

Development of global aerosol models using cluster analysis of Aerosol Robotic Network (AERONET) measurements

Ali H. Omar,¹ Jae-Gwang Won,² David M. Winker,¹ Soon-Chang Yoon,² Oleg Dubovik,³ and M. Patrick McCormick⁴

Received 6 April 2004; revised 7 July 2004; accepted 3 August 2004; published 1 March 2005.

[1] We use radiance measurements and inversions of the Aerosol Robotic Network (AERONET) (Dubovik and King, 2000; Holben et al., 1998; Holben et al., 2001) to classify global atmospheric aerosols using the complete archive of the AERONET data set as of December 2002 and dating back to 1993 for some sites. More than 143,000 records of AERONET solar radiance measurements, derived aerosol size distributions, and complex refractive indices are used to generate the optical properties of the aerosol at more than 250 sites worldwide. Each record is used in a clustering algorithm as an object, with 26 variables comprising both microphysical and optical properties to obtain six significant clusters. Using the mean values of the optical and microphysical properties together with the geographic locations, we identified these clusters as desert dust, biomass burning, urban industrial pollution, rural background, polluted marine, and dirty pollution. When the records in each cluster are subdivided by optical depth class, the trends of the class size distributions show that the extensive properties (mode amplitude and total volume) vary by optical depth, while the intensive properties (mean radius and standard deviation) are relatively constant. Seasonal variations of aerosol types are consistent with observed trends. In particular, the periods of intense biomass burning activity and desert dust generation can be discerned from the data and the results of the analyses. Sensitivity and uncertainty analyses show that the clustering algorithm is quite robust. When subsets of the data set are randomly created and the clustering algorithm applied, we found that more than 94% of the records retain their classification. Adding 10% random noise to the microphysical properties and propagating this error through the scattering calculations, followed by the clustering algorithm, results in a misclassification rate of less than 9% when compared with the noise-free data.

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

1. Introduction

[2] Global aerosol properties are important not only for use in radiative transfer models but also as inputs to the inversion algorithms of satellite-based measurements of aerosol optical properties. As we approach an era of unprecedented global coverage of aerosol profile measurements (Geoscience Laser Altimeter System (GLAS) and Cloud Aerosol Lidar and Infrared Pathfinder Spaceborne Observations (CALIPSO)) to augment the satellite measurements of aerosol optical depths by passive instruments

(Moderate Resolution Imaging Spectrometer (MODIS), Multiangle Imaging Spectroradiometer (MISR), POLDER, Total Ozone Mapping Spectrometer (TOMS), and Stratospheric Aerosol and Gas Experiment (SAGE)), there is a need for well characterized aerosol properties. In its most recent report, the *Intergovernmental Panel on Climate Change (IPCC)* [2001] found that radiative forcing by aerosols is the most uncertain of all radiative forcing estimates. Reducing these uncertainties calls for expanded aerosol measurements and studies to characterize different types of aerosols and sources [*National Research Council (NRC)*, 1996] (see also the National Aerosol-Climate Interactions Program (NACIP) White Paper; available at <http://www-NACIP/ucsd.edu>). Tropospheric aerosols are diverse and their properties depend on sources, emission rates, and removal mechanisms which can be highly variable. To understand and quantify aerosol effects, there have been numerous domestic and international campaigns to characterize aerosol physical and chemical properties and processes. These include the Aerosol Characterization Experiments (ACE-1, ACE-2, and ACE-Asia), the Tropospheric Aerosol

¹NASA Langley Research Center, Atmospheric Sciences, Hampton, Virginia, USA.

²School of Earth and Environmental Sciences, Seoul National University, Seoul, Korea.

³Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, Maryland, USA.

⁴Center for Atmospheric Sciences, Hampton University, Hampton, Virginia, USA.

Radiative Forcing Observational Experiment (TARFOX), the Smoke, Clouds, and Radiation-B (SCAR-B) experiment, and the Indian Ocean Experiment (INDOEX). ACE-1 took place south of Australia in November–December 1995 and measured properties of the natural aerosol in the remote marine boundary layer [Bates *et al.*, 1998]. ACE-2 took place in the north Atlantic ocean in June–July 1997, and focused on the radiative effects and processes controlling anthropogenic aerosols from Europe and desert dust from Africa as they were transported over the north Atlantic ocean [Raes *et al.*, 1999]. ACE-Asia took place during the spring of 2001 off the coast of China, Japan and Korea. This region includes many types of aerosols of widely varying composition and sizes derived from one of the largest aerosol source regions on Earth. TARFOX [Russell *et al.*, 1999], designed to measure and analyze aerosol properties and effects along the US eastern seaboard, took place on 10–31 July 1996. INDOEX [Ramanathan *et al.*, 2001] measured aerosol properties over the tropical Indian Ocean in 1998 and 1999. The large aerosol optical thicknesses (~ 0.5) and the prominent role of the carbonaceous aerosols in the extinction budget during most of INDOEX underline the need to develop long-term records of specific species.

[3] In many cases, mean properties sorted by location or type to represent aerosol characteristics are used in radiative transfer calculations [Kiehl and Briegleb, 1993; Kiehl and Rodhe, 1995; Nemesure *et al.*, 1995; Penner *et al.*, 1992] or inversion algorithms of satellite measurements [Tanré *et al.*, 1999]. Assigning a set of mean aerosol physical and chemical properties to a given location based on a long-term average has significant shortcomings. This is because at any given location, the aerosol type can be variable on timescales as short as a few hours [Sheridan *et al.*, 2001]. These variations result from transport of distinct air masses to a site and nonsystematic events such as fires, wind gusts, hurricanes, tornadoes, and land clearing and development activities. These variations lead to diverse aerosol characteristics at each site on timescales as short as a few hours and preclude the long term average of properties from being a good representation of the characteristics for a site or region. On the other hand, the frequency of occurrence of a given type of aerosol at a location is an indication of the likelihood of finding that type of aerosol at that location if an adequately large sample (years) is used to calculate this frequency. Aerosol optical measurements must therefore be made at short timescales (on the order of a few hours) to develop a large database which can be used to derive statistically significant correlations. The Aerosol Robotic Network (AERONET) [Holben *et al.*, 1998, 2001] measurements provide a database with a fine temporal resolution albeit for column rather than vertically resolved measurements. AERONET is an automatic robotic Sun and sky scanning measurement network that has grown rapidly to over 200 sites worldwide. AERONET uses multiangle radiance measurements to retrieve the discrete aerosol size distributions in 22 size bins ranging from 0.05 to 15 μm and the complex refractive index. The network has the important features of uniform data collection, calibration, and data processing procedures. This study uses the whole AERONET archive (up to December 2002) of measurements and inversions to develop a type-specific set of mean

optical properties of aerosols. Cluster analysis is used for categorization of atmospheric aerosol types. Six significant types are suggested by the cluster analysis which we identify as: desert dust, biomass burning, polluted continental, clean continental, polluted marine aerosol, and dirty pollution. In this classification, clean continental refers to a lightly loaded soot-free pollution normally found in rural areas and is good approximation for background aerosol. Dirty pollution refers to pollution containing significant amounts of absorbing species.

