.atmosenv.2015.06.029.

Global aerosol mixtures and their multiyear and seasonal characteristics

Taylor, M.[1]*, Kazadzis, S.[2], Amiridis V.[3], Kahn, R.A.[4]

[1] Institute for Environmental Research and Sustainable Development (IERSD), National Observatory of Athens (NOA), Metaxa & Vas Pavlou, Penteli, 15236, Athens, Greece.

[2] Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center (PMOD/WRC), Dorfstrasse 33, 7260 Davos Dorf, Switzerland.

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

[4] NASA Goddard Space Flight Centre (GSFC), Greenbelt, MD 20771, U.S.A.

* Correspondence: Tel: +30 2108109170 / Fax : +30 2108103236 / Email address: mtaylor@noa.gr

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 x 2 degree 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 colours to constituent aerosol types enables true colour-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 (1x1 degree) mean compositions of aeorosol 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 facililate further study of the global distribution of aerosol mixtures and possibly for the selection of new AERONET site locations.

Keywords

Atmospheric aerosol | Cluster analysis | Classification | GOCART | AERONET

1. Introduction

It is now known that aerosols contribute strongly to the change in radiative climate forcing and prediction uncertainty (**Mischenko et al, 2007; 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 et al, 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 contrain 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 behaviour in time (**Chin et al, 2014**). It should be borne in mind that the convergence of results between IPCC/AR4 and AR5 in model inter-comparison 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 & Nakajima, 2000**, **2002**; **Jeong & Li, 2005**; **Kahn et al, 2010**; **Kaufman et al, 2005**; **Kim et al, 2007, 2013**; **Lee et al, 2009**, **Remer et al, 2005, 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 inter-continental transport of aerosol (e.g. **Yu et al, 2012**).

A proper characterization of aerosol requires knowledge of aerosol size, shape, and composition provided by optical and microphysical parameters (**Dubovik & 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 unevently 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 centres 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 ascertions, 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, multi-spectral observations with MISR (**Kalashnikova & 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 x 2 degree longitude x latitude) provided by the Goddard Chemistry Aerosol Radiation and Transport (GOCART) model (**Chin et al, 2000; 2002; 2014,** and **Ginoux et al, 2001**). The GOCART model provides global, continuous, gridded 3-hourly values of total AOD at 500nm 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 colour-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 x 2 degrees (longitude x latitude) 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, aborption, 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 Secion 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 degrees for the central spectral wavelengths (440, 675, 870 and 1020nm) 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 1x1 degrees, the GOCART data was downscaled from 2.5 x 2 degree longitude x latitude to 1x1 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 64800 x 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 labelling the result as 'BB' (corresponding to 'Biomass Burning'). Figs. 1 and 2 show the global spatial distribution of the mean total AOD (500nm) 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.

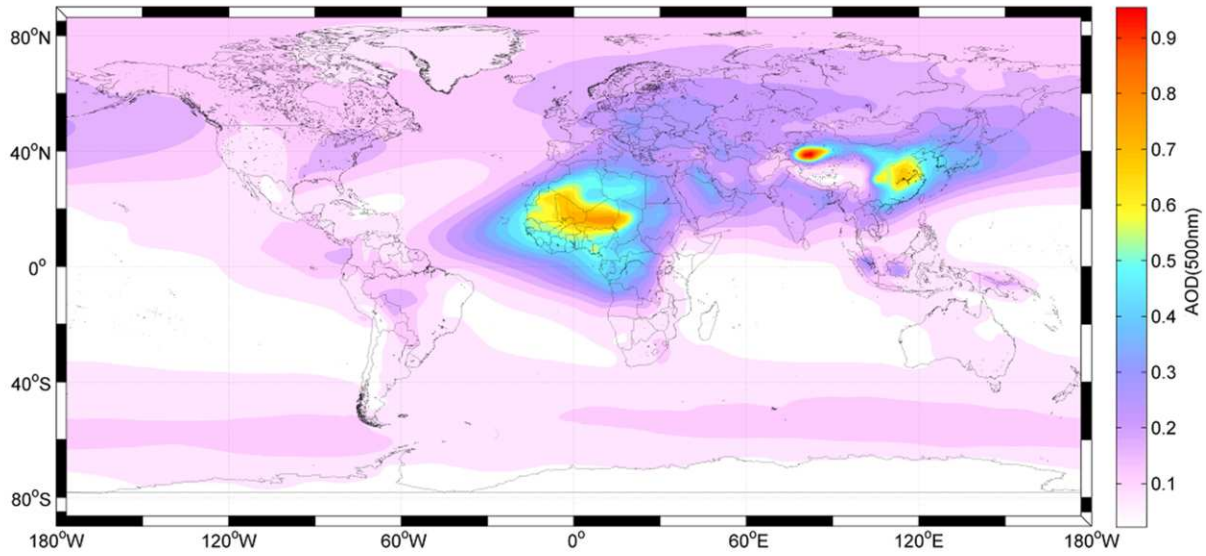


Fig. 1. The multiyear (2000-2006 inclusive) mean AOD (500nm) from the GOCART model.

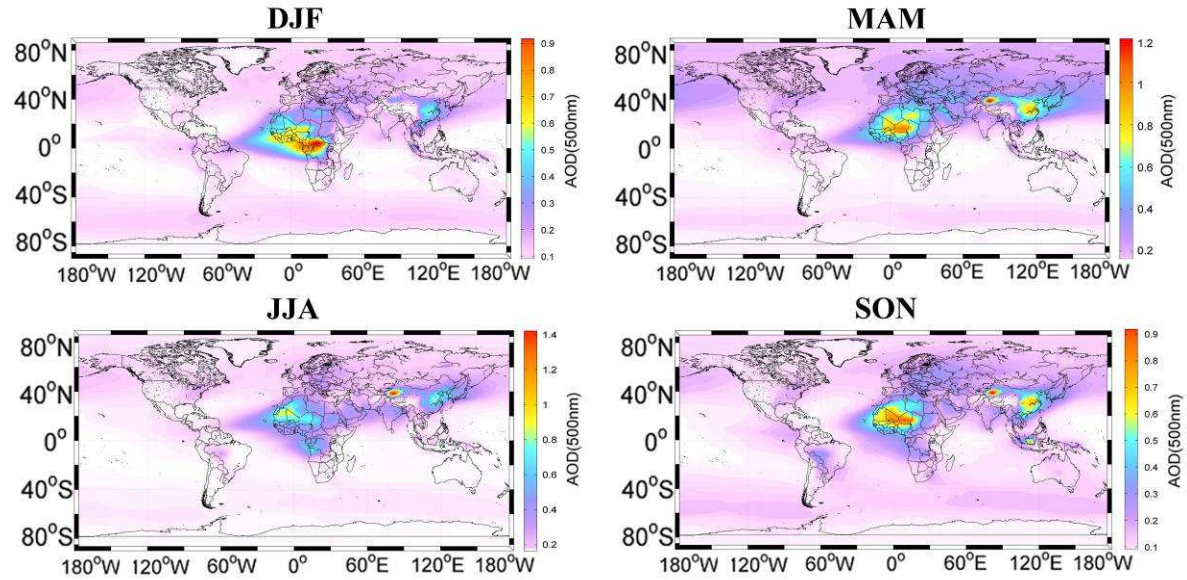


Fig. 2. The seasonal mean AOD (500nm) from the GOCART model for the monthly triplets: DJF, MAM, JJA and SON.

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. colour is used only to represent magnitude), the mean bias and root-mean squared error (RMSE) can be calculated directly by averaging differences over all 64800 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**.

Table 1. Statistics resulting from a comparison of the gridded (1x1 degree) mean total global AOD (500nm) 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 (1x1 degree) 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 mutiyear 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.

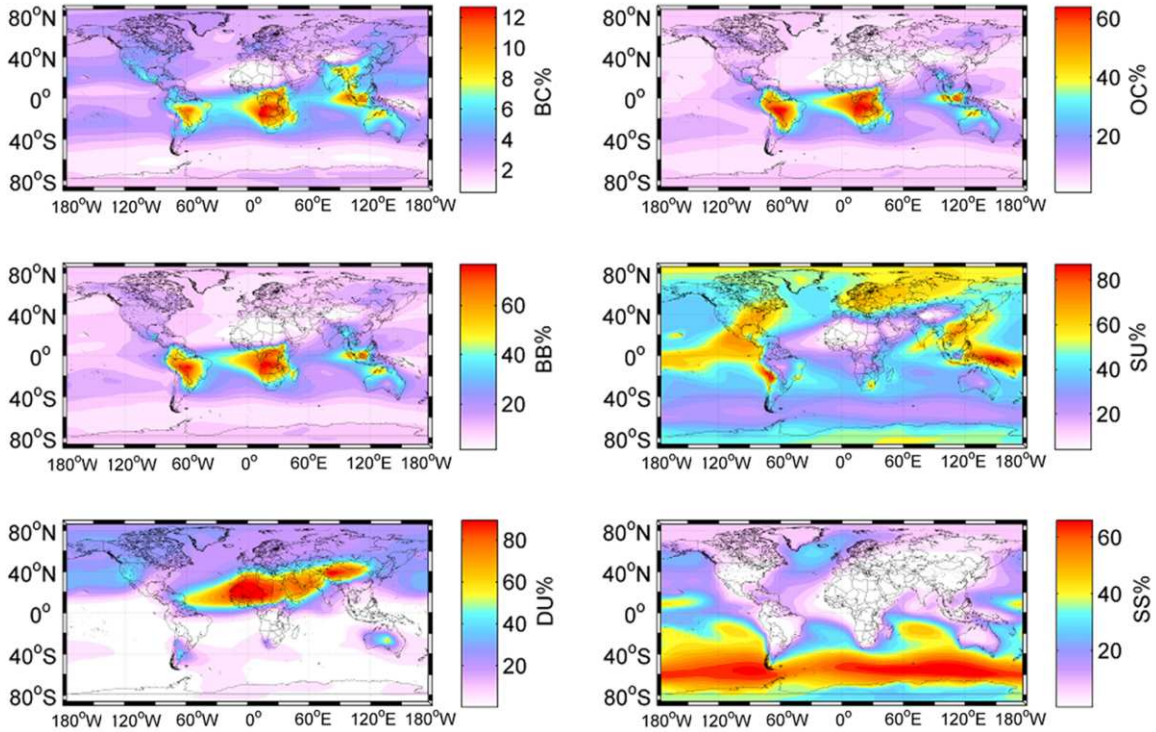


Fig. 3. The multiyear (2000-2006 inclusive) mean percentage contribution of BC, OC, BB=BC+OC, SU, DU and SS to the total AOD (500nm) derived from GOCART model output data. Colour axes are normalized to the maximum value in each case.

Fig. 3 shows that, based on the model simulations, the largest contributions to the mean global AOD are from dust and sulfate (whose colour 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 degrees longitude in extent, situated between 10-40°N. Marine sea salt aerosol appears to be confined strongly to the Southern hemisphere and its peak occupies a band about 20-30 degrees 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 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 centres to every point in the 4D space of percentage contributions. To mitigate the effect of centres being confined to local minima, initial cluster centres were chosen randomly during multiple (we used 10 here) independent runs (for the same number of clusters) as per the prescription of **David & 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 centres. 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.

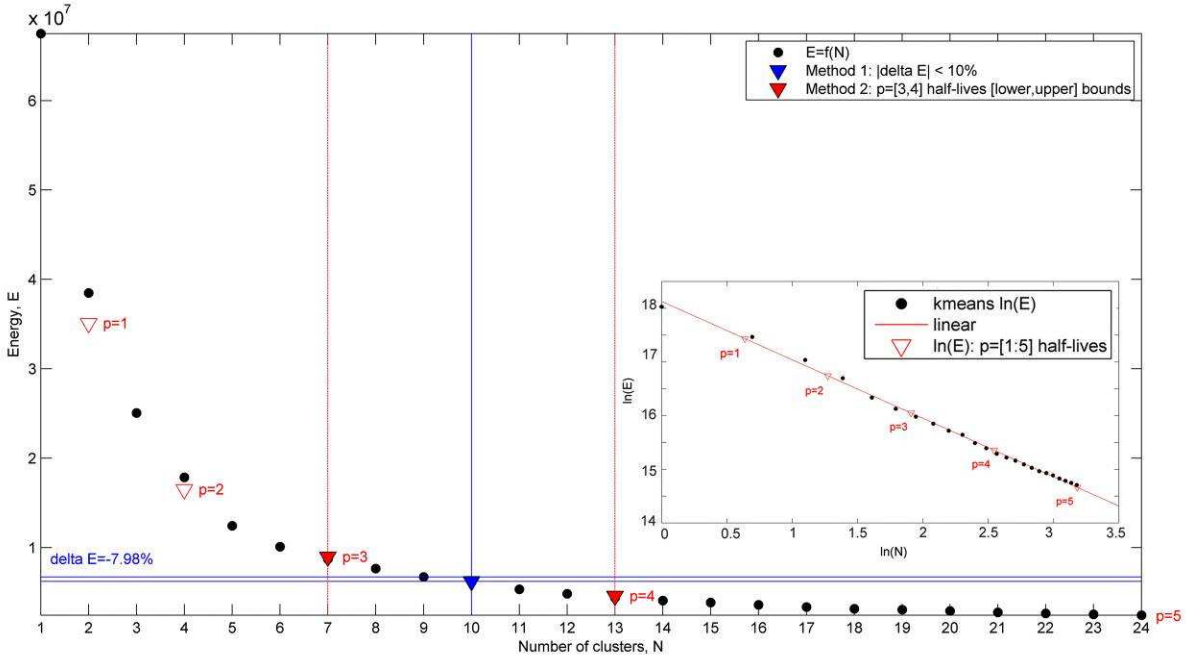


Fig. 4. Graph of the norm of the Euclidean distance ('energy', E) with the number (N) of clusters (1-24) for the multiyear mean percentage contribution of BB, SU, DU and SS to the total AOD (500nm) with points shown as black dots. Red triangles indicate the values of (N,E) associated with 'half-lives' 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.

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 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*.

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	16020	15486	2486	2019	2645	6958	4153	10338	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

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 colour 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 colours and are visually easy to distinguish (i.e. they do not create ambiguities in perception due to colour-blindness for example). With this colour-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:

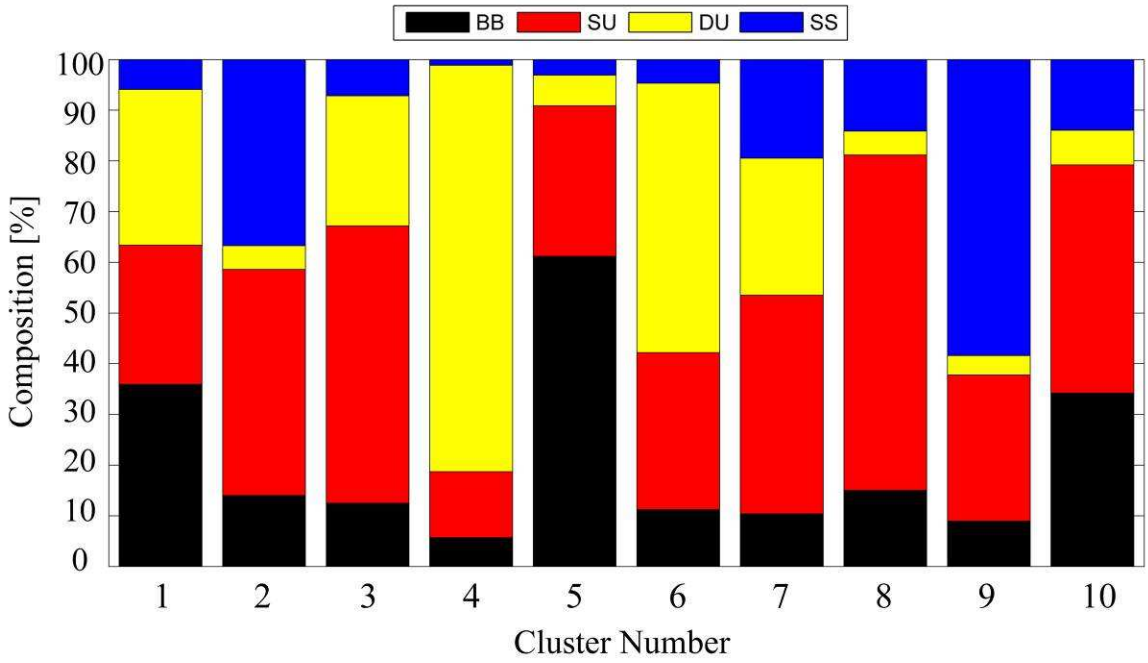


Fig. 5. A (primary) colour-coded stacked bar chart displaying the aerosol composition of each cluster resulting from application of the k-means algorithm and the stopping condition to the global multiyear mean GOCART data.

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."

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**

Stacked bar charts of mean aerosol mixtures obtained at the seasonal timsescale 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 followin 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 occupany (34%). The composition of the southern mid-latitude ocean group (1c) is: 40%SU / 32%SS / 17%BB and 11%DU. The highest occupancy analogue in our study is the "Marine SULFATE" (cluster 2) which has a somewhat lower 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 credenence 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 (1x1 degree) global maps to see how they are spatially distributed worldwide. A challenge here again is the use of colour, and in particular, the creation of a colour scheme for representing mixtures of aerosols. In order to retain a visual link with the colours assigned to pure aerosol types (BB=black, SU=red, DU=yellow and SS=blue) for each cluster, we simply mixed these primary colours 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 colour 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 colour-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.

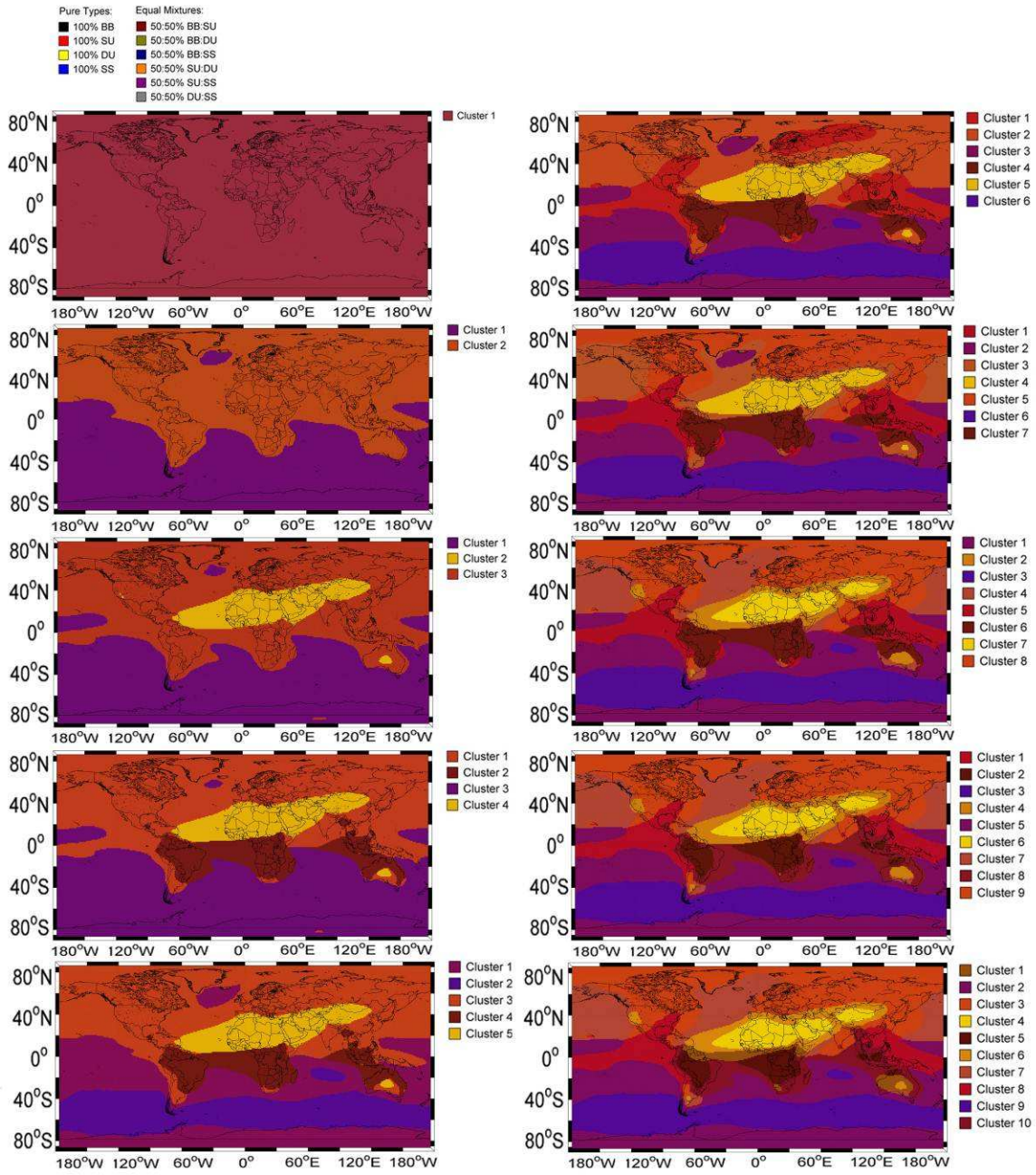


Fig. 6. The evolution of the global partition with the number of clusters for the case of the multiyear global mean. The spatial distribution of aerosol mixtures results from application of the k-means clustering algorithm to the multiyear mean of global GOCART chemical data spanning the period 2000-2006 (inclusive) for a different number (N) of clusters in each case. Note that colours 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.