2. Data Screening and Preparation

[4] An attempt is made to use all the available AERONET measurements dating back to 1993 for some sites. These unscreened measurements are frequently contaminated by clouds and depending on the cloud reflectivity, can have a significant effect on the Sun-sky radiance measurements. This study used the AERONET level 1 size distribution data and applies a two-part cloud screening scheme. The first part checks the symmetry of the almucantar measurements and the second part is a statistical screening procedure. The almucantar measurement is made at several azimuthal angles with the elevation angle of the direct Sun. For the aureole measurement, the increment of angle change is set to be quite small near the direction of the Sun. In order to ensure that the sky is clear at the time of the measurement, we calculate the relative error between the seven pairs of data measured on either side of the direct Sun. The azimuthal angles of these measurements are 2.0°, 2.5°, 3.0°, 3.5°, 4.0°, 5.0°, and 6.0°. Using this method ensures that the sky is clear near the measurements because the presence of cloud cover would result in disparities between symmetrical pairs of measurements.

[5] The statistical screening procedure is based on the cloud screening method for the direct solar measurement of the AERONET Sun/sky radiometer. In this procedure, records with optical depths (τ) and Angstrom coefficients (\AA) that exceed a fixed number of standard deviations (σ) on either side of the mean of the distribution are not included in the analyses. For this study, the range of acceptable optical depths is $\tau_{\text{mean}} - \sigma$ to $\tau_{\text{mean}} + 3\sigma$. The acceptable Angstrom coefficients are greater than $\text{\AA}_{\text{mean}} - 3\sigma$. These conditions assume that unrealistically large aerosol extinction values are due to cloud contamination or other transient phenomena. A very small Angstrom coefficient is also indicative of cloud contamination. This cloud screening scheme on the average rejects about 60% of the data.

3. Cluster Analysis

[6] Cluster analysis is a statistical tool used for grouping large data sets into several categories using predefined variables. In this study, cluster analysis by partitioning [cf. Kaufman and Rousseeuw, 1990] is used to categorize the AERONET data set based on several optical and physical characteristics of the aerosol. The cluster algorithm uses the 26 parameters in Table 1. Note that the optical depths are not used in the cluster analysis because these depend on the amount of aerosol rather than type. The first two columns of the table are properties from the AERONET inversion algorithms. In the third column are optical properties that

Table 1. Parameters Used in the Cluster Analysis^a

Composition	Size Distribution ^b	Optical Properties
Complex refractive index (8)	mean radius (2)	single scattering albedo (4)
	standard deviation (2) mode total volume (2)	asymmetry factor (4) extinction/backscatter ratio (4)

^aThe numbers in parentheses denote the number of variables for the given property, e.g., there are two geometric mean radii, a fine and coarse radius. Note that the number of variables for the optical properties is four denoting the 441, 673, 873, and 1022 nm wavelengths of the AERONET measurements.

^bThe size distribution parameters are based on a bimodal log normal distribution.

can be generated from AERONET inversions of size distribution and complex refractive indices using scattering calculations such as Mie modeling [Dubovik and King, 2000; Mie, 1908], T-matrix [Mishchenko et al., 1995] or geometric-optics integral equations [Dubovik et al., 2002b; Yang and Liou, 1996] depending on the scattering calculation used in the AERONET inversion algorithm. For this data set, the optical properties were generated using Mie calculations. After clustering, the individual sites are inspected to find the frequency of records in the categories. A site belongs to a category x if more than 30% of the records at that site are grouped in cluster x . It is therefore possible for a site to belong to more than one category, i.e., the site experiences a high frequency of more than one type of aerosol.

[7] The AERONET algorithm for size distribution retrieval provides the volume distribution data of 22 size bins ($dV/d\ln r$, V is volume, and r is radius) from $0.05 \mu\text{m}$ to $15 \mu\text{m}$. The best fit for the size distribution data is a two-mode log-normal size distribution described by equation (1),

$$\frac{dV}{d\ln r} = \frac{C_f}{\sqrt{2\pi} \ln \sigma_f} \exp\left(\frac{[\ln r - \ln r_{mf}]^2}{2(\ln \sigma_f)^2}\right) + \frac{C_c}{\sqrt{2\pi} \ln \sigma_c} \cdot \exp\left(\frac{[\ln r - \ln r_{mc}]^2}{2(\ln \sigma_c)^2}\right), \quad (1)$$

where the C_s are total mode volume ($\mu\text{m}^3/\mu\text{m}^2$) and the subscripts f and c denote fine and coarse modes, respectively. σ is the geometric standard deviation and r_{mf} (r_{mc}) is the geometric fine (coarse) mean radius. This partition of the size distribution into fine and coarse modes yields six parameters by which the size distribution can be described.

[8] Natural and anthropogenic aerosols can be classified into five main types: marine aerosol, desert dust, biomass burning aerosol, urban aerosol and rural background aerosol. This forms the minimum number of aerosol clusters. The actual number of clusters is determined by calculating the clusters using successively larger cluster numbers until the calculation does not yield any new significant clusters. For the AERONET records we found that six categories were the most representative. The clustering algorithm involves the following main steps. First the number of clusters, k , is specified. The algorithm then calculates the mean (μ) and standard deviation (σ) of each of the 26 variables. For each variable k points are randomly chosen between $\mu + \sigma$ and $\mu - \sigma$. These points form the initial centers of the k clusters. For a given number of

clusters (k), the results are not sensitive to the initial “seed” points. The only difference caused by varying the chosen starting point is the number of iterations it takes to converge. Our choice of points within one standard deviation ensures that convergence will be achieved relatively quickly. The next step is to calculate the Euclidean distance of each record from the center of the k clusters. Initially each record is assigned to the cluster whose center is “closest” (using a Euclidean distance metric) to the record. When all records have been assigned to individual clusters, new cluster centers are determined by averaging the variables in each cluster. The data clusters formed in this way group all records that have statistically significant similarities in one category. The process is repeated until the change in position between the old and new centers (shift) is less than a prescribed threshold value, i.e., the position of the new center is unchanged. The threshold value of the shift in the cluster centers was set at 0.1%. Figure 1 shows a flowchart of the clustering algorithm. The objects in the clustering algorithm are the cloud-screened AERONET inverted products (size distributions and complex refractive indices) and optical properties determined using scattering calculations. The distance (normalized by the standard deviation to eliminate bias resulting from the different magnitudes of the variables) between a record and the center of category j , is calculated using equation (2),

$$\text{normalized distance} = \sum_{i=1}^{i=26} \frac{\text{data record}(i) - \text{center}(j, i)}{(\text{standard deviation})^2}, j = 1, 6 \quad (2)$$

Convergence of the clustering algorithm is achieved in less than 40 iterations for the various initial conditions.

4. Clusters of Aerosol Types

[9] The AERONET data set yielded six distinct clusters described by the column optical and physical properties. Six clusters provided the largest number of clusters in which each cluster had a reasonable number of members. All the six clusters were well populated, i.e., the number of members of each cluster is at least 4% of the total. When we go to seven or eight clusters, the smallest cluster is composed of a very small number of points. As we go to a larger number of clusters, the members of the main clusters are relatively unchanged. The clusters do however lose a few members to form one or two new clusters. On closer examination, these additional clusters do not look real, e.g., geographic locations of the records are inconsistent with the type of aerosol expected at that location. Ideally, the algorithm should be objective and automated if cluster analyses are to be used to classify optical measurements for the purpose of assigning types to aerosols in near real time. Though the initialization and choice of the number of clusters is subjective in our algorithm, the consistency and error analysis described in sections 6 and 8, respectively, show that the method is quite robust. In general, a well partitioned data set with distinct groups should have “tight” clusters, i.e., clusters with small n -dimensional spheres, and the centers of these spheres should be as far from each other as possible (well separated groups).

[10] After the clustering algorithm has generated aerosol categories, each category is assigned an aerosol type.

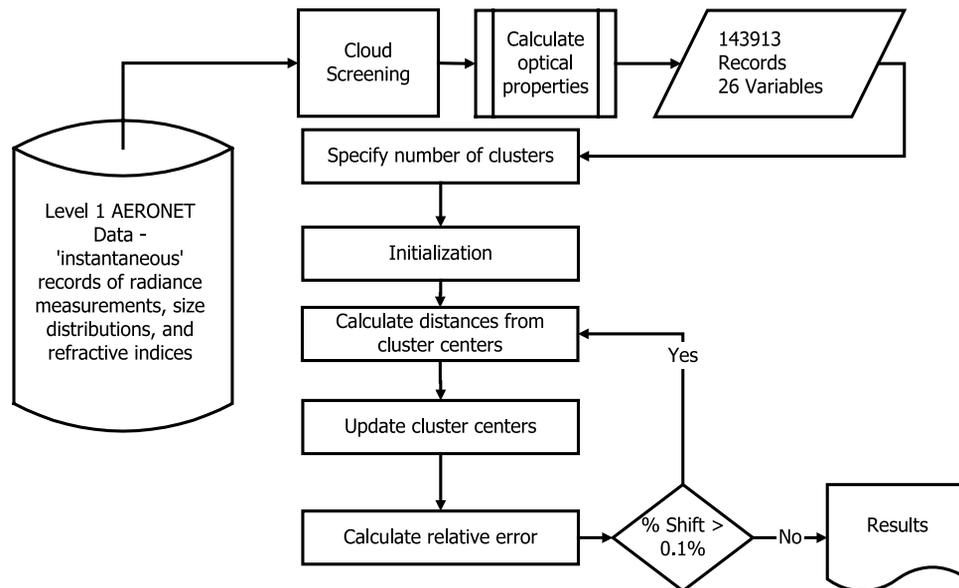


Figure 1. Flowchart of the clustering algorithm used to group 143,913 AERONET records into clusters of with similar composition, size distribution, and optical properties. The convergence is achieved when the cluster centers in 26-parameter space do not change by more than 0.1% in successive iterations.

Aerosol typing of the categories relies on type-indicative characteristics such as the fine and coarse fractions, particle size, optical depth, geographic location and in some cases seasonal variation. The effects of nonsphericity on the cluster analysis are unknown but not negligible. Table 2 shows the properties of the six clusters. The values shown in the table are some of coordinates in 26-parameter space of the centers (medoids) of the clusters. The membership coefficient is the mean Euclidean distance between the center of each cluster and every member of the cluster normalized by standard deviation of each variable. Therefore the smaller the membership coefficient, the better the correlation between the records within a cluster.

[11] For category 1, the membership coefficient of 0.92 is the average value of normalized distance between the center of category 1 and 22,202 points. The properties of the medoids along with the geographic locations of the sites

suggest that categories 1 to 6 are desert dust, biomass burning, rural (background), industrial pollution, polluted marine, and dirty pollution, respectively. Table 2 shows the number of records in each cluster. The category 5 (marine) cluster has the least number of records ($\sim 5\%$ of the total) while category 4 (continental pollution) has the largest number ($\sim 40\%$) of the total. This indicates that the locations of AERONET sites worldwide are more likely to experience continental pollution more than other aerosol types. In general, anthropogenically influenced categories (categories 2, 4, and 6, biomass burning, continental pollution, and dirty pollution, respectively) account for two thirds of the records. Natural sources (categories 1 and 5, dust and marine, respectively) account for 20% of the records and the remaining 14% is attributable to rural/background (category 3) aerosol types.