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 (1x1 degrees) 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.

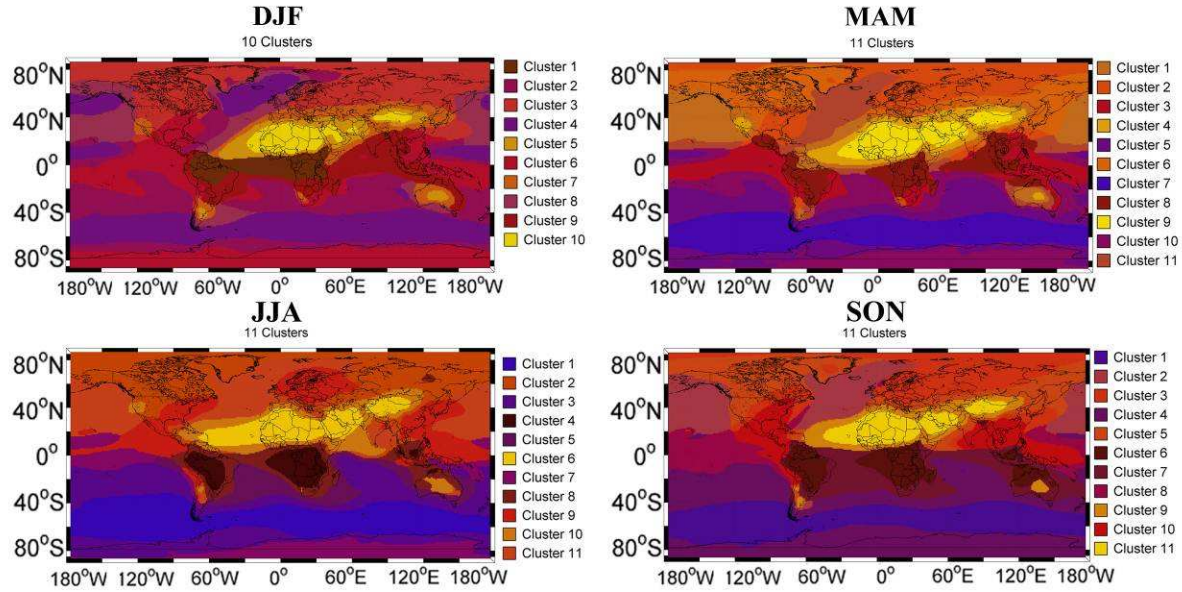


Fig. 7**.** The spatial distribution of aerosol mixtures resulting from application of the k-means clustering algorithm to the seasonal mean of global GOCART chemical data for the monthly triplets: DJF, MAM, JJA and SON.

These maps echo several key features (including the location and extent of dust-dominated aerosol 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**).

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.

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	16020	24.7%	10842	16.7%	8592	13.3%	12154	18.8%	6154	9.5%
Cluster 3	15486	23.9%	12092	18.7%	4296	6.6%	10096	15.6%	8472	13.1%
Cluster 4	2486	3.8%	11315	17.5%	2481	3.8%	1894	2.9%	15186	23.4%

Cluster 5	2019	3.1%	2016	3.1%	9099	14.0%	3194	4.9%	6949	10.7%
Cluster 6	2645	4.1%	12251	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	10338	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	64800	100.0%	64800	100.0%	64800	100.0%	64800	100.0%	64800	100.0%

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.

Table 5. The percentage contribution of BB / SU / DU / SS to the total AOD (500nm) for the multiyear mean (shown in figure 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

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 (1x1 degree) 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 inter-quartile 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" data set 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 (1x1 degree) 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.

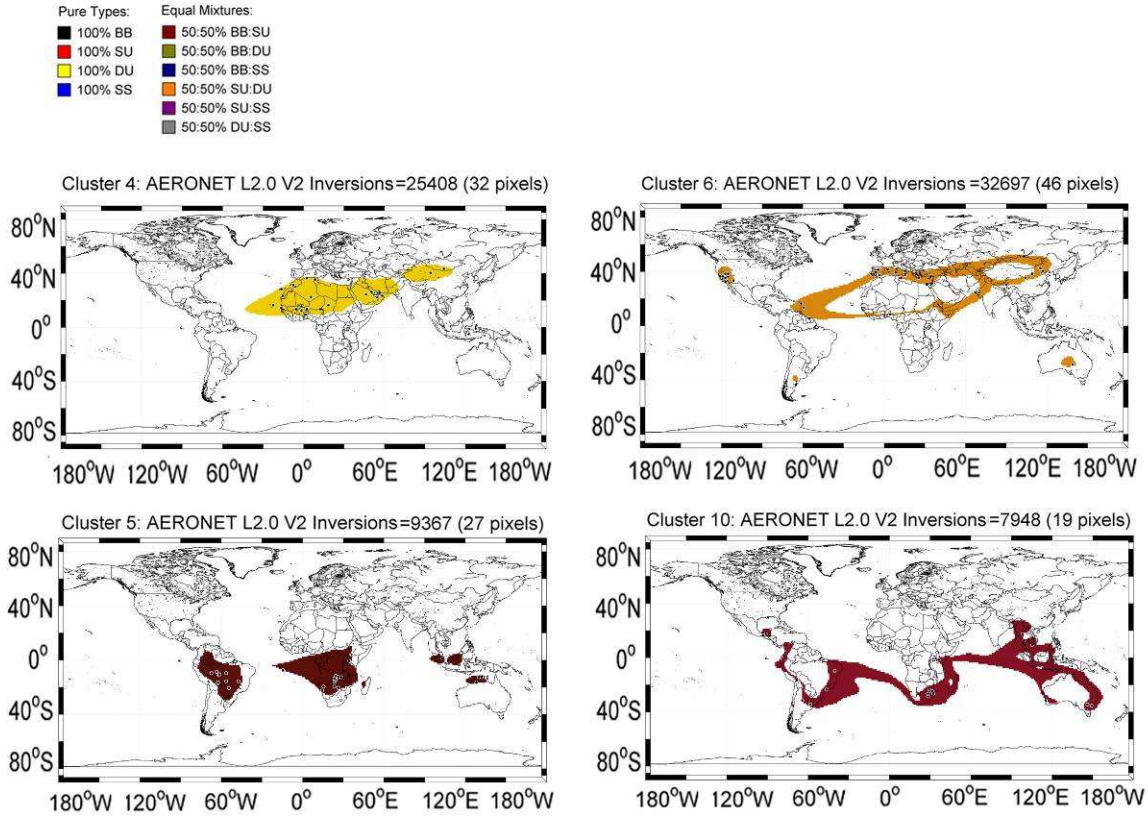


Fig. 8. Maps showing the spatial distribution of 4 aerosol mixtures associated with clusters 4, 5, 6 and 10 coloured according to their compositional mixture, with AERONET sites contributing Level 2.0 Version 2 inversion products wthin each cluster shown as small circles. The total number of complete L2.0 V2 inversion records and the number of gridded 1x1 degree pixels are also shown.

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 degrees (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 1020nm. Although not provided directly by AERONET's inversion algorithm, spectral values of the LR were appended to the extratced dataset using the empirical relationship (**Boyouk et al, 2010**):

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

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 degrees). Retrievals of the SSA and CRI in particular require large aerosol loads where AOD(440nm)>0.4 and four coincident spectral measurements at 440, 675, 870 and 1020nm. Aerosol loading is also important for accurate determination of % Sphericity which requires AOD(440nm)>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.

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.

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	81489	46462	14194	56391	4901	27319	378*	11206
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
H2O	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

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. 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 to 1100nm for each cluster together with a least squares fit using cubic splines. The mean GOCART AOD (500nm) is also plotted for comparison purposes.

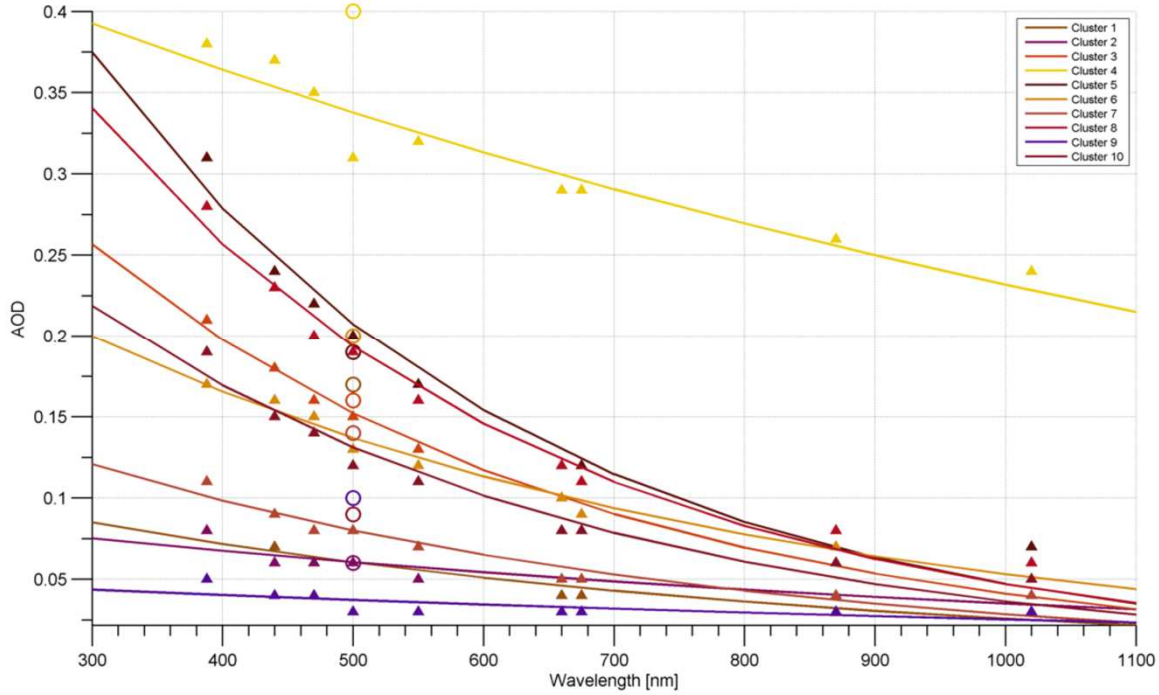


Fig. 9. Optimized least square regression fits to the spectral behaviour of mean values of the AERONET-retrieved AOD for each cluster at the central wavelengths: 380, 440, 470 (extrapolated), 500, 550 (extrapolated), 660 (extrapolated), 675, 870 and 1020nm (triangles). Note that the extrapolated values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

Two features stand out. Firstly, with the exception of clusters 2, 5 and 10, the GOCART AOD (500nm) is much higher than the mean AOD (500nm) 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 500nm 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 (500nm) being much lower than the mean AOD (500nm) 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 behaviour of several optical and microphysical parameters (ASYM, AAOD, SSA and CRI) commonly 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**).

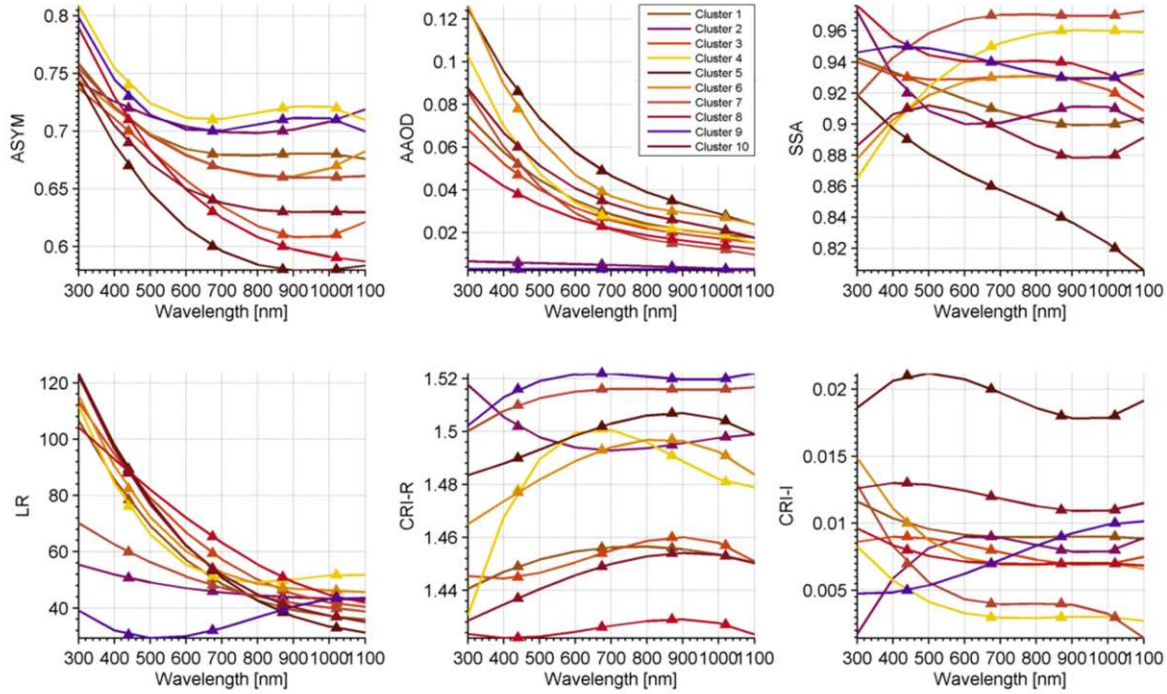


Fig. 10. Spectral behaviour of the global mean values of key optical parameters (ASYM, AAOD, SSA, LR) and microphysical parameters (CRI-R and CRI-I) for each cluster from extracted AERONET data at 440, 675, 870 and 1020nm. Interpolated lines are constructed using cubic spline fits to the extracted data.

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 ≥ 870nm. As described earlier, the AEROENT 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 550nm 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**; **Muller 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.

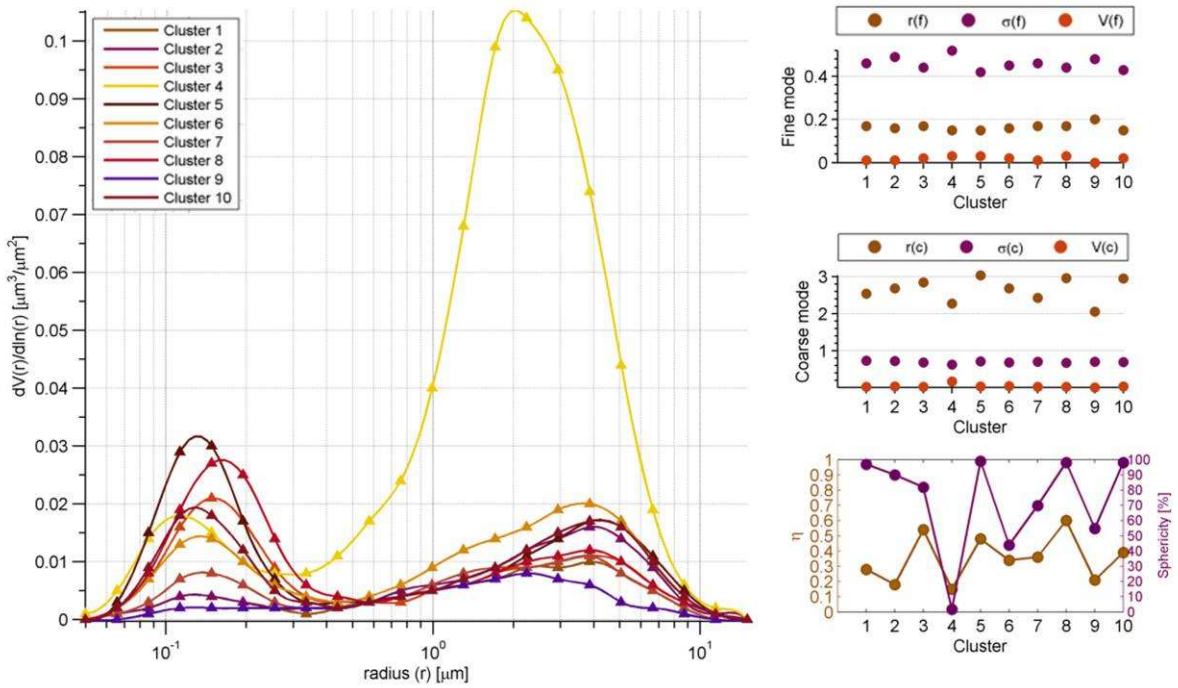


Fig. 11. Global mean size distributions for each cluster together with the derived fine and coarse mode parameters (r_f, σ_f, V_f and r_c, σ_c, V_c respectively). The percentage of spherical particles ('% Sphericity') is plotted with the fine mode fraction (η).

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 (η=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 $43.9 \leq$ % Sphericity ≤ 98.9. Another feature of **Fig. 11** is that the coarse mode peaks appear to cluster around two distinct geometrical radii, the first ($r_c \approx$ 2.0-2.3μm) associated with the DU-dominated and SS-dominated clusters 4 and 9 and the second ($r_c \approx$ 2.4-3.1μm) associated with the other clusters. The location of the fine mode does not share this division and instead is confined to the narrow range: $0.15 \leq r_f \leq 0.20$μm. 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 anymmetry. To the contrary, the behaviour of the AVSD in the coarse mode region (> 0.6μm) either reveals a strong skew (clusters 2, 4 and 5) or the existence of a 'shoulder' on the AVSD in the range 0.6 - 2μm. 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 **Fig. 9**, **Fig. 10** and **Fig. 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**.

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

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

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 (1-2nm) to those reported by **Omar et al (2005)**.

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

In order to ascertain whether there is a correpondence 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 **Table 7** and **Table 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**.

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

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 identifed 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 proption 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 440nm and 1020nm 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 preceeding sections will provide impetus for such studies.

4.3.2 Bivariate classification models

In section 4.2 we saw how the spectral behaviour 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 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.

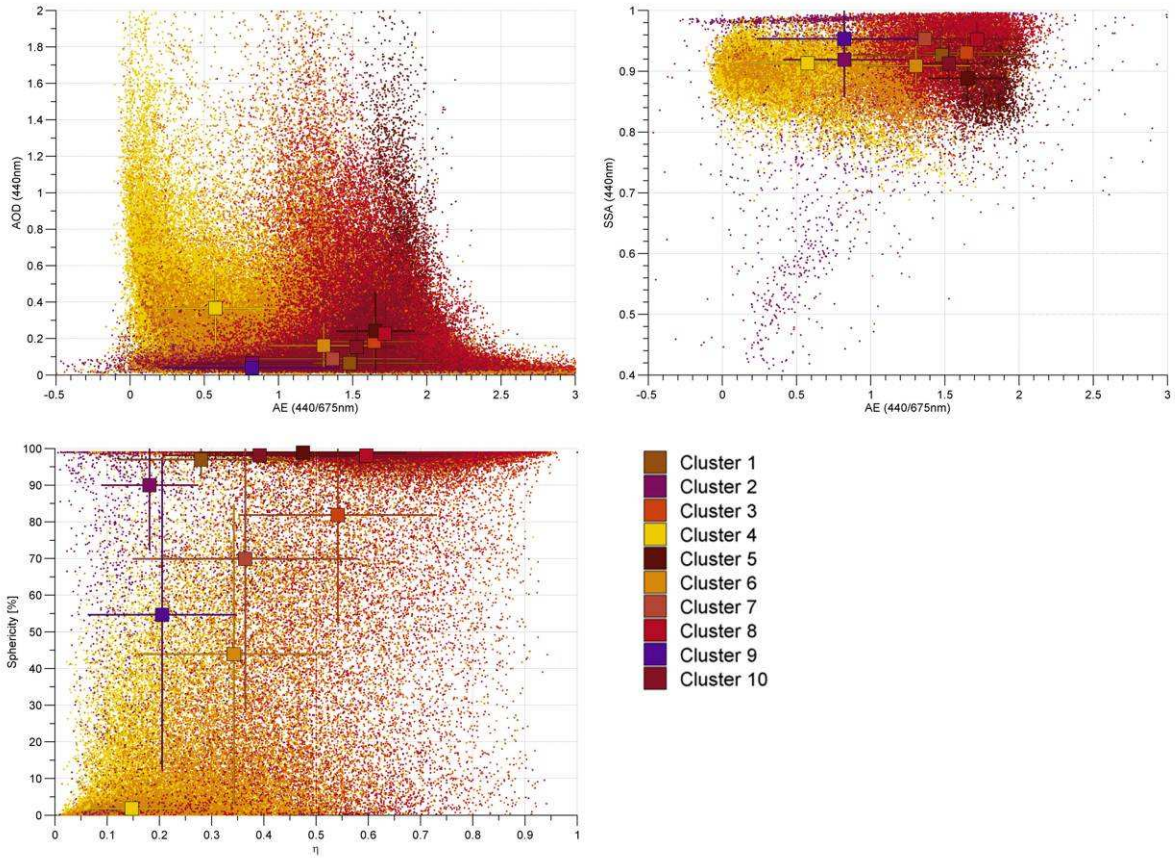


Fig. 12. The location of cluster centers overlaid on AERONET inversion products for two commonly used pair-combination spaces. {AOD (440), AE (440/675)} (*top left panel*), {SSA (440), AE (440/675)} (*top right panel*) and {% Sphericity, η} (*lower panel*) . Error bars extend out to 1 standard deviation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

In the first two cases, although the points representing the AERONET data record suggest the existence of 3-4 discernable 'shaded' regions, the cluster centres 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 line with a recent assessment of global aerosol type retrieved from MISR (**Kahn & 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 observaitons. 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 (500nm) 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, 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 (500nm). 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 centres, 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 1x1 degree 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 prority 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 producsts (**Shi et al., 2011**) identifed and documented a number of potential regions in the American, Aftican and Asian continents in particlar.