[12] The AERONET sites belonging to category 1 are Banizoumbou, Ouagadougou, Cape Verde, Ilorin, Solar

Table 2. Summary of the Cluster Analyses Results by Category^a

	Category 1	Category 2	Category 3	Category 4	Category 5	Category 6
Number of records	22,202	26,662	20,307	55,667	6527	12,548
Single scattering albedo (673 nm)	0.93	0.80	0.88	0.92	0.93	0.72
Real refractive index (673 nm)	1.4520	1.5202	1.4494	1.4098	1.3943	1.4104
Imaginary refractive index (673 nm)	0.0036	0.0245	0.0092	0.0063	0.0044	0.0337
Optical depth (673 nm)	0.327	0.190	0.036	0.191	0.140	0.100
Angstrom coefficient (441/673)	0.608	1.391	1.534	1.597	0.755	1.402
Angstrom coefficient (673/873)	0.486	1.332	1.381	1.536	0.678	1.232
Angstrom coefficient (873/1022)	0.277	1.043	0.950	1.290	0.531	0.846
Asymmetry factor (673 nm)	0.668	0.603	0.580	0.612	0.711	0.594
Fine mean radius, μm	0.117	0.144	0.133	0.158	0.165	0.140
Geometric standard deviation (fine)	1.482	1.562	1.502	1.526	1.611	1.540
Fine total volume, $\mu\text{m}^3/\mu\text{m}^2$	0.077	0.040	0.013	0.061	0.029	0.032
Coarse mean radius, μm	2.834	3.733	3.590	3.547	3.268	3.556
Geometric standard deviation (coarse)	1.908	2.144	2.104	2.065	1.995	2.134
Coarse total volume, $\mu\text{m}^3/\mu\text{m}^2$	0.268	0.081	0.020	0.054	0.083	0.034
Fine fraction by volume	0.22	0.33	0.38	0.53	0.26	0.49
Membership coefficient	0.92	0.93	0.94	0.93	0.91	0.89

^aThe values are the center of each cluster. The membership coefficient is an indication of the “tightness” of each cluster.

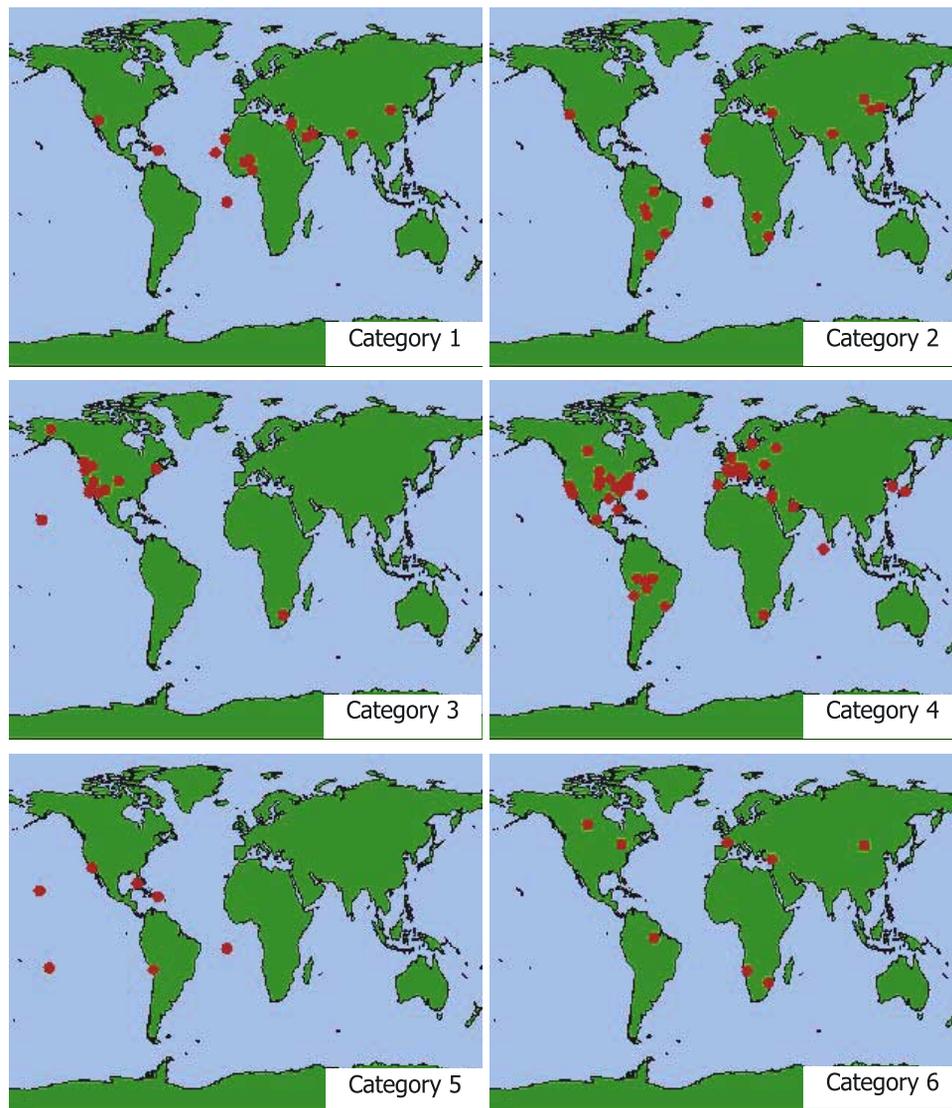


Figure 2. The locations (denoted by red dots) of the six categories. The locations and properties of the categories 1, 2, 4, 5, and 6 are consistent with desert dust, biomass burning, continental pollution, polluted marine, and dirty pollution, respectively.

Village, Bahrain, Kanpur, Dahkla, Sede Boker, Nes Ziona, Maricopa, Ascension Island, La Paguera, and Yulin shown in Figure 2. All these sites are in desert regions, close to deserts, or in regions where desert dust as a result of long-range transport has been observed [Colarco *et al.*, 2003; Di Iorio *et al.*, 2003; Dubovik *et al.*, 2002a; Tanré *et al.*, 2001]. Figure 3a shows a mean radius and geometric standard deviation for the fine (coarse) mode of $0.12 \mu\text{m}$ and 1.5 ($2.8 \mu\text{m}$ and 1.91), respectively, for this category. The fine fraction by volume is 0.22 , i.e., coarse sizes are dominant. The medoid properties of these aerosols are consistent with other measurements of dust aerosols. The global aerosol model of *d'Almeida et al.* [1991] estimate dry accumulation and coarse mode radii for dust of $0.27 \mu\text{m}$ and $4.00 \mu\text{m}$, respectively. Tanré *et al.* [2001] used spectral measurements from both Landsat TM and AERONET to find size distributions for dust aerosols of $0.5 \mu\text{m}$ for the fine mode and $1\text{--}5 \mu\text{m}$ for the coarse mode. Real refractive indices from the same study ranged from $1.46\text{--}1.52$.