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 other to reproduce and extend these findings.

Acknowledgements

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 – KRIPIS" 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.

References

Ansmann A, Müller D, Wandinger U, Mamouri RE. Lidar profiling of aerosol optical and microphysical properties from space: overview, review, and outlook. In: First International Conference on Remote Sensing and Geoinformation of Environment 2013: 879502-879502. International Society for Optics and Photonics.

Arden Pope III C, Dockery DW. Health effects of fine particulate air pollution: lines that connect. J Air Waste Manage Assoc 2006; 56:709–742.

Boyouk N, Léon JF, Delbarre H, Podvin T, Deroo C. Impact of the mixing boundary layer on the relationship between PM2. 5 and aerosol optical thickness. Atmospheric Environment 2010; 44(2):271-277.

Burton SP, Ferrare RA, Vaughan MA, Omar AH, Rogers RR, Hostetler CA, Hair JW. Aerosol classification from airborne HSRL and comparisons with the CALIPSO vertical feature mask. Atmos Meas Tech 2013; 6:1397-1412.

865 Burton SP, Vaughan MA, Ferrare RA, Hostetler CA. Separating mixtures of aerosol types in airborne
866 High Spectral Resolution Lidar data. Atmos Meas Tech 2014; 7:419-436.
867
868 Cattrall C, Reagan J, Thome K, Dubovik O. Variability of aerosol and spectral lidar and backscatter
869 and extinction ratios of key aerosol types derived from selected Aerosol Robotic Network locations. J
870 Geophys Res 2005; Atmospheres (1984–2012):110(D10S11), doi:10.1029/2004JD005124.
871
872 Chen WT, Kahn RA, Nelson D, Yau K, Seinfeld JH. Sensitivity of multiangle imaging to the optical
873 and microphysical properties of biomass burning aerosols. J Geophys Res 2008; 113: D10203.
874
875 Chin M, Jacob DJ, Gardner GM, Foreman-Fowler MS, Spiro PA, Savoie DL. A global three-
876 dimensional model of tropospheric sulfate. J Geophys Res 1996; 101:18667–18690.
877
878 Chin M, Rood RB, Lin SJ, Müller JF, Thompson AM. Atmospheric sulfur cycle simulated in the
879 global model GOCART: Model description and global properties. J Geophys Res 2000; 105(D20):
880 24671-24724.
881
882 Chin M, Ginoux P, Kinne S, Torres O, Holben BN, Duncan BN, Nakajima T. Tropospheric aerosol
883 optical thickness from the GOCART model and comparisons with satellite and Sun photometer
884 measurements. J Atmos Sci 2002; 59(3):461-483.
885
886 Chin M, Diehl T, Ginoux P, Malm W. Intercontinental transport of pollution and dust aerosols:
887 implications for regional air quality. Atmos Chem Phys 2007; 7:5501–5517.
888
889 Chin M, Diehl T, Dubovik O, Eck TF, Holben BN, Sinyuk A, Streets DG. Light absorption by
890 pollution, dust, and biomass burning aerosols: a global model study and evaluation with AERONET
891 measurements. Ann Geophys 2009; 27:3439–3464.
892
893 Chin M, Diehl T, Tan Q, Prospero JM, Kahn RA, Remer LA, Yu H, Sayer AM, Bian H,
894 Geogdzhayev IV, Holben BN, Howell SG, Huebert BJ, Hsu NC, Kim D, Kucsera TL, Levy RC,
895 Mishchenko MI, Pan X, Quinn PK, Schuster GL, Streets DG, Strode SA, Torres O, Zhao X-P. Multi-
896 decadal aerosol variations from 1980 to 2009: a perspective from observations and a global model.
897 Atmos Chem Phys 2014; 14:3657-3690.
898
899 David A, Vassilvitskii S. The Advantages of Careful Seeding: SODA '07. Proceedings of the
900 Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms 2007; 1027–1035.
901
902 Dubovik O, King MD. A flexible inversion algorithm for retrieval of aerosol optical properties from
903 Sun and sky radiance measurements. J Geophys Res 2000; 105(D16):20673-20696.
904
905 Dubovik O, Smirnov A, Holben BN, King MD, Kaufman YJ, Eck TF, Slutsker I. Accuracy
906 assessment of aerosol optical properties retrieval from AERONET sun and sky radiance
907 measurements. J Geophys Res 2000; 105: 9791–9806.
908
909 Dubovik O, Holben B, Eck TF, Smirnov A, Kaufman YJ, King MD, Tanré D, Slutsker I. Variability
910 of absorption and optical properties of key aerosol types observed in worldwide locations. J Atmos
911 Sci 2002; 59:590-608.
912
913 Dubovik O, Sinyuk A, Lapyonok T, Sinyuk A, Mishchenko MI, Yang P, Eck TF, Volten H, Munoz
914 O, Veihelmann B, van der Zander WJ, Sorokin M, Slutsker I. Application of light scattering by
915 spheroids for accounting for particle non-sphericity in remote sensing of desert dust. J Geophys Res
916 2006; 111:D11208, doi:10.1029/2005JD006619.
917

Dubovik O, Herman M, Holdak A, Lapyonok T, Tanre D, Deuze JL, Ducos F, Sinyuk A, Lopatin A. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties from spectral multi-angle polarimetric satellite observations. Atmos Meas Tech 2011; 4:975–1018.

Eck TF, Holben BN, Reid JS, Dubovik O, Smirnov A, O'neill NT, Kinne S. Wavelength dependence of the optical depth of biomass burning, urban, and desert dust aerosols. J Geophys Res 1999; Atmospheres (1984–2012), 104(D24):31333-31349.

Eck TF, Holben BN, Reid JS, Giles DM, Rivas MA, Singh RP, Goloub P. Fog-and-cloud- induced aerosol modification observed by the Aerosol Robotic Network (AERONET). J Geophys Res 2012; Atmospheres (1984–2012):117(D7).

Ginoux P, Chin M, Tegen I, Prospero JM, Holben BN, Dubovik O, Lin SJ. Sources and distributions of dust aerosols simulated with the GOCART model. J Geophys Res 2001; 106(D17):20255-20320.

Gobbi GP, Kaufman YJ, Koren I, Eck TF. Classification of aerosol properties derived from AERONET direct sun data. Atmos Chem Phys 2007; 7(2):453-458.

Granados-Muñoz MJ, Guerrero-Rascado JL, Bravo-Aranda JA, Navas-Guzmán F, Valenzuela A, Lyamani H, Alados-Arboledas L. Retrieving aerosol microphysical properties by Lidar-Radiometer Inversion Code (LIRIC) for different aerosol types. J Geophys Res 2014; Atmospheres, 119(8):4836-4858.

Grubbs FE. Procedures for detecting outlying observations in samples. Technometrics 1969; 11(1):1-21.

Hansen JE, Travis LD. Light scattering in planetary atmospheres. Space Sci Rev 1974; 16(4):527–610.

Hansen J. Earth’s energy imbalance: confirmation and implications. Science 2005; 308:1431–1435.

Higurashi, A., Nakajima, T., Holben, B. N., Smirnov, A., Frouin, R., & Chatenet, B. (2000). A study of global aerosol optical climatology with two-channel AVHRR remote sensing. Journal of Climate, 13(12), 2011-2027.

Higurashi A, Nakajima T. Detection of aerosol types over the East China Sea near Japan from four-channel satellite data. Geophys Res Lett 2002; 29(17):17-21.

Holben BN, Eck TF, Slutsker I, Tanré D, Buis JP et al. AERONET - A federated instrument network and data archive for aerosol characterization. J Rem Sens Environ 1998; 66:1-16.

Holben BN, Eck TF, Slutsker I, Smirnov A, Sinyuk A, Schafer J, Dubovik O. AERONET's version 2.0 quality assurance criteria. In: Asia-Pacific Remote Sensing Symposium 2006; 64080Q, International Society for Optics and Photonics.

Intergovernmental Panel on Climate Change (IPCC). 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, 2013.

Jeong MJ, Li Z. 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 2005; Atmospheres (1984–2012):110(D10).

Kahn RA, Banerjee P, McDonald D. Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean. J Geophys Res 2001; Atmospheres (1984–2012), 106(D16):18219-18238.

Kahn RA, Chen Y, Nelson D, Leung F-Y, Li Q, Diner D, Logan J. Wildfire Smoke Injection Heights: Two perspectives from Space. Geophys Res Lett 2008; 35:L04809, doi:10.1029/2007GL032.

Kahn RA, Gaitley BJ, Garay MJ, Diner DJ, Eck TF, Smirnov A, Holben BN. Multiangle Imaging Spectroradiometer global aerosol product assessment by comparison with the Aerosol Robotic Network. J Geophys Res 2010; 115:D23209, doi:10.1029/2010JD014601.

Kahn RA. Reducing the uncertainties in direct aerosol radiative forcing. Surveys in geophysics 2012; 33(3-4):701-721.

Kahn, RA, and Gaitley, BJ. An analysis of global aerosol type as retrieved by MISR. J Geophys Res Atmos 2015; 120:4248–4281. *doi:* 10.1002/2015JD023322.

Kalashnikova OV, Kahn RA, Sokolik IN, Li W-H. 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 2005; 110, doi: 10.1029/2004JD004550.

Kalashnikova OV, Kahn RA. Ability of multiangle remote sensing observations to identify and distinguish mineral dust types: 2. Sensitivity over dark water. J Geophys Res 2006; Atmospheres (1984–2012):111(D11).

Kaufman YJ, Tanré D, Boucher O. A satellite view of aerosols in the climate system. Nature 2002 (12 September); 419.

Kaufman YJ, Boucher O, Tanré D, Chin M, Remer LA, Takemura T. Aerosol anthropogenic component estimated from satellite data. Geophys Res Lett 2005; 32(17).

Kim SW, Yoon SC, Kim J, Kim SY. Seasonal and monthly variations of columnar aerosol optical properties over east Asia determined from multi-year MODIS, LIDAR, and AERONET Sun/sky radiometer measurements. Atmospheric Environment 2007; 41(8):1634-1651.

Kim MH, Kim SW, Yoon SC, Omar AH. Comparison of aerosol optical depth between CALIOP and MODIS-Aqua for CALIOP aerosol subtypes over the ocean. J Geophys Res 2013; Atmospheres, 118(23):13-241.

Kinne S, O'Donnel D, Stier P, Kloster S, Zhang K, Schmidt H, Stevens B. MAC-v1: A new global aerosol climatology for climate studies. J Adv Mod Earth Systems 2013; 5(4):704-740.

Lamarque J-F, Bond TC, Eyring V, Granier C, Heil A, Klimont Z, Lee D, Liousse C, Mieville A, Owen B, Schultz MG, Shindell D, Smith SJ, Stehfest E, Van Aardenne J, Cooper OR, Kainuma M, Mahowald N, McConnell JR, Naik V, Riahi K, van Vuuren DP. Historical (1850–2000) gridded anthropogenic and biomass burning emissions of reactive gases and aerosols: methodology and application. Atmos Chem Phys 2010; 10:7017–7039.

Lee KH, Li Z, Kim YJ, Kokhanovsky A. Atmospheric aerosol monitoring from satellite observations: a history of three decades. In: Atmospheric and biological environmental monitoring 2009; 13-38, Springer, Netherlands.

Lee J, Kim J, Song CH, Kim SB, Chun Y, Sohn BJ, Holben BN. Characteristics of aerosol types from AERONET sunphotometer measurements. J Atmos Env 2010; 44:3110-3117.

MacQueen JB. Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability 1 1967; 281–297, University of California Press.

Mann GW, Carslaw KS, Reddington CL, Pringle KJ, Schulz M, Asmi A, Hansson HC. Intercomparison and evaluation of global aerosol microphysical properties among AeroCom models of a range of complexity. Atmos Chem Phys 2014; 14(9):4679-4713.

Mattis I, Ansmann A, Müller D, Wandinger U, Althausen D. Dual-wavelength Raman lidar observations of the extinction-to-backscatter ratio of Saharan dust. Geophys Res Lett 2002; 29(9):1306, doi:10.1029/2002GL0147.

Mielonen T, Arola A, Komppula M, Kukkonen J, Koskinen J, de Leeuw G, Lehtinen KEJ. Comparison of CALIOP level 2 aerosol subtypes to aerosol types derived from AERONET inversion data. Geophys Res Lett 2009; 36:L18804, doi:10.1029/2009GL039609.

Millennium Ecosystem Assessment. Millennium ecosystem assessment: living beyond our means: natural assets and human well-being (statement from the Board). Millennium Ecosystem Assessment 2005; India, France, Kenya, UK, USA, Netherlands, Malaysia, http://www.millenniumassessment.org/: 1-28.

Mishchenko MI, Cairns B, Kopp G, Schueler CF, Fafaul BA, Hansen JE, Hooker RJ, Itchkawich T, Maring HB, Travis LD. Accurate monitoring of terrestrial aerosols and total solar irradiance: Introducing the Glory mission. Bull Amer Meteorol Soc 2007; 88:677-691.

Mishchenko MI, Geogdzhayev IV, Rossow WB, Cairns B, Carlson BE, Lacis AA, Liu L, Travis LD. Long-term satellite record reveals likely recent aerosol trend. Science 2007; 315:1543, doi:10.1126/science.1136709.

Mishchenko MI, Liu L, Geogdzhayev IV, Li J, Carlson BE, Lacis AA, Cairns B, Travis LD. Aerosol retrievals from channel-1 and -2 AVHRR radiances: Long-term trends updated and revisited. J Quant Spectrosc Rad Trans 2012; 113:1974–1980.

Mishchenko MI, Geogdzhayev IV, Cairns B, Carlson BE, Chowdhary J, Lacis AA, Liub L, Rossow WB, Larry D, Travis LD. Past, present, and future of global aerosol climatologies derived from satellite observations: a perspective. J Quant Spectrosc Rad Trans 2007; 106:325–347.

Omar AH, Won JG, Winker DM, Yoon SC, Dubovik O, McCormick MP. Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements. J Geophys Res 2005; Atmospheres (1984–2012):110(D10).

Omar AH, Winker DM, Kittaka C, Vaughan MA, Liu Z, Hu Y, Trepte CR, Rogers RR, Ferrare RA, Lee KP, Kuehn RE, Hostetler CA. The CALIPSO automated aerosol classification and Lidar ratio selection algorithm. J Atmos Oceanic Tech 2009; 26:1994-2014.

Quinn PK, Shaw G, Andrews E, Dutton EG, Ruoho-Airola T, Gong SL. Arctic haze: current trends and knowledge gaps. Tellus 2007; 59:99–114.

Penning de Vries MJM, Beirle S, Hörmann C, Tilstra LG, Tuinder ONE, Stammes P, Wagner T. Towards a global aerosol type climatology from MODIS and GOME-2 observations of aerosols and trace gases. Poster: 2012 EUMETSAT Meteorological Satellite Conference, 3/9/2012-7/9/2012, Sopot, Poland, EUMETSAT.

Ramanathan V, Crutzen PJ, Kiehl JT, Rosenfeld D. Aerosols, climate, and the hydrological cycle. Science 2001; 294:2119–2124.

Remer LA, Kaufman YJ, Tanré D, Mattoo S, Chu DA, Martins JV, Holben BN. The MODIS aerosol algorithm, products, and validation. J Atmos Sci 2005; 62(4):947-973.

Remer LA, Kleidman RG, Levy RC, Kaufman YJ, Tanré D, Mattoo S, Martins JV, Ichoku C, Koren I, Yu H, Holben BN. Global aerosol climatology from the MODIS satellite sensors. J Geophys Res 2008; Atmospheres 113: D14S07, doi:10.1029/2007JD009661.

Russell PB, Bergstrom RW, Shinozuka Y, Clarke AD, DeCarlo PF, Jimenez JL, Strawa A. Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition. Atmos Chem Phys 2010; 10(3):1155-1169.

Russell PB, et al. A multiparameter aerosol classification method and its application to retrievals from space-borne polarimetry. J Geophys Res 2014; Atmospheres 119, doi:10.1002/2013JD021411.

Samset BH, Myhre G, Schulz M. Upward adjustment needed for aerosol radiative forcing uncertainty. Nat Clim Change 2014; 4(4):230–232, doi:10.1038/nclimate2170.

Sayer AM, Smirnov A, Hsu NC, Holben BN. A pure marine aerosol model, for use in remote sensing applications. J Geophys Res 2012; 117: 1-25.

Schwartz SE. Uncertainty requirements in radiative forcing of climate change. J Air Waste Manag Assoc 2004; 54:1351–1359.

Schwartz SE, Charlson RJ, Kahn RA, Rodhe H. Earth's climate sensitivity: Apparent Inconsistencies in recent analyses. Earth's Future 2014; doi:10.1002/2014EF000273.

Seinfeld JH, Pandis SN. Atmospheric Chemistry and Physics: From Air Pollution to Climate Change (1st edition). Wiley, New York, 1998.

Shi Y, Zhang J, Reid JS, Hyer EJ, Eck TF, Holben BN, Kahn RA. A critical examination of spatial biases between MODIS and MISR aerosol products – application for potential AERONET deployment. Atmos Meas Tech 2011; 4:2823-2836, doi:10.5194/amt-4-2823-2011.

Smirnov A, Holben BN, Eck TF, Dubovik O, Slutsker I. Cloud screening and quality control algorithms for the AERONET database. Rem Sens Envir 2000; 73:337-349.

Smirnov A, Holben BN, Giles DM, Slutsker I, O'Neill NT, Eck TF, Macke A, Croot P, Courcoux Y, Sakerin SM, Smyth TJ, Zielinski T, Zibordi G, Goes JI, Harvey MJ, Quinn PK, Nelson NB, Radionov VF, Duarte CM, Losno R, Sciare J, Voss KJ, Kinne S, Nalli NR, Joseph E, Krishna Moorthy K, Covert DS, Gulev SK, Milinevsky G, Larouche P, Belanger S, Horne E, Chin M, Remer LA, Kahn RA, Reid JS, Schulz M, Heald CL, Zhang J, Lapina K, Kleidman RG, Griesfeller J, Gaitley BJ, Tan Q, Diehl TL. Maritime aerosol network as a component of AERONET – first results and comparison with global aerosol models and satellite retrievals. Atmos Meas Tech 2011; 4:583–597, doi:10.5194/amt-4-583-2011.

Stier P, Schutgens NAJ, Bellouin N, Bian H, Boucher O, Chin M, Zhou C. Host model uncertainties in aerosol radiative forcing estimates: results from the AeroCom Prescribed intercomparison study. Atmos Chem Phys 2013; 13(6):3245-3270.

Taylor M, Kazadzis S, Gerosopoulos E. Multi-modal analysis of aerosol robotic network size distributions for remote sensing applications: dominant aerosol type cases. Atmos Meas Tech 2014; 7:839-858.

Tomasi C, Kokhanovsky AA, Lupi A, Ritter C, Smirnov A, O'Neill NT, Piketh SJ. Aerosol remote sensing in polar regions. Earth-Science Reviews 2014.

1137 Tsekeri A, Amiridis V, Kokkalis P, Basart S, Chaikovsky A, Dubovik O, Gross B. Application of a
1138 synergetic lidar and sunphotometer algorithm for the characterization of a dust event over Athens,
1139 Greece. British J Envir Clim Change 2013; 3(4):531-546.
1140
1141 Val Martin M, Logan JA, Kahn RA, Leung F-Y, Nelson D, Diner D. Fire smoke injection heights
1142 over North America constrained from the Terra Multi-angle Imaging SpectroRadiometer. Atmos
1143 Chem Phys 2010; 10:1491-1510.
1144
1145 Wang Z, Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to
1146 structural similarity. IEEE Trans Imag Process 2004; 13(4):600-612.
1147
1148 Winker DM, Tackett JL, Getzewich BJ, Liu Z, Vaughan MA, Rogers RR. The global 3-D distribution
1149 of tropospheric aerosols as characterized by CALIOP. Atmos Chem Phys 2013; 13(6):3345-3361.
1150
1151 Xia X. Variability of aerosol optical depth and Angstrom wavelength exponent derived from
1152 AERONET observations in recent decades. Envir Res Lett 2011; 6:44011-44020.
1153
1154 Yu H, Remer LA, Chin M, Bian H, Kleidman RG, Diehl T. A satellite-based assessment of
1155 transpacific transport of pollution aerosol. J Geophys Res 2008; 113:D14S12,
1156 doi:10.1029/2007JD009349.
1157
1158 Yu H, Chin M, Winker DM, Omar AH, Liu Z, Kittaka C, Diehl T. Global view of aerosol vertical
1159 distributions from CALIPSO lidar measurements and GOCART simulations: Regional and seasonal
1160 variations. J Geophys Res 2010; Atmospheres (1984–2012):115(D4).
1161
1162 Yu H, Chin M, West JJ, Atherton CS, Bellouin N, Bergmann D, Bey I, Bian H, Diehl T, Forberth G,
1163 Hess P, Schulz M, Shindell D, Takemura T, Tan Q. A multimodel assessment of the influence of
1164 regional anthropogenic emission reductions on aerosol direct radiative forcing and the role of
1165 intercontinental transport. J Geophys Res 2013; 118:700–720, doi:10.1029/2012JD018148, 2013.
1166
1167 Zhang J, Reid JS. A decadal regional and global trend analysis of the aerosol optical depth using a
1168 data-assimilation grade over-water MODIS and Level 2 MISR aerosol products. Atmos Chem Phys
1169 2010; 10: 10949–10963, doi:10.5194/acp-10-10949-2010.
1170

Highlights for review

Global aerosol mixtures and their multiyear and seasonal characteristics

Taylor, M.[1]*, Kazadzis, S.[2], Amiridis V.[3], Kahn, R.A.[4]

[1] Institute for Environmental Research and Sustainable Development (IERSD), National Observatory of Athens (NOA), Metaxa & Vas Pavlou, Penteli, 15236, Athens, Greece.

[2] Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center (PMOD/WRC), Dorfstrasse 33, 7260 Davos Dorf, Switzerland.

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

[4] NASA Goddard Space Flight Centre (GSFC), Greenbelt, MD 20771, U.S.A.

[*] Correspondence: Tel: +30 2108109170 / Fax : +30 2108103236 / Email address: mtaylor@noa.gr

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 classication studies

Supplement of

Global aerosol mixtures and their multiyear and seasonal characteristics

Taylor, M.[1]*, Kazadzis, S.[2], Amiridis V.[3], Kahn, R.A.[4]

[1] Institute for Environmental Research and Sustainable Development (IERSD), National Observatory of Athens (NOA), Metaxa & Vas Pavlou, Penteli, 15236, Athens, Greece.

[2] Physikalisch-Meteorologisches Observatorium Davos, World Radiation Center (PMOD/WRC), Dorfstrasse 33, 7260 Davos Dorf, Switzerland.

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

[4] NASA Goddard Space Flight Centre (GSFC), Greenbelt, MD 20771, U.S.A.

* Correspondence to: mtaylor@noa.gr / Tel: +30 2108109170 / Fax : +30 2108103236

In order to facilitate the uptake of the findings of this work, spreadsheets (EXCEL) of gridded (1x1 degree) cluster values at the multiyear and seasonal timescales are made freely available at: http://apcg.space.noa.gr/aeromap/. In particular,

- Gridded (1x1 degree) cluster indices for multiyear mean global aerosol mixtures
- Gridded (1x1 degree) cluster indices for seasonal mean global aerosol mixtures: DJF, MAM, JJA and SON.
- A table of descriptive statistics (mean, standard deviation, median and inter-quartile range) of the 5-component (BC, OC, SU, DU and SS) percentage composition for each cluster of the multiyear mean global partition.
- Tables of descriptive statistics (mean, standard deviation, median and inter-quartile range) of the 5-component (BC, OC, SU, DU and SS) percentage composition for each cluster of seasonal mean global partitions: DJF, MAM, JJA and SON.
- A table of 50 optical and microphysical aerosol parameters extracted from the AERONET inversion data record for each cluster of the mean multiyear global partition including: spectral parameters (AOD, ASYM, AAOD, SSA, CRI-R, CRI-I, LR), the mean percentage of spherical particles ('%Sphericity'), the mean AVSD (in 22 logarithmic radial bins from 0.05μm to 15 μm), AVSD-derived microphysical parameters (r_f, σ_f, V_f, r_c, σ_c, V_c associated with the fine and coarse modes 'f''and 'c' respectively), and the fine mode fraction η.
- Tables of 50 optical and microphysical aerosol parameters extracted from the AERONET inversion data record for each cluster of the mean seasonal (DJF, MAM, JJA and SON) global partitions including: spectral parameters (AOD, ASYM, AAOD, SSA, CRI-R, CRI-I, LR), the mean percentage of spherical particles ('%Sphericity'), the mean AVSD (in 22 logarithmic radial bins from 0.05μm to 15 μm), AVSD-derived microphysical parameters (r_f, σ_f, V_f, r_c, σ_c, V_c associated with the fine and coarse modes 'f''and 'c' respectively), and the fine mode fraction η.

S1. The 2000-2006 mean global AOD (500nm) as a proxy

In this section we present the results of comparing the mean total global AOD (500nm) with annual means for the study period (2000-2006 inclusive).

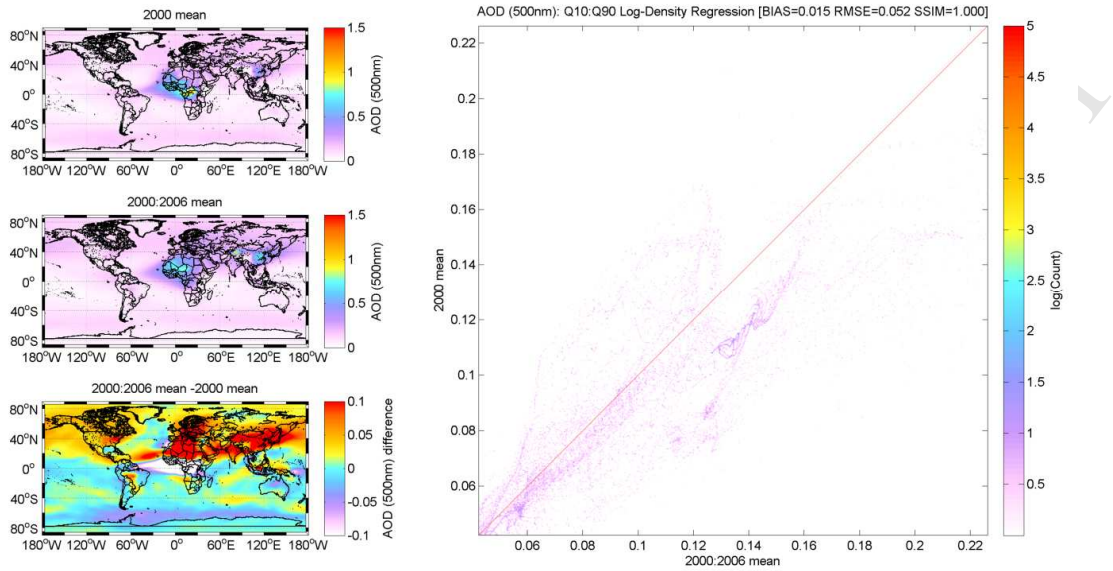


Fig. S1. *left panels*: Maps of the mean total global AOD (500nm) for the year 2000 (top), 2000-2006 inclusive (centre) and the difference (bottom). *right panel*: a log-density plot of the year 2000 mean total AOD (500nm) against the 2000-2006 (inclusive) mean. 'RMSE' is the root mean squared error and 'SSIM' is the value of the mean structural similarity index over all 64800 gridded (1x1 degree) pixels.

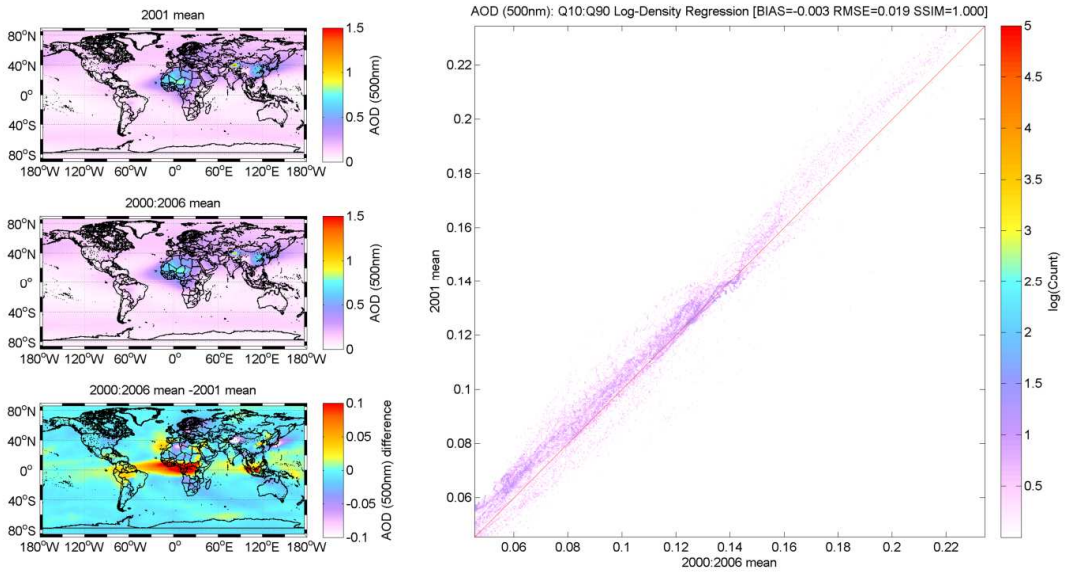


Fig. S2. *left panels*: Maps of the mean total global AOD (500nm) for the year 2001 (top), 2000-2006 inclusive (centre) and the difference (bottom). *right panel*: a log-density plot of the year 2001 mean total AOD (500nm) against the 2000-2006 (inclusive) mean. 'RMSE' is the root mean squared error

and ‘SSIM’ is the value of the mean structural similarity index over all 64800 gridded (1x1 degree) pixels.

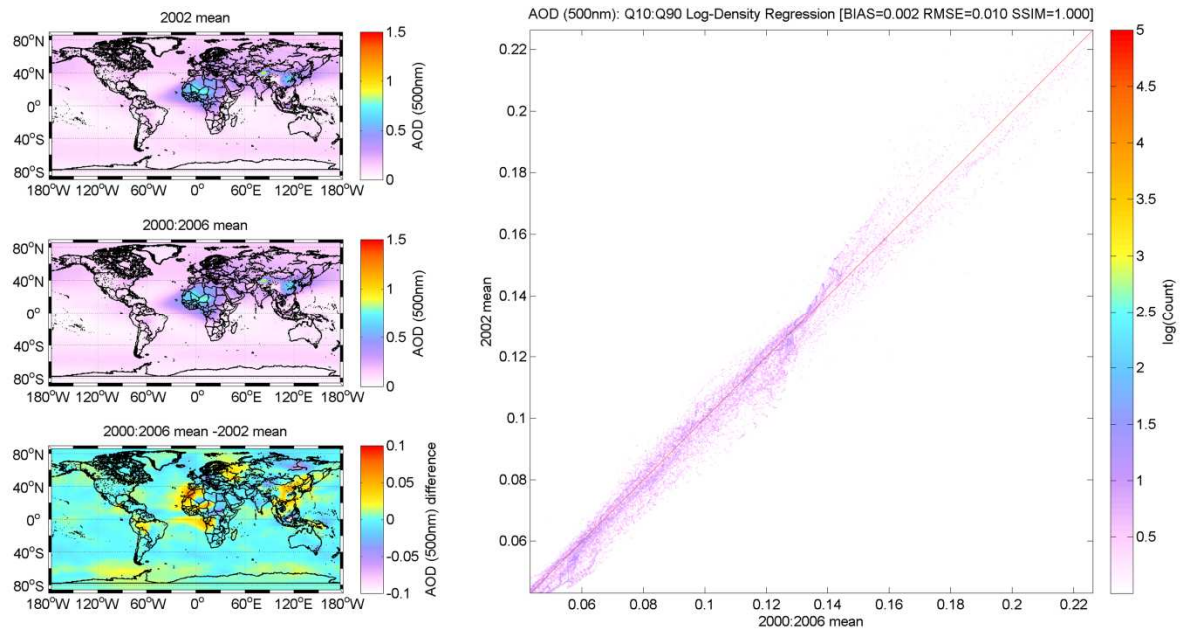


Fig. S3. *left panels*: Maps of the mean total global AOD (500nm) for the year 2002 (top), 2000-2006 inclusive (centre) and the difference (bottom). *right panel*: a log-density plot of the year 2002 mean total AOD (500nm) against the 2000-2006 (inclusive) mean. ‘RMSE’ is the root mean squared error and ‘SSIM’ is the value of the mean structural similarity index over all 64800 gridded (1x1 degree) pixels.

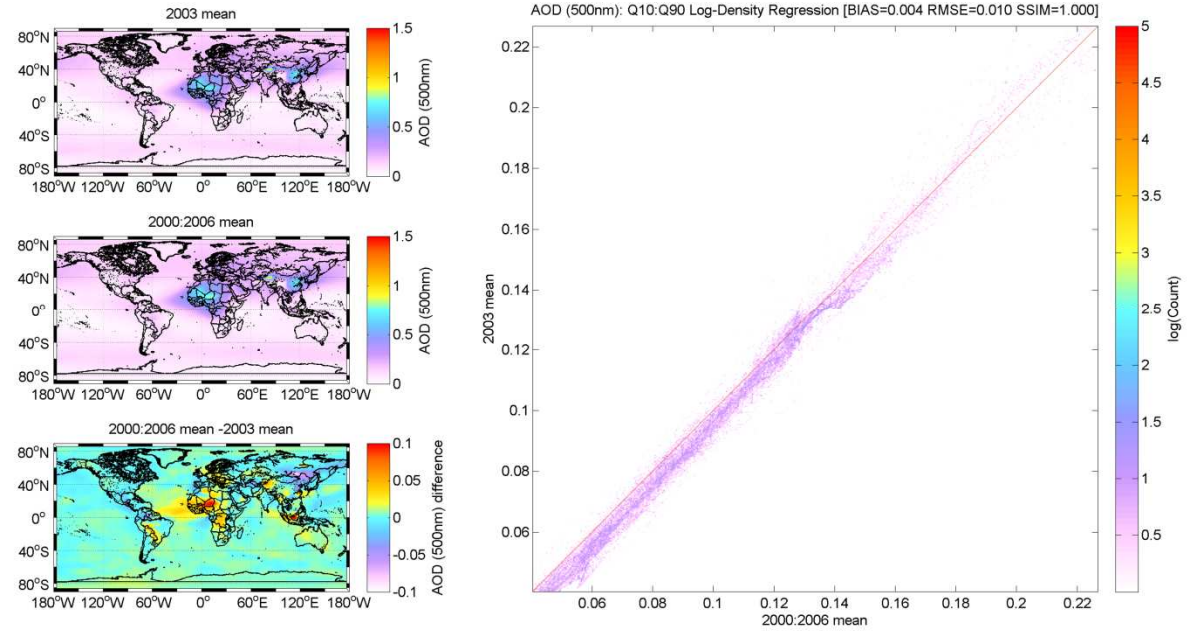


Fig. S4. *left panels*: Maps of the mean total global AOD (500nm) for the year 2003 (top), 2000-2006 inclusive (centre) and the difference (bottom). *right panel*: a log-density plot of the year 2003 mean total AOD (500nm) against the 2000-2006 (inclusive) mean. ‘RMSE’ is the root mean squared error

and 'SSIM' is the value of the mean structural similarity index over all 64800 gridded (1x1 degree) pixels.

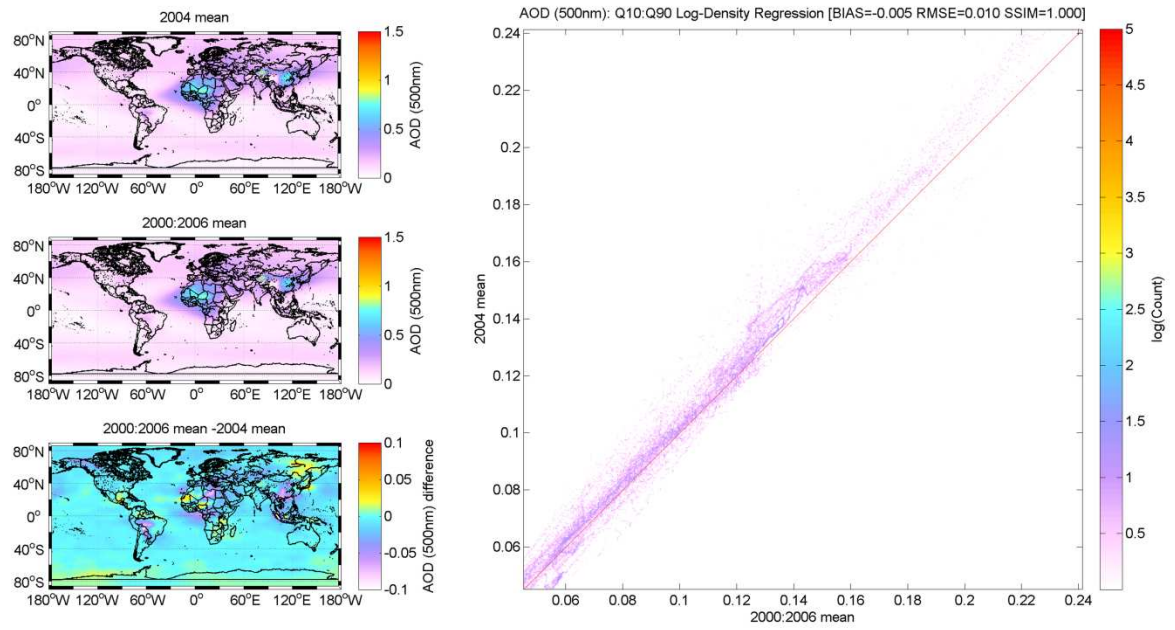


Fig. S5. *left panels*: Maps of the mean total global AOD (500nm) for the year 2004 (top), 2000-2006 inclusive (centre) and the difference (bottom). *right panel*: a log-density plot of the year 2004 mean total AOD (500nm) against the 2000-2006 (inclusive) mean. 'RMSE' is the root mean squared error and 'SSIM' is the value of the mean structural similarity index over all 64800 gridded (1x1 degree) pixels.

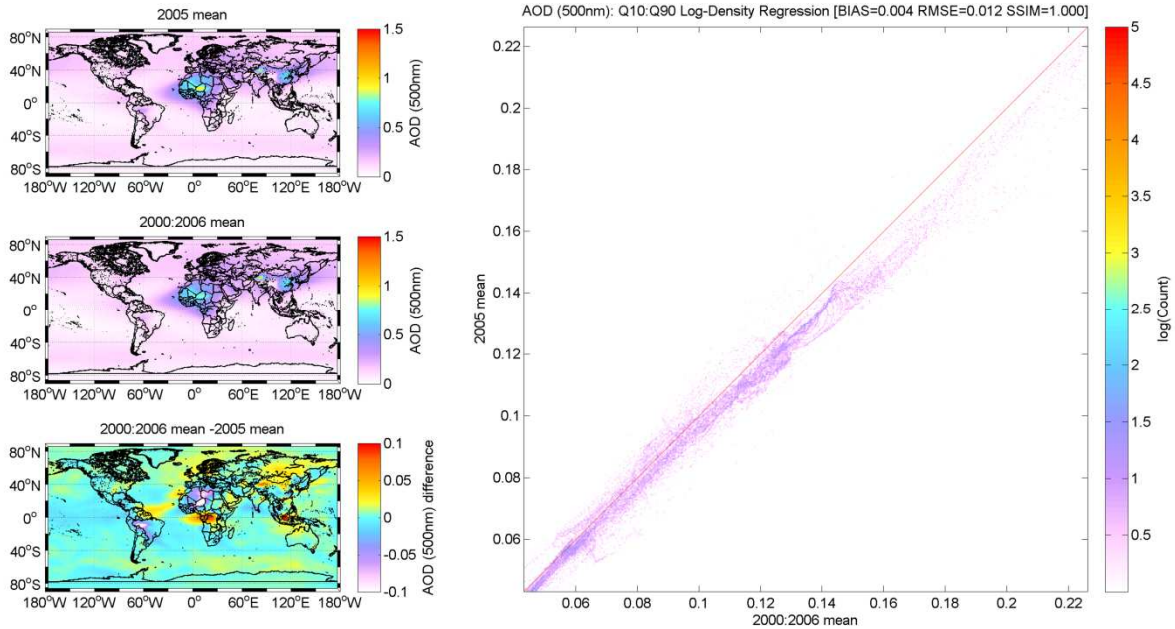


Fig. S6. *left panels*: Maps of the mean total global AOD (500nm) for the year 2005 (top), 2000-2006 inclusive (centre) and the difference (bottom). *right panel*: a log-density plot of the year 2005 mean total AOD (500nm) against the 2000-2006 (inclusive) mean. 'RMSE' is the root mean squared error

and ‘SSIM’ is the value of the mean structural similarity index over all 64800 gridded (1x1 degree) pixels.