[13] Using airborne particle counters, a ground based lidar and sunphotometer, *Di Iorio et al.* [2003] estimated real and imaginary parts (m_r and m_i) of the refractive indices of $(1.52 \text{ to } 1.58) - (0.005 \text{ to } 0.007)i$ for dust aerosols. In the same study they estimate single scattering albedos ranging from 0.71 to 0.75 . Category 1 aerosols have a mean single scattering albedo of 0.93 which is closer to the values reported by others [cf. Tanré *et al.*, 2001]. In studies of the mineralogical composition of dust, *Sokolik and Toon* [1999] report m_r values of 1.5 for kaolinite, quartz, calcite, and gypsum and 1.35 for illite. In the same study, the m_i ranged from 0.000044 for kaolinite and montmorillonite to 0.001 for illite and nearly 1.0 for hematite at wavelengths near 600 nm . These studies show that there is intricate spatial variability of the composition of dust aerosols depending on sources, transport and mixing which cannot be resolved by AERONET measurements and these analyses alone. In general the elemental composition of dust from various geographical locations are highly variable resulting in significant differences in the complex refractive indices.

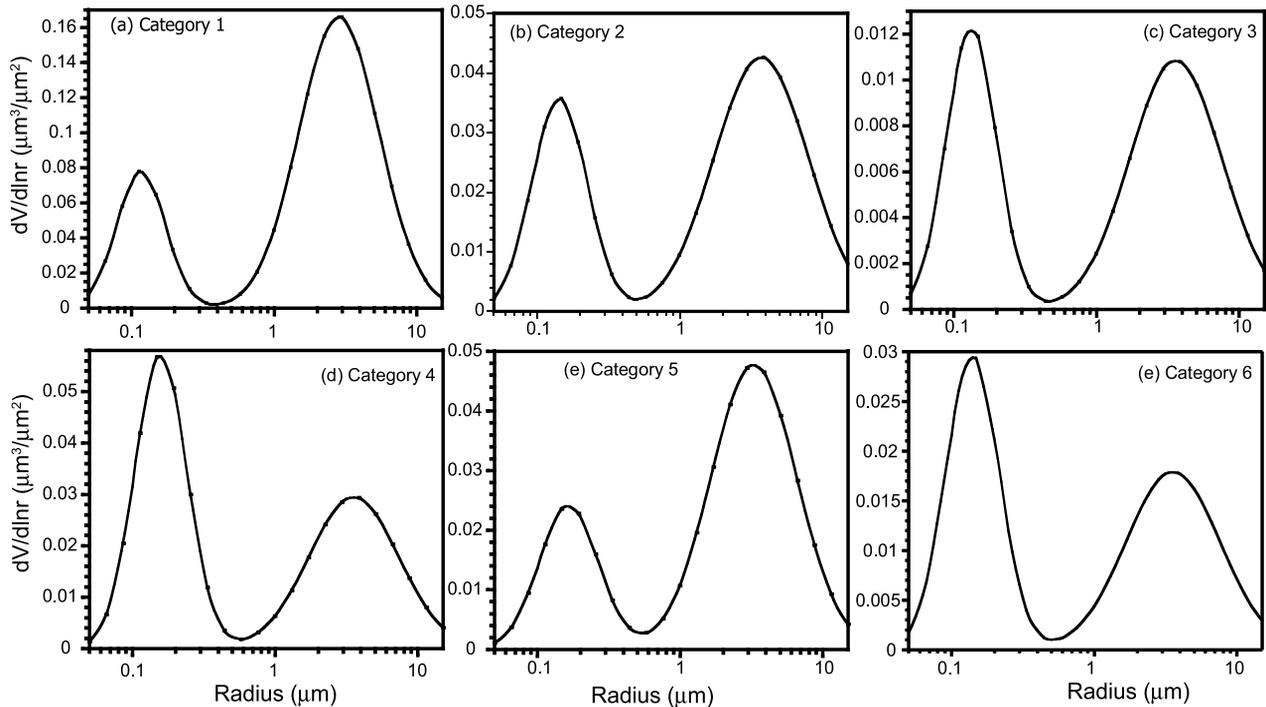


Figure 3. Size distributions of the cluster medoids of the six aerosol categories.

As has been noted by others [e.g., Sokolik *et al.*, 2001], the characterization of dust aerosols will require a concerted effort involving in situ and ground based measurements, and satellite remote sensing at targeted regions.

[14] Category 2 sites include Skukuza, IMS Metu Erdemli, Abracos Hill, Mongu, Cuiaba, Los Fieros, Yulin, CEILAP-BA. Figure 3b shows the size distributions described by fine (coarse) radius and standard deviation of 0.14 μm , and 1.56 (3.7 μm , and 2.1), respectively. The fine fraction based on the volume distribution is 0.33. Real and imaginary parts of the refractive index of 1.52 and 0.02 are consistent with measurements of Dubovik *et al.* [2002b] and Ansmann *et al.* [2000]. Furthermore, a comparison of the properties of category 2 aerosols are comparable to the observations of predominantly fine mode aerosol observed at the IMS Metu Erdemli site and believed to have a significant biomass burning component [Kubilyay *et al.*, 2003].

[15] Using spectrally resolved particle backscatter and extinction measurements, the effective particle radius of free tropospheric aerosol advected from the Indian subcontinent and primarily composed of biomass burning pollution is estimated at 0.17 μm [Ansmann *et al.*, 2000]. Müller *et al.* [2000] found effective radii ranging between 0.14 and 0.22 μm , complex refractive indices of $1.65 - (0.03 \text{ to } 0.08)i$ and single scattering albedos of 0.79 to 0.86 at 532 nm for similar aerosol plumes. A series of measurements of both microphysical and optical properties of biomass burning aerosols made during the SCAR-B campaign [Kaufman *et al.*, 1998] found values of size distributions ($r_f \sim 0.13 - 0.17 \mu\text{m}$, $\sigma_f \sim 1.5 - 1.8$), single scattering albedos (0.79–0.85), and real refractive indices (1.37–1.55), not inconsistent with the values reported for category 2 in Table 2.

[16] The category 3 aerosol is characterized by a low optical depth (0.04) and a high frequency of incidence in

areas where the atmosphere is expected to be relatively clean, i.e., clean continental or rural background. The members of this category include Mauna Loa, Railroad Valley, Saturn Island, Sevilleta, Rogers Dry Lake, Bonanza Creek, Rimrock, Maricopa, H. J. Andrews, Howland, Konza EDC, Bethlehem, San Nicolas shown in Figure 2. The fine fraction by volume for category 5 is 0.4. Figure 3c shows the size distribution described by a fine (coarse) radius and standard deviation of 0.13 μm , and 1.50 (3.6 μm , and 2.1), respectively, for this category. Most of these sites are continental and are mostly in the western United States. Though Mauna Loa is an oceanic site, the site elevation of 3397 m above mean sea level means measurements are above the marine boundary layer most of the time and explains why the site experiences mostly background aerosol. Though this category has a large number of records (20,307), the low optical depth of the measurements means necessarily that the retrieval accuracy is significantly deteriorated. For all AERONET measurements the accuracy of the single scattering albedo, and refractive indices decreases significantly for optical depths of less than 0.2 (at 441 nm) [Dubovik *et al.*, 2002a, 2000]. It is therefore likely that the single scattering albedos reported in Table 2 are unrealistically low.

[17] Category 4 aerosols are most prevalent at Walker Branch, Venice, ISDGM CNR, Moscow MSU MO, MD Science Center, Columbia SC, GSFC, Wallops, SERC, Avignon, UCLA, Cart Site, Mexico City, Toulouse, Anmyon, Lille, Moldova, Stennis, Rome Tor Vergat, Konza EDC, Rio Branco, Nes Ziona, Ispra, Shirahama, La Jolla, Bondville, COVE, Dry Tortugas, Alta Floresta, Arica, Fresno, GISS, Sao Paulo, Gotland, Bethlehem, Abracos Hill, Bahrain, Sioux Falls, Los Fieros, Corcoran, Sede Boker, Kaashidhoo, El Arenosillo, Bermuda, Bordeaux, Mongu, Waskesiu. The fine fraction by volume is 0.53

and the size distribution is described by a fine (coarse) radius and standard deviation of $0.16\ \mu\text{m}$, and $1.53\ (3.5\ \mu\text{m}, \text{ and } 2.1)$, respectively (Figure 3d). The location of the aerosols (near or in urban centers), and the composition (mean refractive index of $1.41-0.006i$) are consistent with urban pollution consisting of mainly sulfate particles with a small absorbing (soot) component.

[18] Category 5 sites are Lanai, Tahiti, Arica, La Paguera, Dry Tortugas, Ascension Island, La Jolla. The fine fraction is 0.26 and the refractive index is $1.39-0.004i$ at $673\ \text{nm}$. The global aerosol models of *d'Almeida et al.* [1991] estimate an oceanic aerosol refractive index of 1.38 with a negligibly small imaginary part. All the sites are islands or coastal sites. The size distribution is described by a fine (coarse) radius and geometric standard deviation of $0.17\ \mu\text{m}$, and $1.6\ (3.3\ \mu\text{m}, \text{ and } 2.0)$, respectively (Figure 3e). In case studies of three clean maritime sites, *Smirnov et al.* [2003] found effective fine and coarse mode radii ranges of $0.11-0.14\ \mu\text{m}$ and $1.8-2.1\ \mu\text{m}$, respectively, and complex refractive indices of $1.37-0.001i$ resulting in a single scattering albedo (ω) of 0.98. All these parameters except the imaginary index of refraction (and consequently the ω) are consistent with the mean values of this cluster. While the coarse mode is seasalt, the fine mode, whose contribution is most likely responsible for the enhancement of the imaginary part of the total refractive index, is most likely biomass burning aerosol or from other anthropogenic sources. It is therefore likely that category 5 is marine aerosol mixed with biomass burning smoke or industrial pollution. We refer to this type as polluted marine aerosol.

[19] Category 6 sites are Egbert, Skukuza, Bordeaux, Etosha Pan, Belterra, IMS Metu Erdeml, Waskesiu, and Dalanzadgad. Three of these sites (Skukuza, IMS Metu Erdeml, and Dalanzadgad) are also found in category 2. This class of aerosols (referred to as dirty pollution) is characterized by a size distribution (Figure 3f) similar to category 4 (industrial pollution) with a larger imaginary index of refraction (0.034) and small single scattering albedos. The low single scattering value (0.72) implies that these are most likely aerosols with a large fraction of elemental carbon. In addition to the mixing mechanism of soot, single scattering albedo depends on the age of the smoke, the combustion phase and fuel moisture content. The single scattering albedo for fresh smoke plumes near forest fires ranges from 0.35 to 0.9 with a mean value of 0.75 for flaming combustion and 0.82 for smoldering combustion [*Reid et al.*, 1998]. It is therefore likely that category 6 represents flaming combustion while category 2 is the more frequent smoldering type of combustion. Also note that category 6 aerosols have relatively thin layers which necessarily means that uncertainties in the single scattering albedo could be as high as 0.07 [*Dubovik et al.*, 2002a]. The likelihood that the low single scattering albedos for category 6 are not an artifact of low solar zenith angles is explored below.

[20] Figure 4 shows cumulative distributions of (Figure 4a) the optical depth, (Figure 4b) Angstrom coefficients, (Figure 4c) imaginary part of the refractive index, and (Figure 4d) single scattering albedo. At least 50% of the category 1, 2, and 4 aerosol types have an optical depth larger than 0.1 and therefore the retrievals of the microphysical properties have reasonable uncertainties. The distributions of

the Angstrom coefficients shows that more than 60% of category 1 aerosols (dust) have Angstrom coefficients smaller than 0.5, i.e., large particles. The next smallest distribution of Angstrom coefficients are category 5 aerosols (polluted marine). Angstrom coefficients of both category 4 (industrial pollution) and category 2 (biomass burning) are mostly between 1 and 2. Angstrom coefficients of 70% the category 4 aerosols fall between 1 and 2, and more than 50% of the category 2 aerosols fall in this Angstrom coefficient range. This is the expected Angstrom coefficient range for small particles.