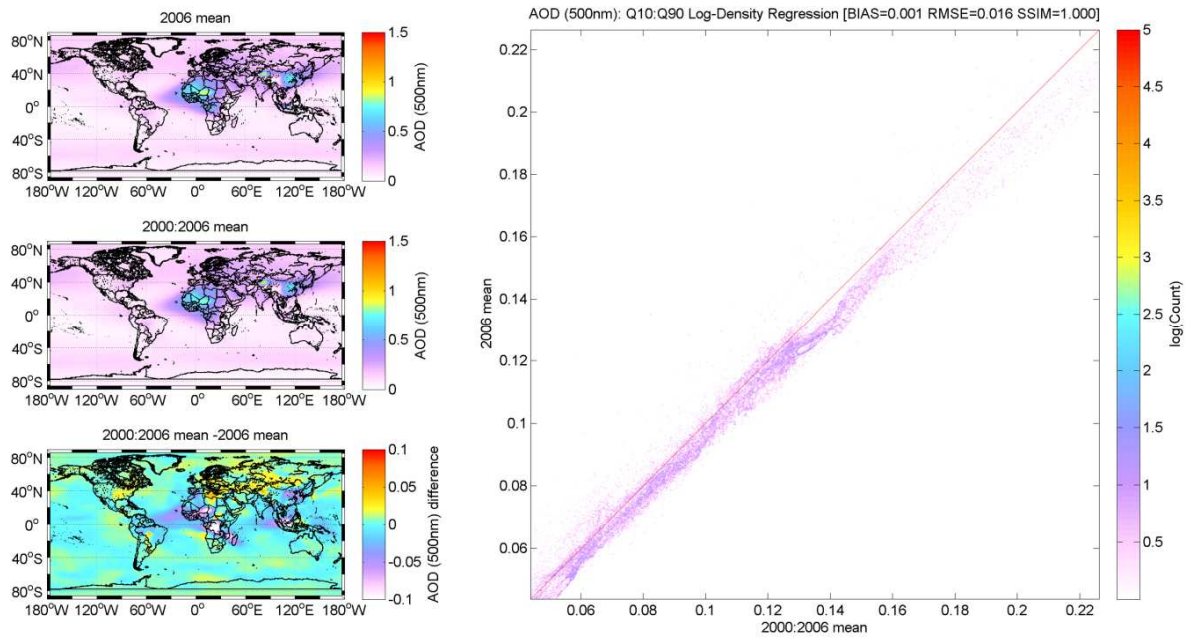


Fig. S7. *left panels*: Maps of the mean total global AOD (500nm) for the year 2006 (top), 2000-2006 inclusive (centre) and the difference (bottom). *right panel*: a log-density plot of the year 2006 mean total AOD (500nm) against the 2000-2006 (inclusive) mean. ‘RMSE’ is the root mean squared error and ‘SSIM’ is the value of the mean structural similarity index over all 64800 gridded (1x1 degree) pixels.

S2. Cluster analysis of the multiyear mean

In this section we present the descriptive statistics of the clusters obtained by applying the cluster analysis technique described in Section 3.1 to the multiyear GOCART data, as well as bar charts of the composition and the resulting taxonomy.

Table S1. Descriptive statistics of the global mean mixtures extracted from the multiyear 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	16020	15486	2486	2019	2645	6958	4153	10338	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
%BB (Median)	35.8	12.4	11.5	4.9	60.6	9.3	9.6	14.4	8.2	33.3
%SU (Median)	27.2	44.6	54.8	13.2	28.8	31.1	42.6	64.5	28.4	44.3
%DU (Median)	29.8	4.3	25.4	79.6	3.8	53.5	27.5	3.0	3.5	5.3
%SS (Median)	5.7	36.8	7.5	0.4	2.0	2.2	19.0	13.2	58.9	12.7
%BC (Median)	5.8	2.6	3.2	1.6	9.4	2.8	3.0	3.0	1.5	6.6
%OC (Median)	29.9	9.8	8.3	3.2	51.6	6.4	6.6	11.1	6.7	26.5
%BB (IQR)	10.7	5.2	2.9	3.4	14.5	6.9	2.5	7.3	3.0	10.2
%SU (IQR)	7.6	10.9	6.6	9.0	9.6	11.7	6.0	7.9	6.6	8.6
%DU (IQR)	12.2	1.3	6.3	13.7	6.5	13.2	9.9	5.3	1.5	7.0
%SS (IQR)	4.7	9.1	3.4	0.6	3.0	8.4	6.9	15.6	11.2	13.5
%BC (IQR)	1.7	1.4	0.8	1.2	2.6	1.5	0.9	1.6	0.6	1.5
%OC (IQR)	9.9	3.8	2.3	2.4	12.6	4.9	1.7	5.9	2.4	9.0

Fig. S8 below shows stacked bar charts of mean aerosol mixtures obtained at the seasonal timsescale:

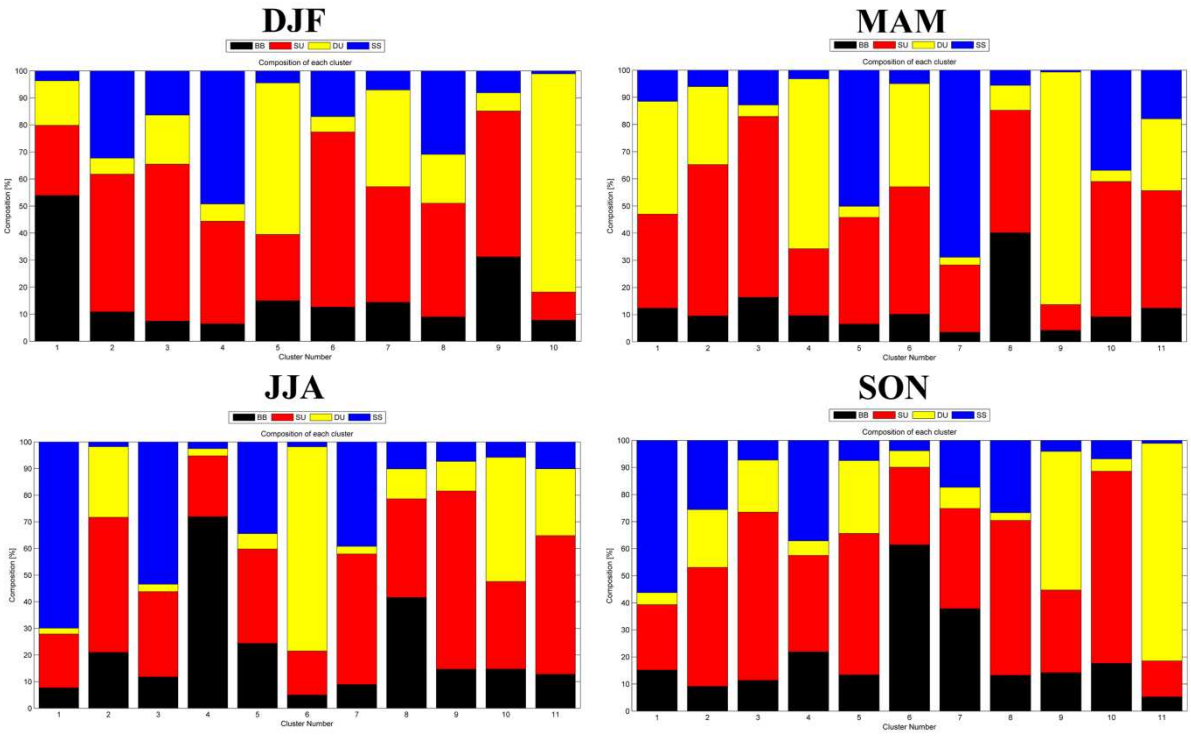


Fig. S8. Colour-coded stacked bar charts displaying the aerosol composition of each cluster for the global mean GOCART data at the seasonal timescale as represented by the monthly triplets: DJF, MAM, JJA and SON.

Applying the naming convention logic described in Section 3.2 to these 4 partitions leads to the following taxonomy:

Table. S2. Taxonomy of aerosol mixtures for the seasonal global mean partitions. 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. For cases where more than one aerosol type is dominant (to within 1 standard deviation) then these are highlighted as multiple 'objects'.

	DJF	MAM	JJA	SON
Cluster 1	Dusty Sulfurous **SMOKE**	SULFUROUS **DUST**	**MARINE**	Sulfurous **MARINE**
Cluster 2	Marine **SULFATE**	Dusty **SULFATE**	Smokey Dusty **SULFATE**	Dusty Marine **SULFATE**
Cluster 3	Dusty **SULFATE**	**SULFATE**	Sulfurous **MARINE**	Dusty **SULFATE**
Cluster 4	Sulfurous **MARINE**	Sulfurous **DUST**	Sulfurous **SMOKE**	Smokey **SULFUROUS MARINE**
Cluster 5	Sulfurous **DUST**	**SULFUROUS MARINE**	Smokey **MARINE SULFATE**	Dusty **SULFATE**
Cluster 6	**SULFATE**	**DUSTY SULFATE**	**DUST**	Sulfurous **SMOKE**
Cluster 7	DUSTY **SULFATE**	Sulfurous **MARINE**	Marine **SULFATE**	**SULFUROUS SMOKE**
Cluster 8	Dusty Marine **SULFATE**	Smokey **SULFATE**	**SULFUROUS SMOKE**	Marine **SULFATE**
Cluster 9	Smokey **SULFATE**	**DUST**	**SULFATE**	Sulfurous **DUST**
Cluster 10	**DUST**	Marine **SULFATE**	Sulfurous **DUST**	**SULFATE**
Cluster 11		Marine Dusty **SULFATE**	Dusty **SULFATE**	**DUST**

S3. Cluster analysis of the seasonal means

In this section we present the results of applying the cluster analysis technique described in Section 3.1 to the multiyear GOCART data extracted for each season (DJF, MAM, JJA and SON).

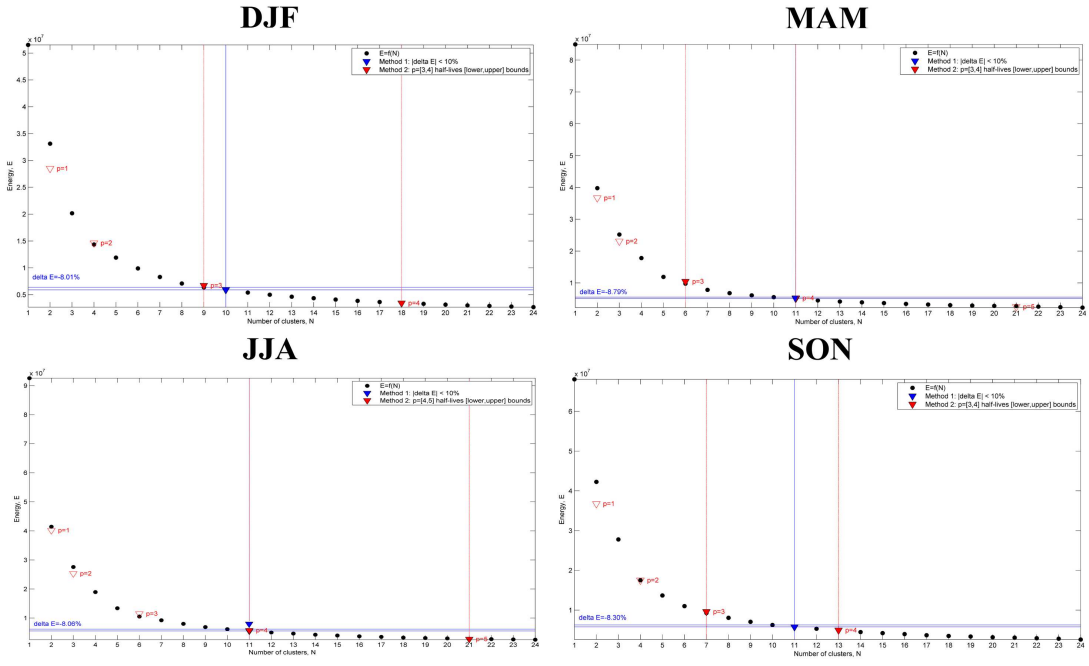


Fig. S9. Graphs of the norm of the Euclidean distance ('Energy', E) with the number (N) of clusters (1-24) for the mean percentage contribution of BB, SU, DU and SS to the total AOD (500nm) at the seasonal timescale as represented by the monthly triplets: DJF, MAM, JJA and SON.

Table S3. Descriptive statistics of the global mean mixtures extracted from the season (DJF) 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	2128	10842	12092	11315	2016	12251	2970	5495	4043	1648
AOD (500)	0.32	0.07	0.12	0.12	0.22	0.06	0.14	0.11	0.11	0.35
%BB (Mean)	53.9	10.9	7.5	6.5	15.0	12.6	14.4	9.0	31.2	7.8
%SU (Mean)	25.9	50.9	58.0	37.9	24.5	64.7	42.8	42.0	53.9	10.5
%DU (Mean)	16.5	5.9	18.2	6.4	56.1	5.7	35.8	18.1	6.8	80.7
%SS (Mean)	3.7	32.3	16.3	49.2	4.5	16.9	7.1	30.9	8.1	1.1
%BC (Mean)	7.5	2.6	3.0	1.6	3.7	2.9	4.5	3.0	6.4	2.0
%OC (Mean)	46.4	8.3	4.5	4.9	11.2	9.8	9.9	6.0	24.8	5.8
%BB (St.D.)	10.1	3.5	2.1	1.9	9.9	3.2	7.9	3.3	6.4	4.9
%SU (St.D.)	9.1	4.5	3.9	3.7	8.3	5.6	7.5	4.0	7.1	5.9
%DU (St.D.)	10.1	2.7	3.3	3.8	7.7	2.9	6.9	4.0	5.0	7.1
%SS (St.D.)	3.4	4.8	4.7	5.3	5.0	5.5	5.6	5.5	5.6	1.5
%BC (St.D.)	2.3	0.8	0.8	0.7	1.8	1.0	2.1	0.8	1.8	1.0
%OC (St.D.)	8.5	2.9	1.4	1.5	8.7	2.6	6.6	2.8	6.2	4.1
%BB (Median)	51.6	10.7	6.5	5.8	11.9	12.7	12.7	8.0	30.0	6.0
%SU (Median)	25.9	50.7	58.3	37.6	25.5	63.5	42.8	41.8	54.6	9.4
%DU (Median)	14.1	5.7	17.6	4.9	56.4	6.7	35.2	18.3	5.1	81.0
%SS (Median)	2.9	32.0	17.9	48.9	2.0	17.1	5.5	30.5	7.7	0.6
%BC (Median)	7.4	2.6	2.7	1.4	3.1	2.8	4.3	2.8	6.1	1.7
%OC (Median)	45.1	8.0	3.9	4.6	7.9	9.7	7.0	5.1	23.5	4.4
%BB (IQR)	12.9	4.2	1.3	2.4	12.9	3.7	13.7	4.1	9.0	4.6
%SU (IQR)	13.1	6.8	5.0	5.6	10.8	3.9	11.2	5.9	9.7	11.0
%DU (IQR)	14.6	3.3	4.2	4.0	11.6	4.8	9.9	5.0	6.4	12.1
%SS (IQR)	4.8	7.8	5.4	8.3	5.7	6.4	5.9	8.8	9.4	0.8
%BC (IQR)	3.0	0.9	0.7	0.9	2.1	0.8	3.1	1.1	2.3	1.0
%OC (IQR)	10.9	3.3	0.7	2.0	11.2	3.0	11.0	3.3	7.5	3.8

Table S4. Descriptive statistics of the global mean mixtures extracted from the season (MAM) 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	Cluster 11
Pixels	5251	8592	4296	2481	9099	7172	8513	3690	2488	9294	3924
AOD (500)	0.20	0.22	0.08	0.28	0.06	0.23	0.10	0.10	0.49	0.04	0.14
%BB (Mean)	12.3	9.5	16.4	9.6	6.5	10.1	3.5	40.1	4.2	9.2	12.4
%SU (Mean)	34.7	55.8	66.5	24.7	39.3	47.0	24.8	45.1	9.5	49.8	43.3
%DU (Mean)	41.5	28.7	4.3	62.5	4.1	38.0	2.8	9.1	85.6	4.1	26.4
%SS (Mean)	11.5	6.1	12.8	3.2	50.1	5.0	68.9	5.6	0.7	36.9	17.9
%BC (Mean)	3.2	2.7	3.5	2.6	1.7	2.9	0.9	6.9	1.3	2.3	3.5
%OC (Mean)	9.1	6.8	12.9	7.0	4.8	7.2	2.6	33.2	2.9	6.9	8.9
%BB (St.D.)	6.0	3.1	7.1	5.3	1.8	3.2	0.7	9.6	2.4	3.8	6.1
%SU (St.D.)	4.0	3.0	8.4	6.3	4.1	3.1	3.6	9.3	4.6	4.3	4.5
%DU (St.D.)	4.5	3.1	3.4	6.6	2.3	3.7	1.0	7.4	6.4	2.3	5.4
%SS (St.D.)	4.6	2.6	7.2	4.1	4.7	2.4	4.8	5.1	1.2	4.8	5.1
%BC (St.D.)	1.0	0.8	1.6	1.1	0.5	0.8	0.2	2.0	0.5	0.9	1.4
%OC (St.D.)	5.4	2.4	5.7	4.5	1.3	2.5	0.5	9.1	2.0	3.0	4.8
%BB (Median)	10.3	8.5	15.7	8.4	6.0	9.1	3.4	39.3	3.7	7.5	10.5
%SU (Median)	34.8	55.3	65.5	24.2	39.1	47.2	24.2	47.1	10.1	49.8	43.4
%DU (Median)	41.4	29.1	3.5	62.3	3.6	37.1	2.6	7.0	85.3	3.7	28.1
%SS (Median)	12.1	6.3	12.4	1.2	49.8	5.5	69.6	3.8	0.3	37.6	17.4
%BC (Median)	2.9	2.4	2.9	2.2	1.6	2.6	0.8	7.0	1.2	2.0	3.0
%OC (Median)	7.1	6.1	12.7	5.9	4.4	6.4	2.5	31.9	2.4	5.5	7.4
%BB (IQR)	4.5	1.2	11.2	6.2	1.5	2.1	1.0	13.4	2.7	3.4	9.5
%SU (IQR)	4.2	3.9	8.3	9.4	6.9	4.9	6.0	13.8	7.3	5.3	5.3
%DU (IQR)	6.0	3.5	4.2	11.2	1.5	4.6	1.0	10.2	10.8	1.4	5.7
%SS (IQR)	5.2	2.2	13.0	3.2	7.5	3.1	8.4	6.1	0.3	5.8	6.9
%BC (IQR)	1.2	0.3	2.1	1.6	0.5	0.5	0.3	2.9	0.8	0.8	1.9
%OC (IQR)	3.0	1.1	9.1	4.7	1.1	1.6	0.8	12.9	1.9	2.7	7.3

Table S5. Descriptive statistics of the global mean mixtures extracted from the season (JJA) 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	Cluster 11
Pixels	8887	12154	10096	1894	3194	3010	6280	2784	6014	3063	7424
AOD (500)	0.10	0.14	0.06	0.23	0.07	0.43	0.05	0.11	0.18	0.19	0.15
%BB (Mean)	7.6	21.0	11.8	72.0	24.4	5.0	8.9	41.6	14.7	14.7	12.7
%SU (Mean)	20.3	50.7	32.0	22.8	35.4	16.5	49.0	37.0	67.0	32.9	52.1
%DU (Mean)	2.2	26.5	2.8	2.7	5.8	76.7	2.8	11.2	11.1	46.6	25.1
%SS (Mean)	69.9	1.8	53.4	2.5	34.4	1.8	39.2	10.1	7.3	5.8	10.1
%BC (Mean)	1.4	4.3	2.7	11.4	4.9	1.7	2.1	7.3	3.6	3.6	3.8
%OC (Mean)	6.2	16.7	9.1	60.6	19.4	3.3	6.8	34.3	11.0	11.1	9.0
%BB (St.D.)	1.8	4.0	3.8	9.5	5.3	2.6	2.0	8.0	4.9	7.0	2.2
%SU (St.D.)	3.1	4.5	4.8	6.8	4.6	5.7	4.7	7.5	6.0	6.0	4.7
%DU (St.D.)	0.5	3.4	1.2	2.4	4.3	8.7	2.3	8.8	6.4	8.5	4.8
%SS (St.D.)	4.4	1.5	4.5	3.1	6.3	2.6	6.0	6.5	6.9	5.8	4.4
%BC (St.D.)	0.3	0.7	1.1	2.6	1.0	0.9	0.7	1.6	1.2	1.2	0.6
%OC (St.D.)	1.6	3.6	2.9	8.0	4.6	2.0	1.5	7.3	4.3	6.3	2.2
%BB (Median)	7.2	19.8	10.9	71.3	24.2	4.5	8.2	41.0	14.0	12.6	12.2
%SU (Median)	19.8	51.4	31.3	23.5	34.6	16.7	48.7	36.2	66.0	32.3	52.0
%DU (Median)	2.0	26.6	2.5	1.9	3.7	76.2	2.2	8.3	12.5	45.6	24.5
%SS (Median)	70.3	1.5	53.1	1.4	35.5	0.3	40.7	9.5	5.0	5.3	9.4
%BC (Median)	1.4	4.0	2.6	11.3	5.0	1.5	1.8	7.1	3.5	3.4	3.6
%OC (Median)	5.9	15.8	8.2	60.6	19.2	2.9	6.4	33.9	10.6	9.0	8.5
%BB (IQR)	1.9	4.2	6.0	16.3	6.5	3.1	1.4	12.9	5.8	9.6	2.3
%SU (IQR)	4.7	5.3	7.4	10.1	6.0	8.2	4.4	10.4	7.9	9.0	6.0
%DU (IQR)	0.7	3.5	1.0	2.1	4.4	13.8	0.6	14.5	12.3	14.2	7.2
%SS (IQR)	7.3	1.0	7.4	2.3	9.7	3.0	5.7	10.7	9.8	7.2	5.5
%BC (IQR)	0.4	0.7	1.9	4.0	1.3	1.2	0.7	2.3	1.2	1.5	0.8
%OC (IQR)	1.6	3.6	4.2	13.7	5.7	1.9	0.9	11.6	5.3	8.4	2.7