[21] The distributions of the imaginary part of the refractive indices shows that categories 1 (dust), 4 (continental pollution) and 5 (polluted marine) have imaginary parts of the refractive indices smaller than 0.02 nearly all the time, while categories 2 (biomass burning) and 6 (dirty pollution) Angstrom coefficients are greater than 0.02 for more than 60% of the records. More than half of the dirty pollution records have single scattering albedos of less than 0.7. Dust, polluted marine, and polluted continental have the highest single scattering albedos and the variation in the frequency of these types is very small as shown in the figure.

[22] Figure 5 shows the distributions of the solar zenith angles and the single scattering albedos for category 6 aerosols. The mean solar zenith angles for all category 6 records is 60.1 degrees with a standard deviation of 12.9 degrees. Most of the measurements, therefore, have a small error due to low solar zenith angles. The mean single scattering albedo of 0.72 with a standard deviation of 0.06 does not show any significant bias due to low solar zenith angle values. The largest uncertainty in these values is due to the retrieval uncertainties associated with the low optical depths (mean optical depth of category 6 is 0.1).

5. Seasonal Variation of Aerosol Types at Selected Sites

[23] The clustering algorithm assigns each cloud-screened and quality-qualified record in the database to one of the six clusters and therefore affords an easy method of examining the seasonal variation of aerosol types at each individual site. The seasonal variation also aids in identifying the type of aerosol in the six clusters at sites where this variation is known. For this paper we examine the seasonal variations at a few sites where one or more of the six categories is either dominant or shows a distinct variation. To do this, we plot the percentage frequency of the records in categories at each of the selected sites, i.e., how many times does a record get classified in a category in a given month, normalized by the total number of records for that month and averaged over the lifetime of the measurements at the site.

[24] Figure 6a illustrates the classification of the aerosol records at Cape Verde between October 1994 and December 2002. The Cape Verde ($16\text{N}, 22\text{W}$) site is an oceanic site off the west coast of Africa and directly in the path of westerly outflows of Saharan dust off the coast of Africa. The figure shows that the aerosol type is almost exclusively category 1 (desert dust) with a small intrusion of category 2 (biomass burning) in May. There is also evidence of a small amount of category 5 (polluted marine) in December. These variations are consistent with the seasonal trends of the aerosol type described by others [*Chiapello et al.*, 1997; *Tanré et*

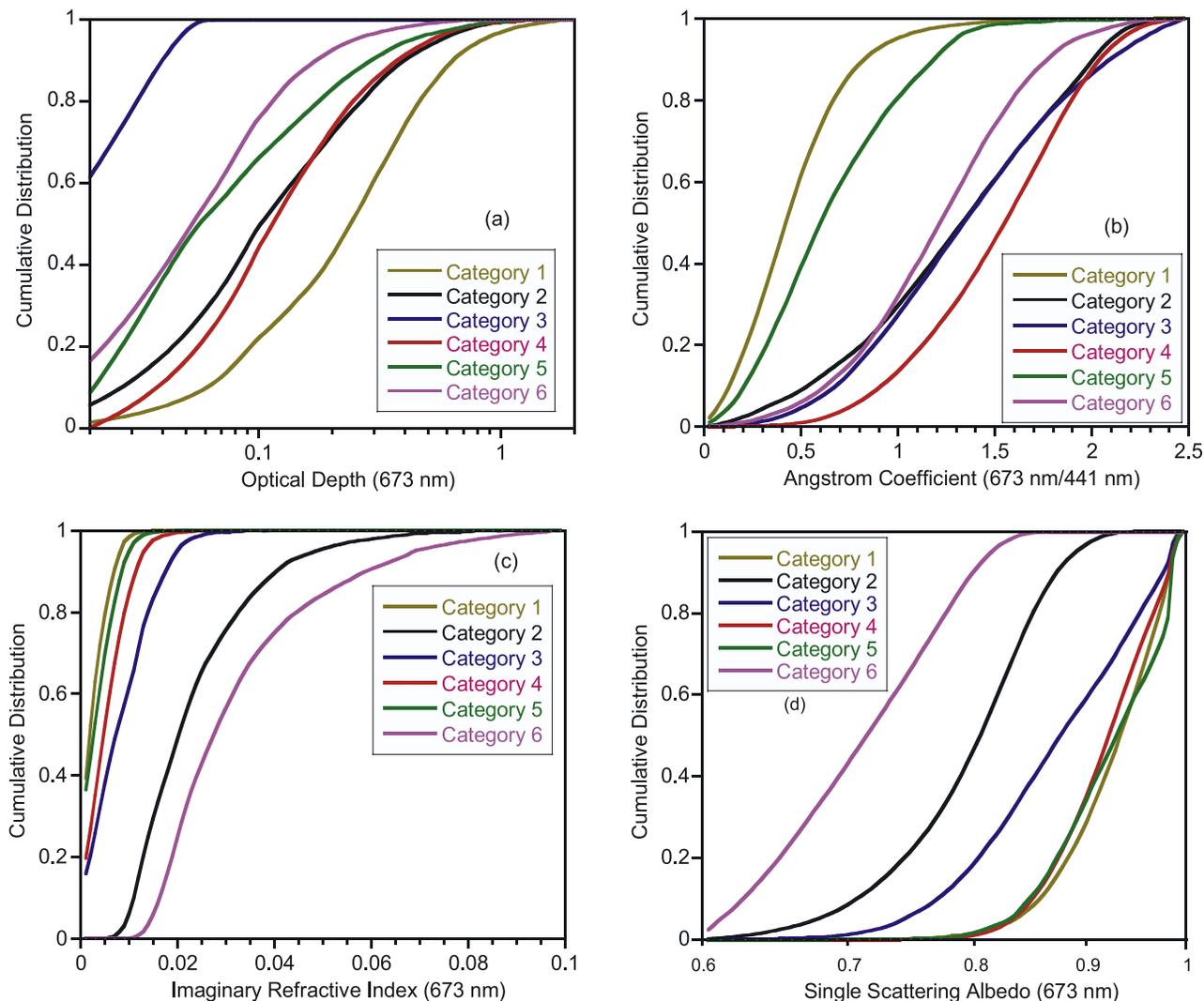


Figure 4. (a) Cumulative distributions of the optical depth, (b) Angstrom coefficient, (c) single scattering albedo, and (d) imaginary refractive index of the six categories of aerosol found by cluster analysis.

al., 2003] at Cape Verde. Banizoumbou (13N, 2E) is an inland site near the southern edge of the Sahara desert. The site experiences a high fraction of desert dust throughout the year as shown in Figure 6b. This site is more likely to experience category 1 aerosols except in December when there is a small fraction of biomass burning aerosol (category 2). The seasonal variation shown in the figure has been derived from measurements between October 1995 and December 2002. Figure 6c shows the trend of aerosol type frequencies at Solar Village (25N, 46E), a site in the Arabian desert. The frequencies are based on measurements between February 1999 and December 2002 and show the desert dust fractions peaking in the months of April and May. Also note that between the months of October and February there is a comparable fraction of polluted continental aerosol (category 4). This is expected since this site is in close proximity to the Riyadh metropolitan area (population 4.7 million) in heartland of the Arabian peninsula.