Table S6. Descriptive statistics of the global mean mixtures extracted from the season (SON) 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	Cluster 11
Pixels	8532	6154	8472	15186	6949	2947	4785	3349	2384	3665	2377
AOD (500)	0.12	0.11	0.14	0.07	0.12	0.16	0.10	0.06	0.17	0.14	0.36
%BB (Mean)	15.2	9.1	11.4	21.9	13.4	61.4	37.8	13.2	14.2	17.6	5.2
%SU (Mean)	24.2	43.9	62.1	35.6	52.2	28.7	37.0	57.2	30.5	71.0	13.3
%DU (Mean)	4.4	21.4	19.2	5.4	26.9	6.1	7.8	2.8	51.2	4.6	80.3
%SS (Mean)	56.2	25.6	7.3	37.1	7.5	3.8	17.3	26.7	4.1	6.8	1.1
%BC (Mean)	2.4	3.4	3.5	4.0	4.1	10.0	7.2	3.3	3.8	4.4	1.7
%OC (Mean)	12.7	5.8	7.9	17.9	9.2	51.4	30.6	9.9	10.4	13.2	3.6
%BB (St.D.)	2.3	1.8	2.3	3.9	3.7	9.2	6.0	4.2	9.7	6.8	3.7
%SU (St.D.)	3.0	4.5	3.8	4.3	4.2	6.3	6.7	5.5	7.9	7.1	6.3
%DU (St.D.)	1.3	4.9	3.1	2.0	4.5	7.1	7.0	3.6	8.9	3.6	8.9
%SS (St.D.)	5.6	5.1	3.8	4.5	4.0	3.6	7.8	5.6	4.4	4.6	1.9
%BC (St.D.)	0.4	0.7	0.8	1.1	1.2	2.1	1.3	0.9	1.6	1.8	0.9
%OC (St.D.)	1.9	1.5	1.9	3.0	2.8	8.0	5.2	3.7	8.6	5.7	3.1
%BB (Median)	15.0	8.8	11.4	20.7	12.5	60.0	37.2	12.8	9.6	15.7	4.6
%SU (Median)	24.0	43.4	61.1	35.3	52.8	28.0	36.5	57.7	30.5	70.9	13.3
%DU (Median)	4.1	22.3	19.6	5.3	25.9	2.8	4.9	0.9	51.4	4.5	79.2
%SS (Median)	56.3	24.6	7.1	37.1	7.0	2.6	19.1	26.4	1.8	6.1	0.3
%BC (Median)	2.3	3.2	3.3	3.6	3.8	9.9	7.1	3.2	3.3	4.2	1.6
%OC (Median)	12.6	5.4	8.2	17.0	8.7	50.7	30.0	9.5	6.2	11.5	2.7
%BB (IQR)	3.2	1.9	2.1	5.2	2.4	13.5	9.4	6.0	12.7	9.7	3.8
%SU (IQR)	4.8	6.0	5.5	6.8	5.9	8.0	7.0	8.6	12.3	9.6	10.9
%DU (IQR)	1.4	9.1	4.4	1.4	3.5	7.2	7.5	2.9	14.0	6.4	15.9
%SS (IQR)	9.8	7.2	3.6	6.7	5.1	4.0	13.4	7.4	7.0	7.4	0.6
%BC (IQR)	0.6	0.9	0.5	1.6	0.6	2.9	1.6	1.6	2.1	2.4	1.4
%OC (IQR)	2.6	1.2	1.9	3.5	1.8	12.5	8.2	5.6	10.1	7.6	2.4

S4. Cluster characteristics of the seasonal partitions

In this section we present the summary tables and graphs of mean values of optical and microphysical aerosol parameters extracted from the AERONET inversion data record for each cluster of the mean seasonal (DJF, MAM, JJA and SON) global partitions including: spectral parameters (AOD, ASYM, AAOD, SSA, CRI-R, CRI-I, LR) at central wavelength of 440, 675, 870 and 1020nm, the mean percentage of spherical particles ('%Sphericity'), the mean AVSD (in 22 logarithmic radial bins from 0.05µm to 15 µm), AVSD-derived microphysical parameters (r_f, σ_f, V_f, r_c, σ_c, V_c associated with the fine and coarse modes 'f''and 'c' respectively), and the fine mode fraction η. Values are calculated from the L2.0 V2 inversion product unless accompanied by a '*' whereby the value is calculated from the L1.5 V2 inversion product.

Table S7. Descriptive statistics of key optical and microphysical parameters extracted from the global AERONET inversion record in each cluster for the season (DJF).

AERONET L2.0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10
Pixels	8	8	33	2	18	33	57	13	23	18
Records	358	1372	4080	744*	4834	7739	14149	929	3056	9800
AOD (440)	0.19	0.08	0.06	0.14*	0.42	0.11	0.12	0.12	0.20	0.27
AE (440/675)	0.92	0.63	1.73	1.14*	1.07	1.76	1.55	1.09	1.41	0.72
H2O	3.62	2.82	0.49	1.56*	1.17	0.95	0.80	1.15	2.34	1.27
ASYM (440)	0.72	0.75	0.72	0.69*	0.72	0.70	0.71	0.73	0.71	0.75
P180 (440)	0.20	0.41	0.15	0.26*	0.19	0.17	0.17	0.23	0.19	0.18
AAOD (440)	0.057	0.005*	0.062	0.017*	0.138	0.056	0.073	0.052	0.070	0.083
SSA (440)	0.90	0.95*	0.90	0.90*	0.87	0.93	0.90	0.93	0.90	0.89
LR (440)	90.1	39.4*	108.4	65.5*	94.5	92.2	96.4	99.8	98.4	96.4
CRI-R (440)	1.424	1.449*	1.418	1.542*	1.432	1.432	1.466	1.412	1.441	1.442
CRI-I (440)	0.011	0.003*	0.018	0.012*	0.014	0.010	0.016	0.015	0.016	0.008
r(f)	0.15	0.18	0.18	0.15*	0.16	0.17	0.17	0.19	0.17	0.16
σ(f)	0.47	0.49	0.46	0.46*	0.48	0.45	0.47	0.50	0.46	0.50
V(f)	0.02	0.01	0.01	0.01*	0.05	0.01	0.02	0.01	0.03	0.02
r(c)	2.35	2.70	2.92	2.69*	2.54	2.96	2.89	2.68	3.02	2.26
σ(c)	0.71	0.69	0.68	0.66*	0.65	0.67	0.68	0.68	0.69	0.63
V(c)	0.06	0.04	0.01	0.058*	0.08	0.02	0.02	0.02	0.03	0.10
η	0.28	0.15	0.50	0.26*	0.31	0.58	0.51	0.44	0.41	0.20
% Sphericity	63.4	95.6	95.5	51.3*	33.8	86.3	72.3	97.9	96.4	2.6

While 1372 Level 2.0 records were extracted for cluster 2, there were no records containing data for the spectral parameters AAOD, SSA, LR and the CRI. In order to estimate the values of these parameters, we extracted a total of 1074 analogous records in this cluster from the Level 1.5 dataset at the sites: Amsterdam Island, Cape_San_Juan, Coconut_Island, Dunedin, Honolulu, Manila_Observatory, Prospect Hill, Rio Pedras, Roosevelt Roads, Sable_Island, Tahiti, and Tudor Hill, and used them to calculate an estimate of the mean values. For cluster 4, only 8 Level 2.0 records were available. As such we extracted all 744 available records in this cluster from the Level 1.5 dataset at the sites: Azores, Bonanza_Creek, CEILAP-RG, Crozet Island, Guam, Kapoho, Mace Head, Marambio, and Santiago, and then proceeded to calculate estimated mean values of all parameters.

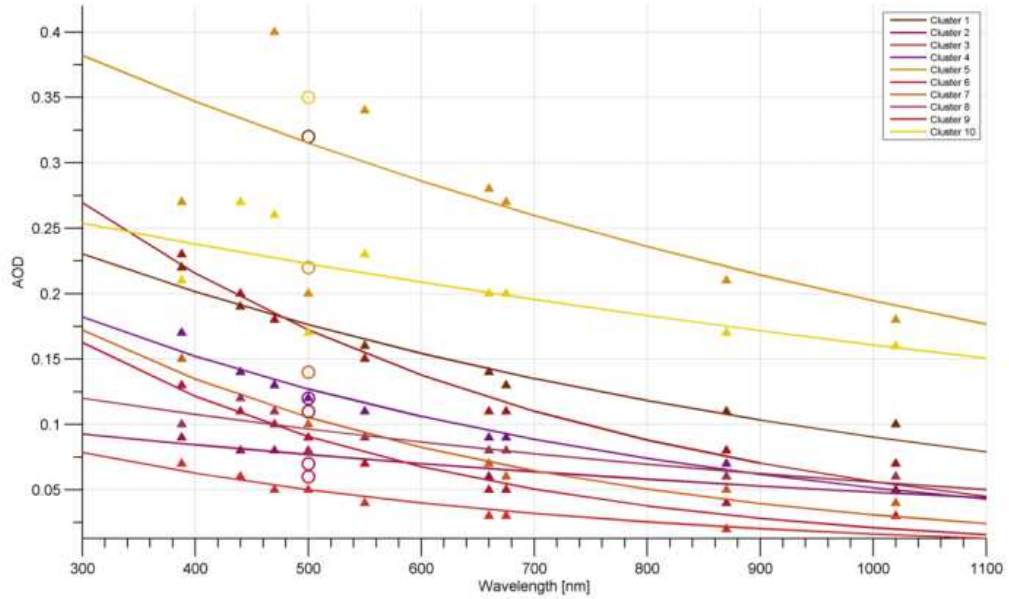


Fig. S10. Optimized least square regression fits to the spectral behaviour of mean values of the AERONET-derived AOD for each cluster during the season (DJF). The circles at 500nm are the mean cluster values of the total AOD obtained by GOCART.

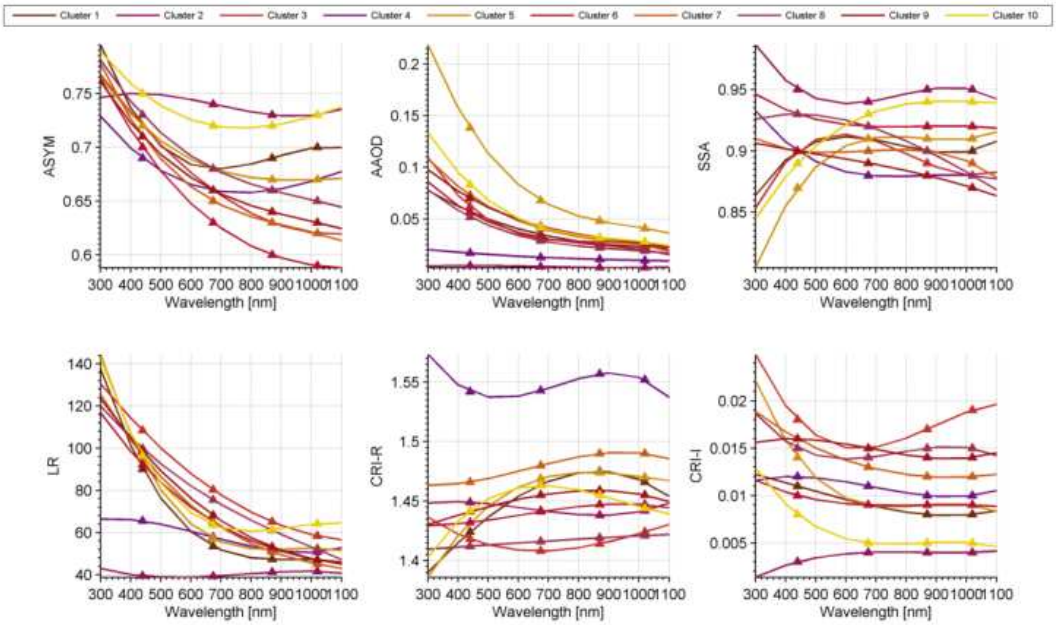


Fig. S11. Spectral behaviour of the global mean values of key optical parameters (ASYM, AAOD, SSA, LR) and microphysical parameters (CRI-R and CRI-I) for each cluster from extracted AERONET data at 440, 675, 870 and 1020nm during the season (DJF).

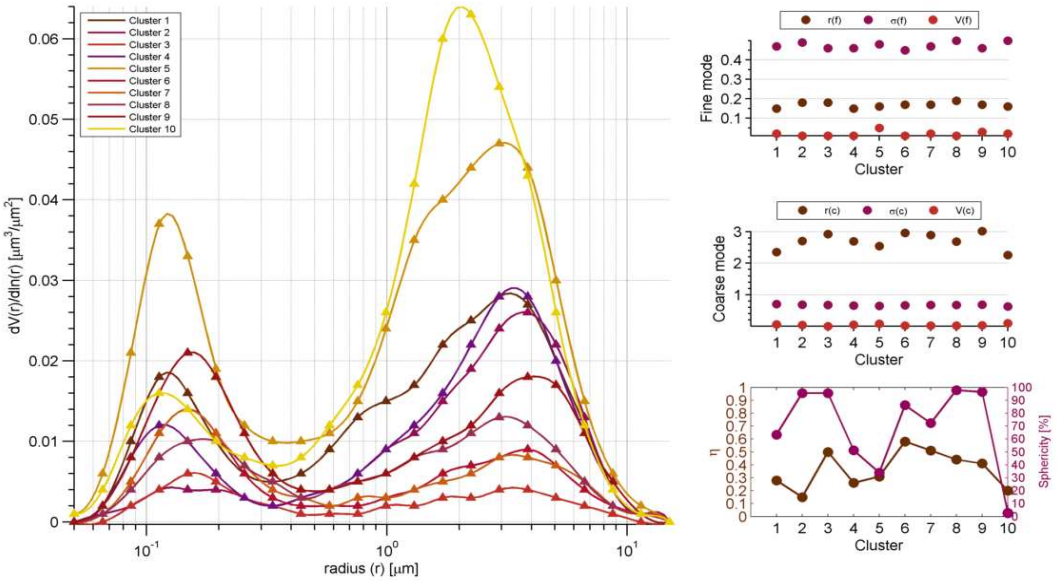


Fig. S12. Global mean size distributions for each cluster together with the derived fine and coarse mode parameters (r_f, σ_f, V_f and r_c, σ_c, V_c respectively) during the season (DJF).

Table S8. Descriptive statistics of key optical and microphysical parameters extracted from the global AERONET inversion record in each cluster for the season (MAM).

AERONET L2.0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11
Pixels	18	58	18	28	2	45	0	25	26	1	15
Records	3380	15023	2645	10340	451*	9083	157*	4170	11823	871*	1909
AOD (440)	0.07	0.19	0.20	0.19	0.05*	0.19	0.03*	0.25	0.39	0.07*	0.07
AE (440/675)	1.19	1.62	1.47	1.09	0.78*	1.52	0.62*	1.62	0.38	0.71*	1.21
H2O	1.13	1.08	2.19	1.25	1.60*	1.05	0.89*	2.36	1.44	3.98*	1.83
ASYM (440)	0.71	0.70	0.71	0.72	0.72*	0.71	0.75*	0.69	0.75	0.73*	0.71
P180 (440)	0.23	0.19	0.18	0.20	0.34*	0.18	0.25*	0.19	0.21	0.31*	0.26
AAOD (440)	0.105	0.033	0.040	0.088	0.004*	0.050	0.010*	0.068	0.064	0.006*	0.030
SSA (440)	0.88	0.95	0.95	0.89	0.92*	0.93	0.64*	0.92	0.91	0.93*	0.94
LR (440)	92.7	81.9	91.3	77.9	45.2*	85.6	78.4*	90.2	71.7	49.6*	69.6
CRI-R (440)	1.458	1.437	1.425	1.507	1.511*	1.458	1.559*	1.445	1.483	1.503*	1.446
CRI-I (440)	0.010	0.007	0.007	0.009	0.008*	0.009	0.194*	0.013	0.004	0.005*	0.003
r(f)	0.16	0.16	0.17	0.16	0.17*	0.16	0.26*	0.16	0.16	0.17*	0.16
σ(f)	0.48	0.44	0.45	0.47	0.49*	0.45	0.56*	0.43	0.58	0.51*	0.48
V(f)	0.01	0.02	0.03	0.02	0.00*	0.02	0.00*	0.03	0.02	0.00*	0.01
r(c)	2.46	2.49	2.84	2.43	2.55*	2.38	2.76*	2.97	2.14	2.78*	2.39
σ(c)	0.73	0.72	0.67	0.70	0.73*	0.71	0.65*	0.69	0.62	0.70*	0.70
V(c)	0.02	0.02	0.03	0.05	0.02*	0.02	0.01*	0.03	0.21	0.04*	0.03
η	0.26	0.52	0.46	0.28	0.18*	0.47	0.36*	0.47	0.11	0.14*	0.35
% Sphericity	9.3	91.0	77.2	27.6	87.0*	56.0	87.1*	98.6	1.0	66.2*	63.1

While 36 and 114 Level 2.0 records were extracted for clusters 5 and 10 (cluster 7 did not encompass any sites), there were no records containing data for AAOD, SSA, LR and the CRI at this level of quality assurance. Furthermore, since the number of Level 2.0 records is low for both clusters, we extracted a total of 451 full records for cluster 5 from the Level 1.5 dataset at the sites: Amsterdam Island, Rottnest Island and Dunedin, and a total of 871 full records for cluster 10 from the Level 1.5 dataset at the sites: Kapoho, McMurdo, Nauru, REUNION_ST_DENIS, South_Pole, and Tahiti. For cluster 7, a total of 157 full records from the Level 1.5 dataset at the sites: CEILAP-RG, Crozet Island, Marambio, Utsteinen, and Vechernaya_Hill was used. We then proceeded to use this data to calculate mean values of all parameters.

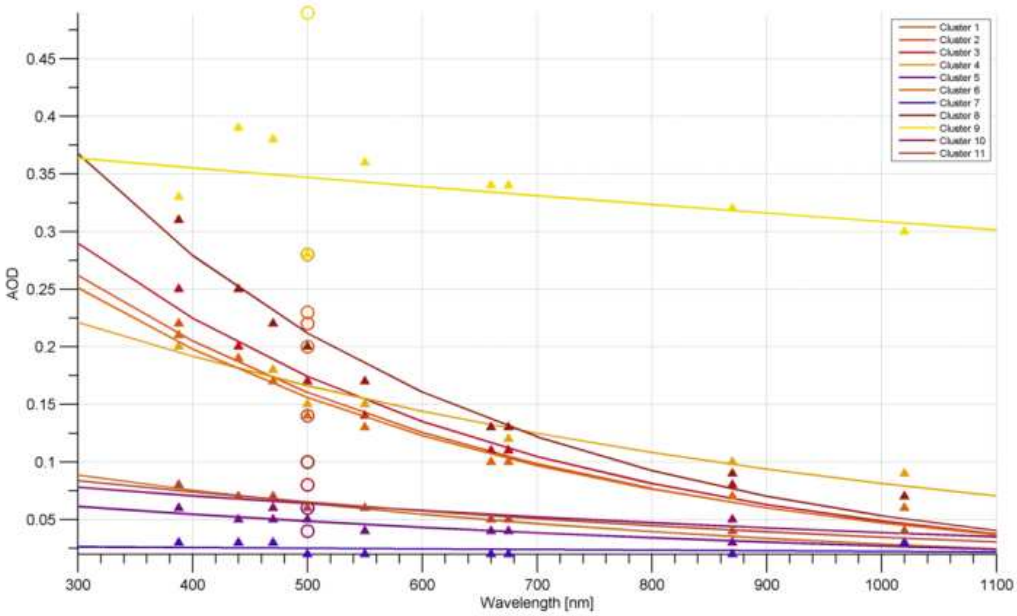


Fig. S13. Optimized least square regression fits to the spectral behaviour of mean values of the AERONET-derived AOD for each cluster during the season (MAM). The circles at 500nm are the mean cluster values of the total AOD obtained by GOCART.

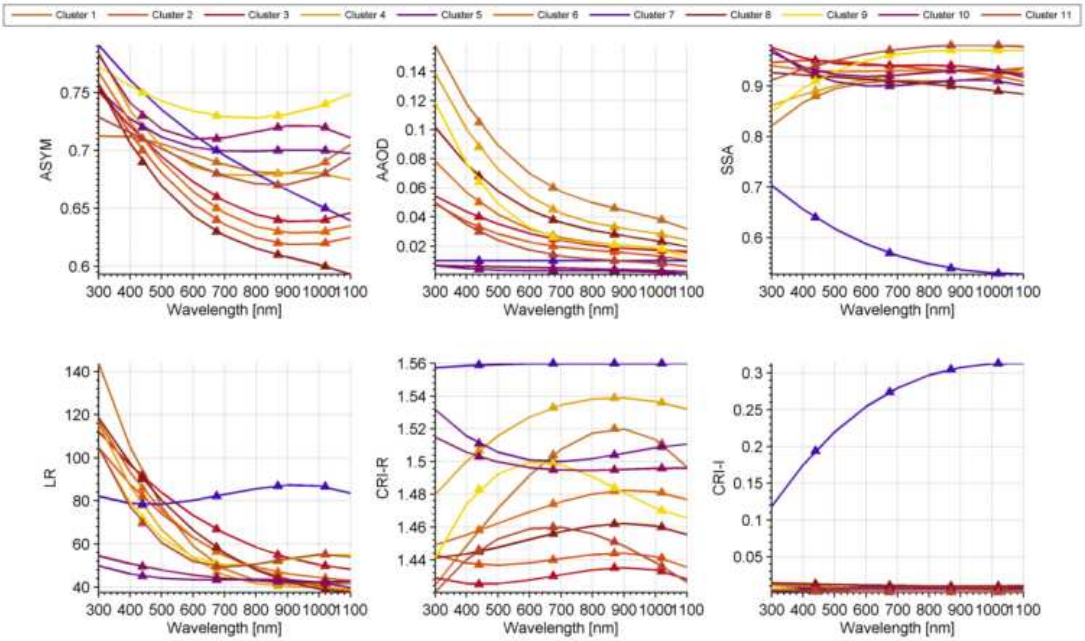


Fig. S14. Spectral behaviour of the global mean values of key optical parameters (ASYM, AAOD, SSA, LR) and microphysical parameters (CRI-R and CRI-I) for each cluster from extracted AERONET data at 440, 675, 870 and 1020nm during the season (MAM).

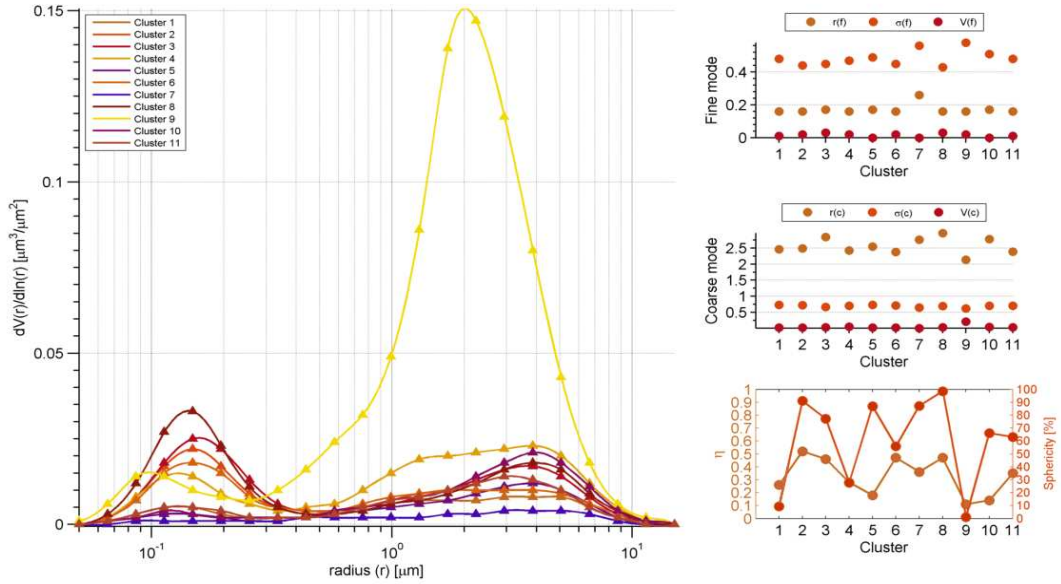


Fig. S15. Global mean size distributions for each cluster together with the derived fine and coarse mode parameters (r_f, σ_f, V_f and r_c, σ_c, V_c respectively) during the season (MAM).

Table S9. Descriptive statistics of key optical and microphysical parameters extracted from the global AERONET inversion record in each cluster for the season (JJA).

AERONET L2.0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11
Pixels	0	51	2	21	5	31	0	13	82	32	16
Records	102*	6466	771*	1792	2419*	3478	333*	1043	6195	2921	1525
AOD (440)	0.03*	0.18	0.06*	0.26	0.07*	0.45	0.07*	0.14	0.28	0.19	0.27
AE (440/675)	0.78*	1.60	0.89*	1.74	1.10*	0.50	0.58*	1.41	1.72	1.34	1.64
H2O	0.66*	1.99	2.91*	2.05	1.38*	2.40	4.37*	1.95	2.53	1.94	2.65
ASYM (440)	0.74*	0.70	0.72*	0.67	0.70*	0.74	0.73*	0.69	0.71	0.72	0.71
P180 (440)	0.21*	0.19	0.31*	0.18	0.22*	0.21	0.33*	0.20	0.17	0.19	0.17
AAOD (440)	0.011*	0.043	0.005*	0.080	0.012*	0.055	0.006*	0.077	0.026	0.051	0.030
SSA (440)	0.63*	0.94	0.92*	0.89	0.90*	0.92	0.92*	0.87	0.97	0.92	0.95
LR (440)	95.9*	80.4	50.4*	89.2	79.1*	67.3	46.8*	100.7	84.8	78.6	87.4
CRI-R (440)	1.558*	1.453	1.498*	1.488	1.519*	1.501	1.505*	1.422	1.415	1.464	1.389
CRI-I (440)	0.137*	0.008	0.006*	0.020	0.020*	0.004	0.006*	0.020	0.005	0.007	0.006
r(f)	0.25*	0.16	0.17*	0.15	0.16*	0.16	0.18*	0.15	0.17	0.16	0.16
σ(f)	0.55*	0.42	0.49*	0.40	0.47*	0.53	0.50*	0.45	0.43	0.45	0.42
V(f)	0.00*	0.02	0.00*	0.03	0.01*	0.03	0.01*	0.02	0.04	0.02	0.04
r(c)	2.40*	2.92	2.60*	3.18	2.65*	2.31	2.82*	2.80	3.02	2.63	2.76
σ(c)	0.67*	0.69	0.69*	0.69	0.73*	0.62	0.69*	0.71	0.67	0.67	0.67
V(c)	0.01*	0.03	0.03*	0.03	0.02*	0.23	0.03*	0.03	0.02	0.04	0.03
η	0.47*	0.47	0.17*	0.53	0.35*	0.12	0.15*	0.39	0.65	0.34	0.58
% Sphericity	92.8*	84.7	84.8*	98.9	92.6*	2.4	86.9*	94.5	98.6	43.1	94.7

While 224 and 294 Level 2.0 records were extracted for clusters 3 and 5 (clusters 1 and 7 did not encompass any sites), there were no records containing data for AAOD, SSA, LR and the CRI at this level of quality assurance. For these four clusters (1, 3, 5 and 7) we therefore extracted a total of 102 full records for cluster 1 from the Level 1.5 dataset at the sites: CEILAP-RG and Marambio, a total of 771 full records for cluster 3 from the Level 1.5 dataset at the sites: Amsterdam_Island, Crozet Island, Dunedin, McMurdo, and Tahiti, a total of 2419 full records for cluster 5 from the Level 1.5 dataset at the sites: Adelaide_7, ARM_Gan_Island, Ascension_Island, Canberra, CEILAP-Bariloche, Coleambally, Lucinda, Milyering, Perth, Petrolina_SONDA, Puerto_Madryn, REUNION_ST_DENIS, Rottnest_Island, and Trelew, and a total of 333 full records for cluster 5 from the Level 1.5 dataset at the sites: Kapoho, Manus, Nauru, South_Pole_Obs, Utsteinen, and Vechernaya_Hill. We then used this data to calculate mean values of all parameters.

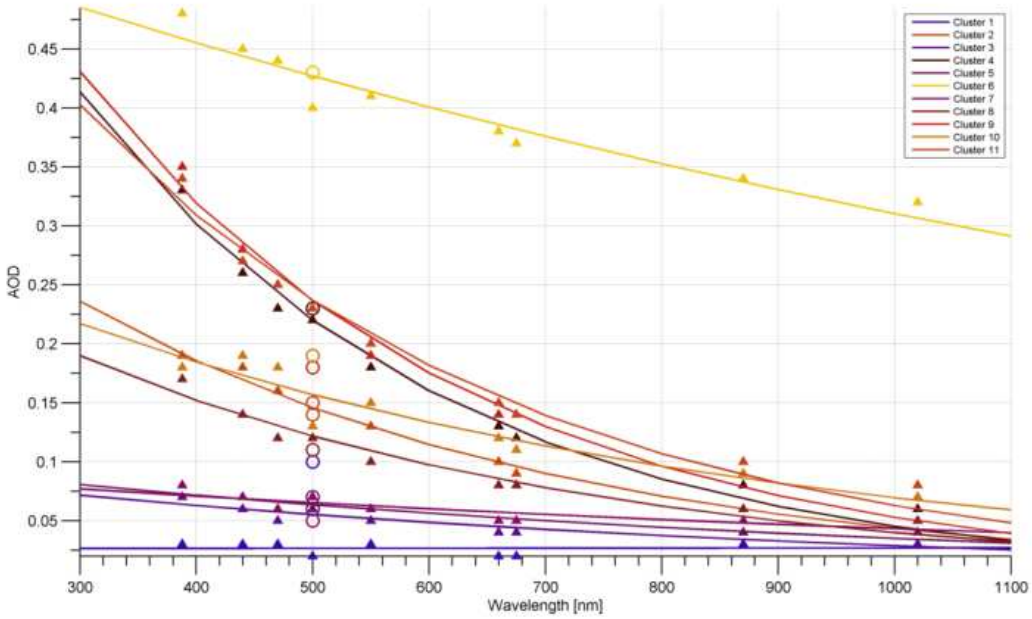


Fig. S16. Optimized least square regression fits to the spectral behaviour of mean values of the AERONET-derived AOD for each cluster during the season (JJA). The circles at 500nm are the mean cluster values of the total AOD obtained by GOCART.

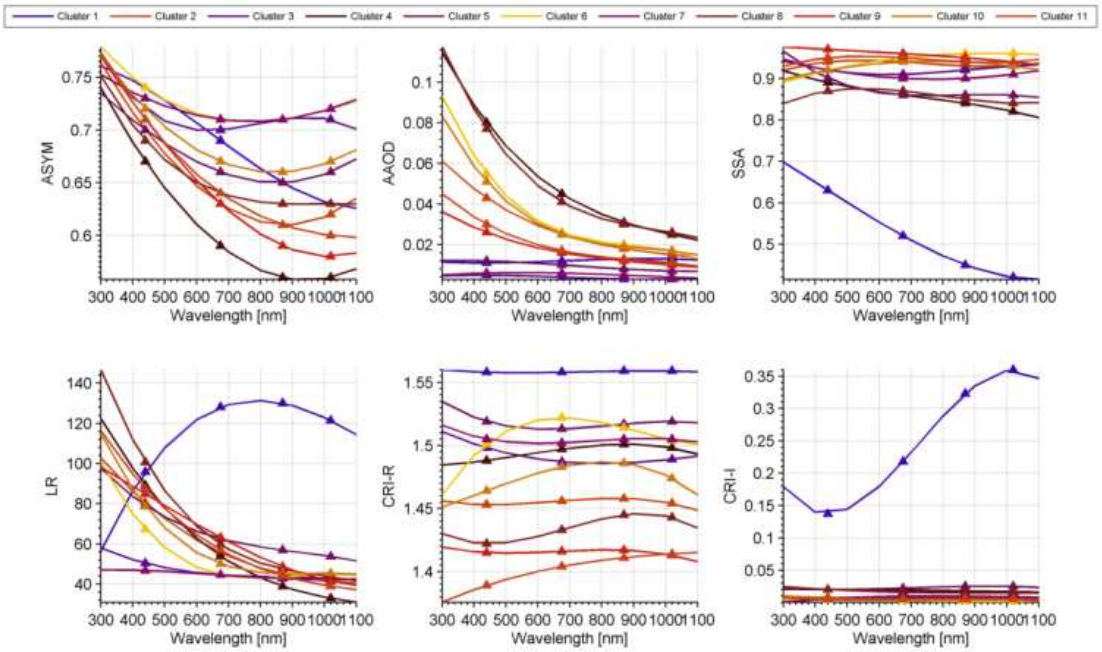


Fig. S17. Spectral behaviour of the global mean values of key optical parameters (ASYM, AAOD, SSA, LR) and microphysical parameters (CRI-R and CRI-I) for each cluster from extracted AERONET data at 440, 675, 870 and 1020nm during the season (JJA).

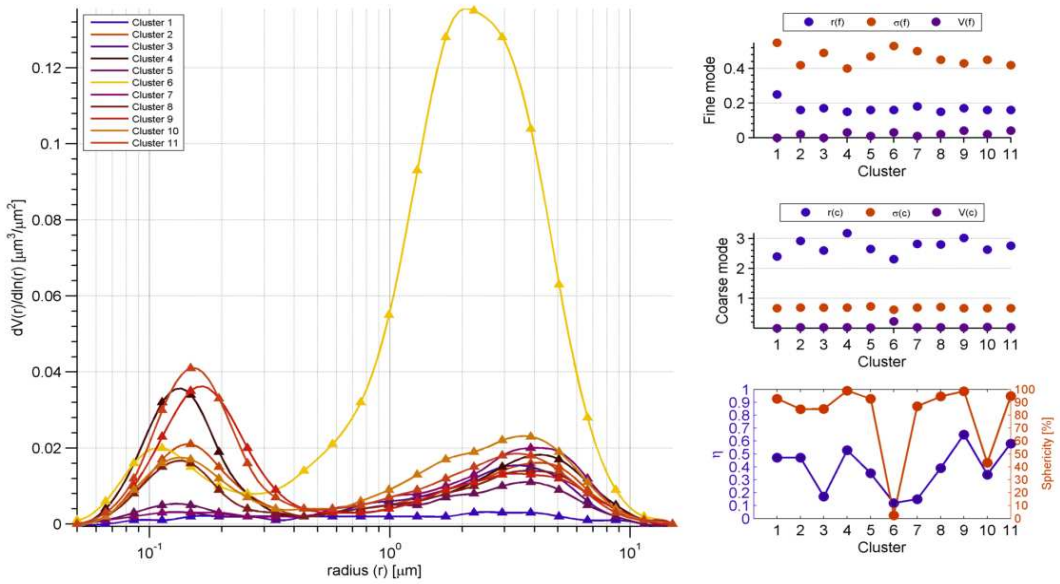


Fig. S18. Global mean size distributions for each cluster together with the derived fine and coarse mode parameters (r_f, σ_f, V_f and r_c, σ_c, V_c respectively) during the season (JJA).

Table S10. Descriptive statistics of key optical and microphysical parameters extracted from the global AERONET inversion record in each cluster for the season (SON).

AERONET L2.0	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6	Cluster 7	Cluster 8	Cluster 9	Cluster 10	Cluster 11
Pixels	1	4	51	3	52	28	15	5	23	43	28
Records	197*	268	9554	626*	15055	5671	2044	930	9738	10516	14096
AOD (440)	0.04*	0.06	0.16	0.07*	0.14	0.44	0.16	0.03	0.20	0.18	0.34
AE (440/675)	0.78*	1.56	1.61	0.97*	1.54	1.62	1.49	2.46	1.26	1.72	0.71
H2O	0.90*	2.09	1.40	2.62*	1.48	2.69	1.50	1.39	1.77	1.85	2.14
ASYM (440)	0.73*	0.70	0.71	0.71*	0.71	0.68	0.69	0.68	0.71	0.71	0.74
P180 (440)	0.15*	0.24	0.18	0.34*	0.18	0.18	0.20	0.26	0.19	0.17	0.19
AAOD (440)	0.016*	0.025	0.040	0.008*	0.055	0.080	0.063	0.001*	0.071	0.039	0.047
SSA (440)	0.69*	0.95	0.94	0.91*	0.92	0.90	0.90	0.98*	0.91	0.95	0.92
LR (440)	121.2*	72.6	90.1	48.9*	94.5	86.4	93.7	55.7*	85.0	91.9	80.2
CRI-R (440)	1.587*	1.450	1.421	1.495*	1.438	1.473	1.427	1.537*	1.470	1.409	1.461
CRI-I (440)	0.103*	0.007	0.008	0.008*	0.011	0.017	0.015	0.002*	0.011	0.007	0.005
r(f)	0.22*	0.16	0.17	0.16*	0.16	0.15	0.15	0.17	0.16	0.17	0.16
σ(f)	0.56*	0.44	0.44	0.48*	0.44	0.40	0.44	0.49	0.45	0.43	0.49
V(f)	0.00*	0.01	0.02	0.01*	0.02	0.06	0.02	0.00	0.02	0.03	0.03
r(c)	2.45*	2.35	2.88	2.70*	2.88	2.92	3.02	3.51	2.66	3.07	2.31
σ(c)	0.70*	0.70	0.68	0.72*	0.68	0.71	0.69	0.67	0.66	0.65	0.63
V(c)	0.01*	0.01	0.02	0.03*	0.02	0.04	0.03	0.01	0.06	0.02	0.14
η	0.46*	0.40	0.52	0.18*	0.47	0.56	0.38	0.65	0.34	0.60	0.17
% Sphericity	99.0*	98.5	93.4	94.8*	66.4	98.6	93.2	61.8	41.6	96.5	7.8

While 930 Level 2.0 records were extracted for cluster 8, there were no records containing data for AAOD, SSA, LR and the CRI at this level of quality assurance. For this cluster, we extracted a total of 527 analogous records from the Level 1.5 dataset at the sites: Andros_Island, Bermuda, Coconut_Island, Guam, Honolulu, Kapocho, Manus, Mauna_Loa, Nauru, Prospect_Hill, and Tudor_Hill, to estimate these parameters. For cluster 1 only 2 Level 2.0 records were available. As such we extracted 197 complete records from Level 1.5 data at the sites: CEILAP_RG, Crozet_Island and Marambio, and then proceeded to calculate mean values of all parameters. In the case of cluster 4, while 142 Level 2.0 records were extracted, there were no records containing data for AAOD, SSA, LR, CRI and the % sphericity at this level of quality assurance. For this cluster, we extracted a total of 626 analogous records from the Level 1.5 dataset at the sites: Amsterdam_Island, Dunedin, McMurdo, REUNION_ST_DENIS, Tahiti, Utsteinen, and Vechernaya_Hill, and used this data to estimate the values of all parameters.

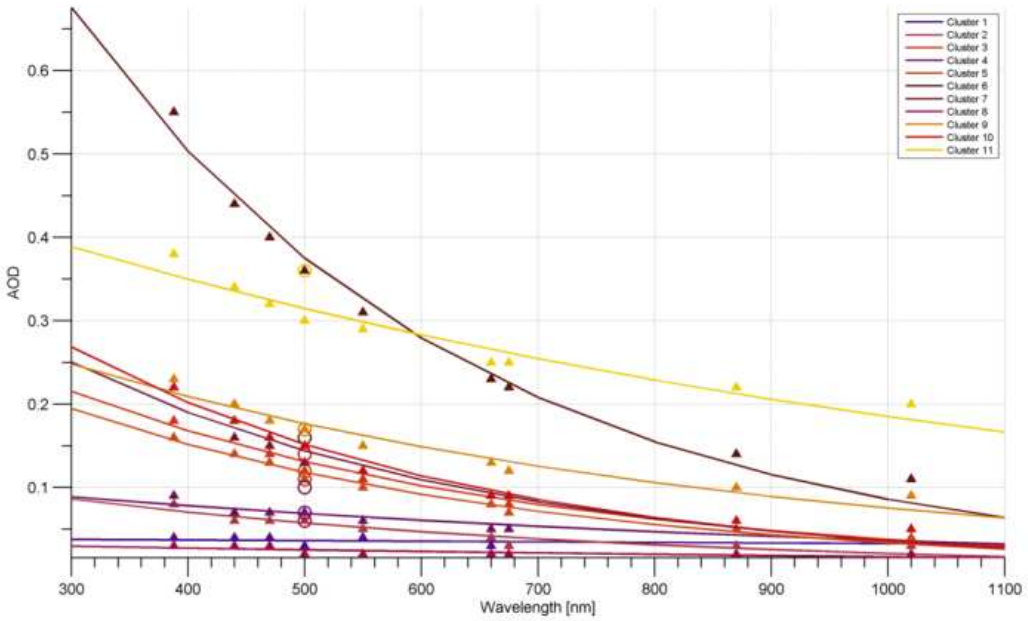


Fig. S19. Optimized least square regression fits to the spectral behaviour of mean values of the AERONET-derived AOD for each cluster during the season (SON). The circles at 500nm are the mean cluster values of the total AOD obtained by GOCART.

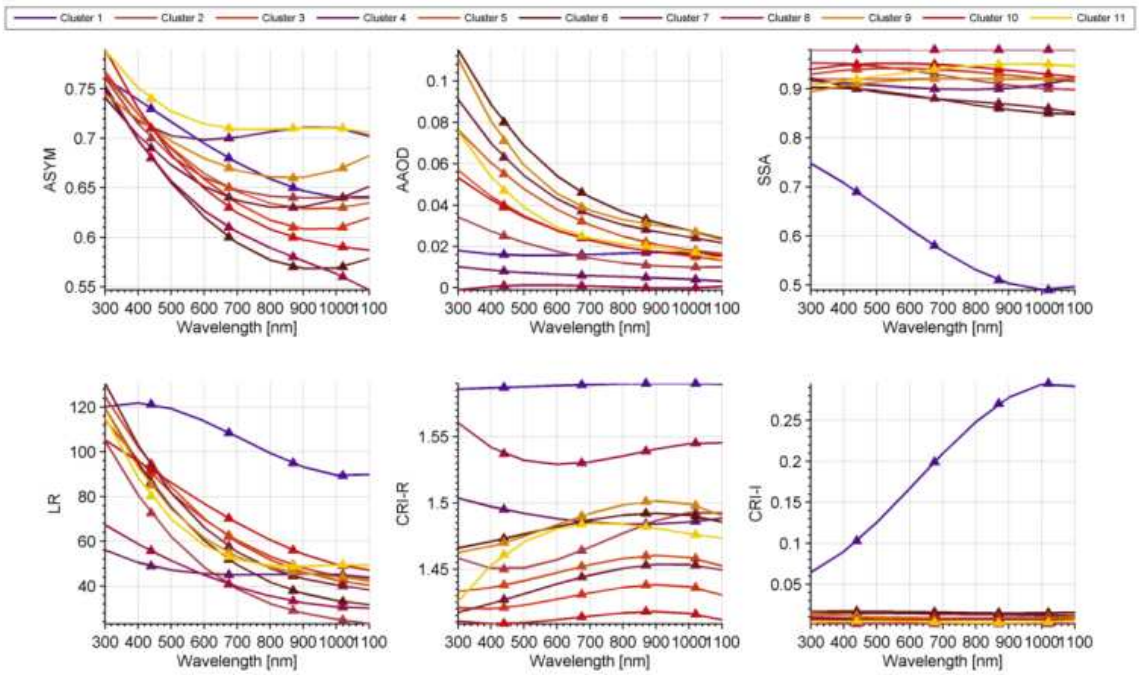


Fig. S20. Spectral behaviour of the global mean values of key optical parameters (ASYM, AAOD, SSA, LR) and microphysical parameters (CRI-R and CRI-I) for each cluster from extracted AERONET data at 440, 675, 870 and 1020nm during the season (SON).

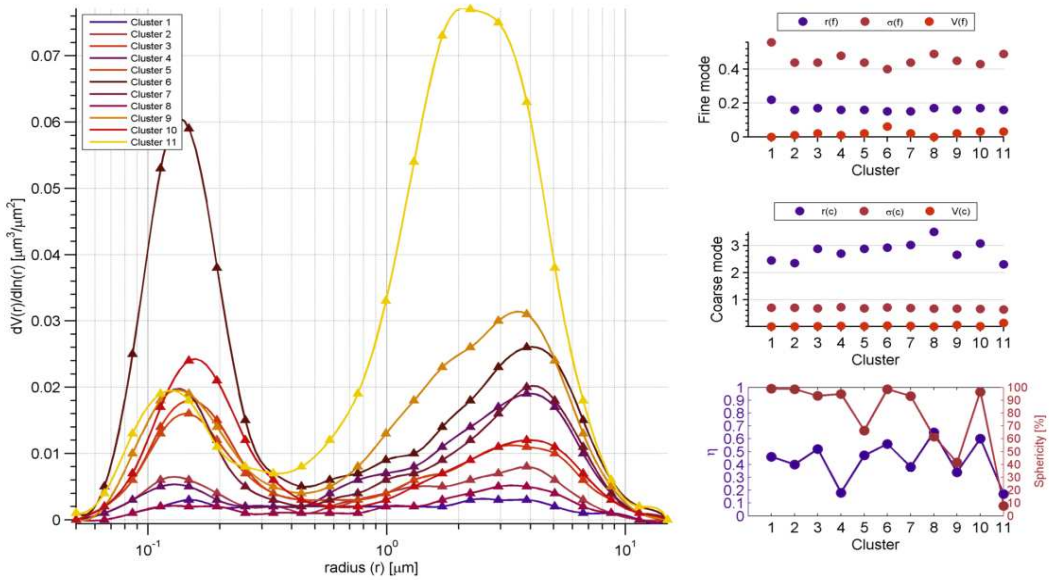


Fig. S21. Global mean size distributions for each cluster together with the derived fine and coarse mode parameters (r_f, σ_f, V_f and r_c, σ_c, V_c respectively) during the season (SON).

S4. Bivariate parameterizations of seasonal global aerosol mixtures

In this section we present the results of parameterizing the AERONET data record with pairs of optical and/or microphysical parameters commonly used in aerosol classification studies described in detail in the main text of the manuscript.

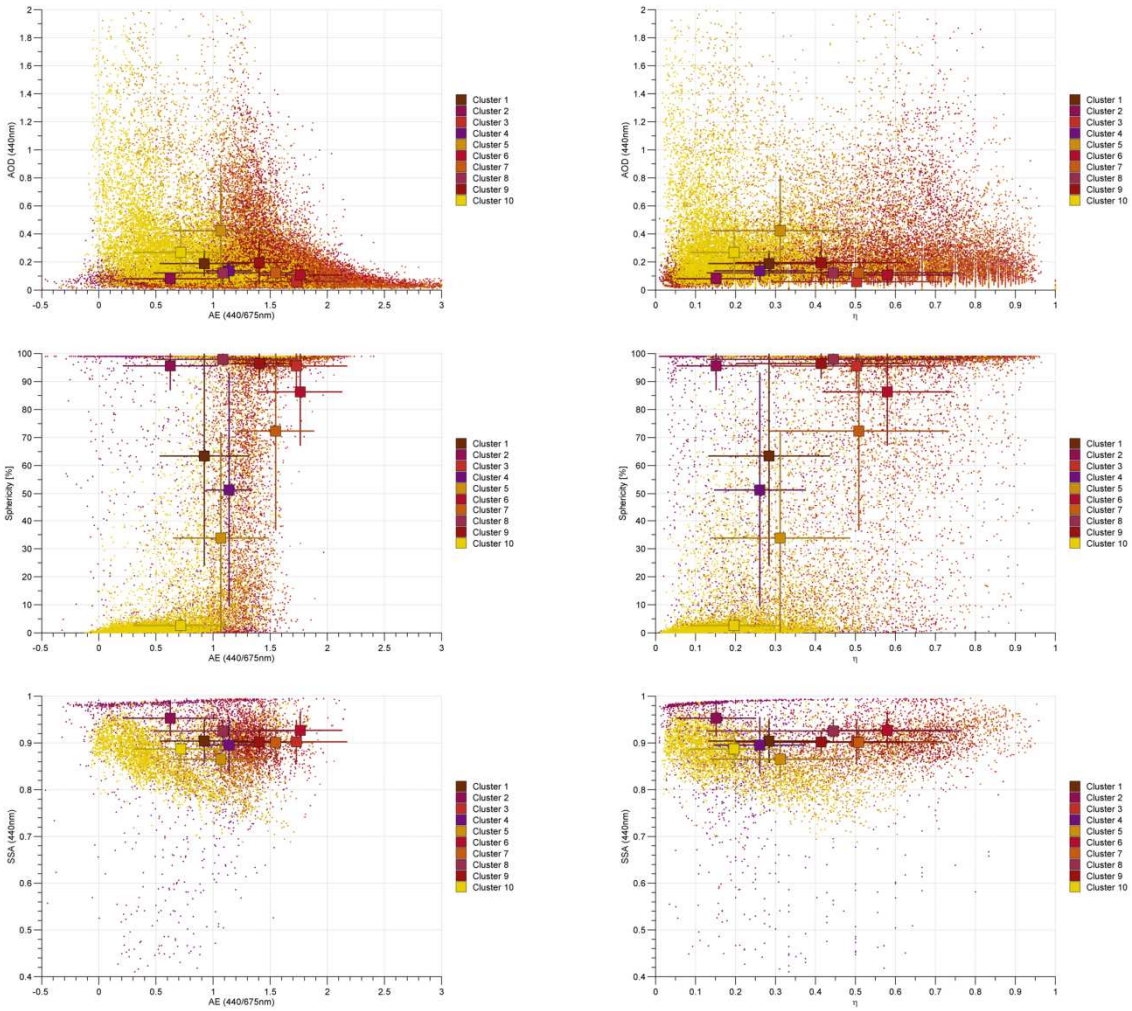


Fig. S22. Cluster means (squares) and standard deviations (lines) overlaid on the AERONET data record (points) for the season (DJF) parameterized by pairs of optical and/or microphysical parameters commonly used in aerosol classification studies.

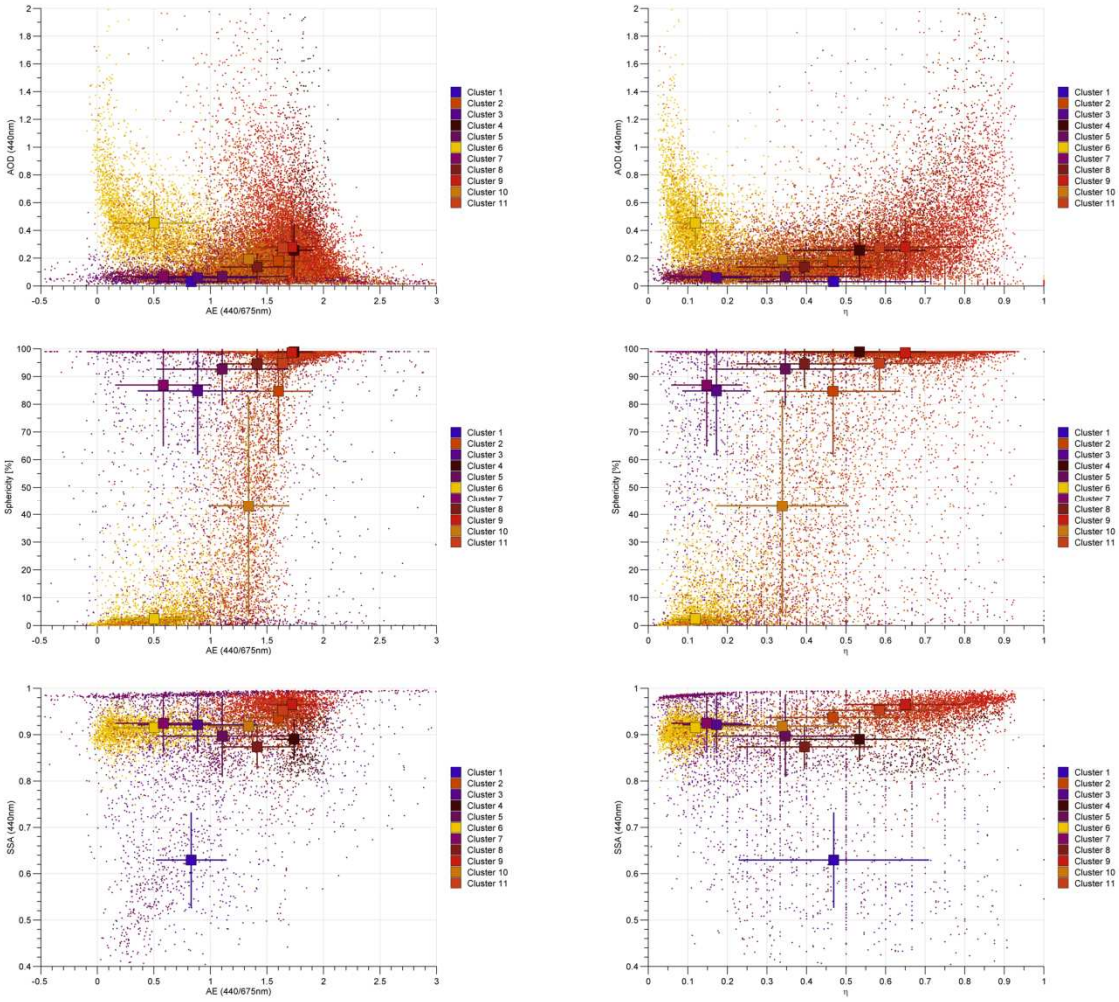


Fig. S23. Cluster means (squares) and standard deviations (lines) overlaid on the AERONET data record (points) for the season (MAM) parameterized by pairs of optical and/or microphysical parameters commonly used in aerosol classification studies.

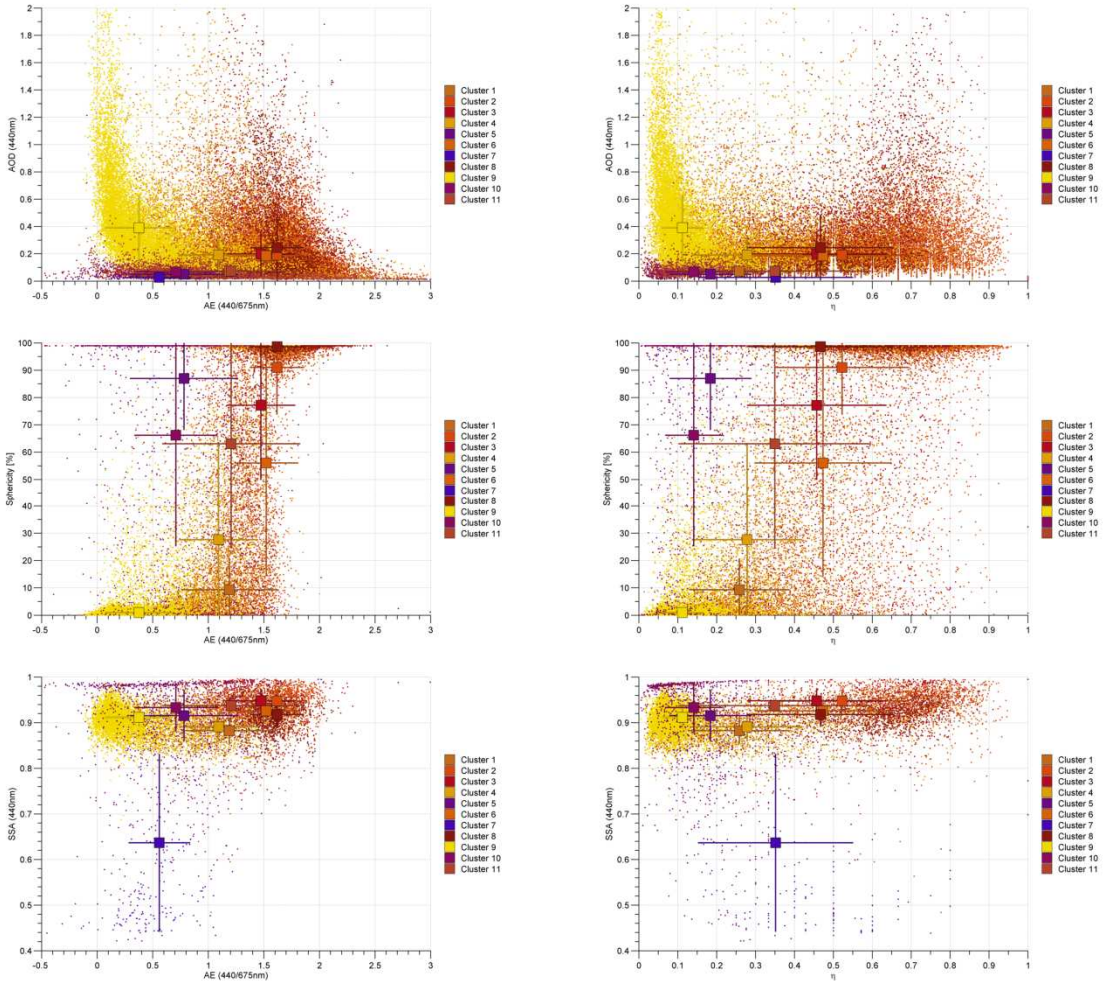


Fig. S24. Cluster means (squares) and standard deviations (lines) overlaid on the AERONET data record (points) for the season (JJA) parameterized by pairs of optical and/or microphysical parameters commonly used in aerosol classification studies.

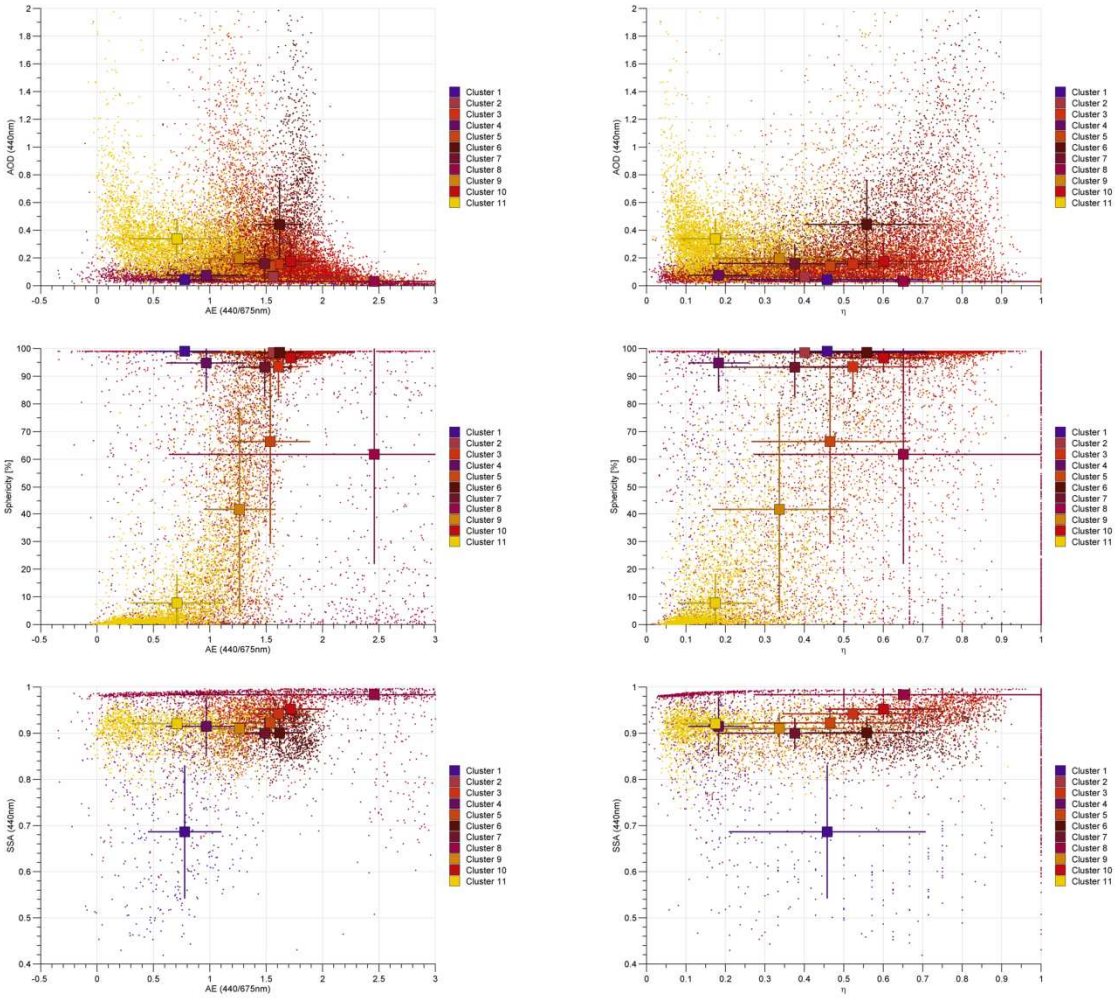


Fig. S25. Cluster means (squares) and standard deviations (lines) overlaid on the AERONET data record (points) for the season (SON) parameterized by pairs of optical and/or microphysical parameters commonly used in aerosol classification studies.