[25] Figure 6d shows the aerosol classification at Mongu (15S, 23E). The aerosol types derived from measurements at this site between June 1995 and December 2002 are

predominantly biomass burning (category 2) in May–October and polluted continental in October–March. In the months of March, April, and May there is a significant fraction of the highly absorbing aerosol type (dirty pollution, category 6). The trends for Mauna Loa (19N,155W), a

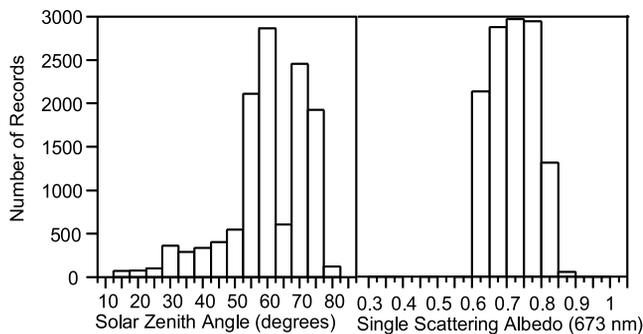


Figure 5. Distributions of the solar zenith angle and the single scattering albedos for the records in category 6.

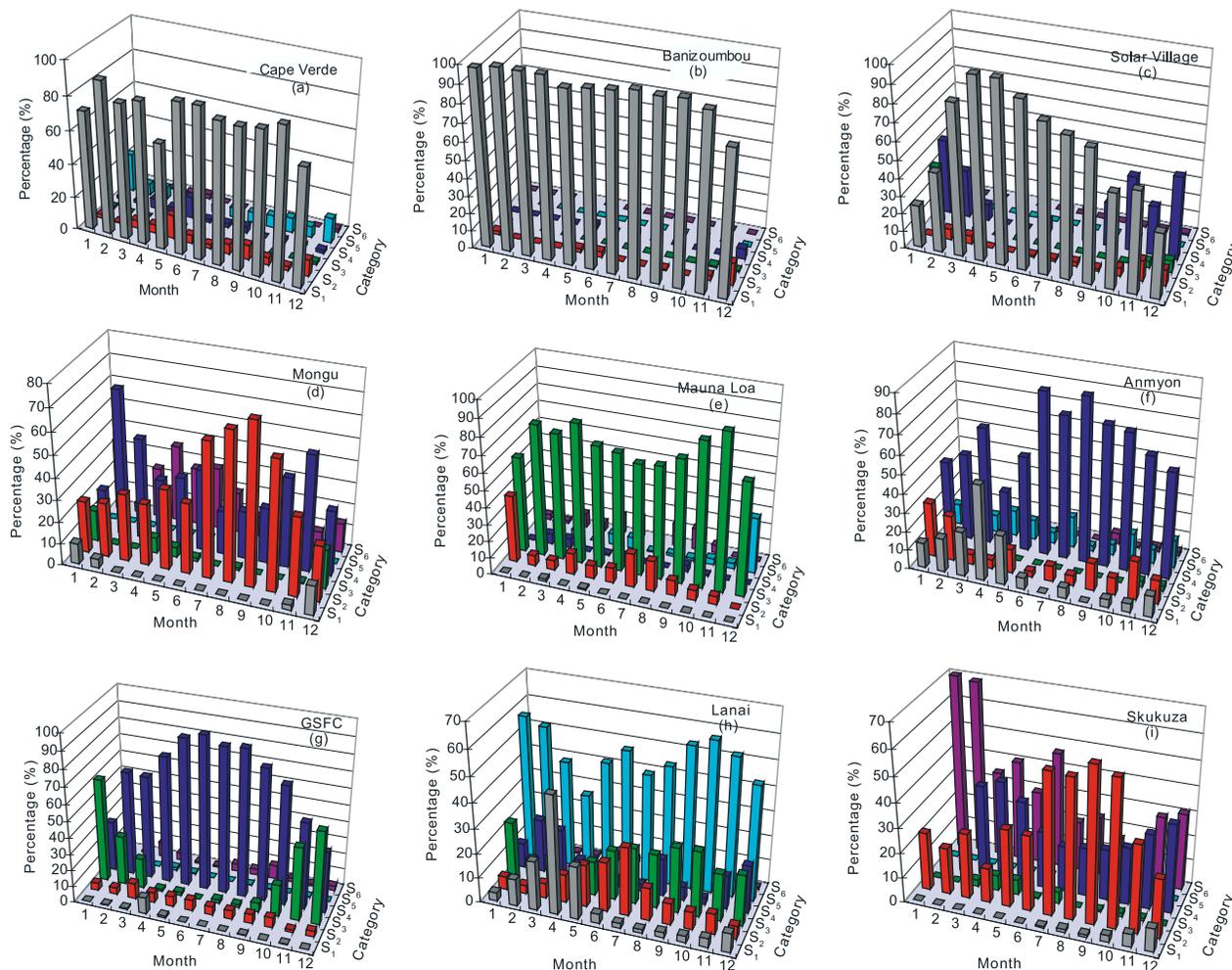


Figure 6. The variation of six categories of aerosol (desert dust, biomass burning, background, polluted continental, polluted marine, and dirty pollution, denoted by S_1 , S_2 , S_3 , S_4 , S_5 , and S_6 , respectively) at (a) Cape Verde between October 1994 and December 2002, (b) Banizoumbou between October 1995 and December 2002, (c) Solar Village between February 1999 and December 2002, (d) Mongu between June 1995 and December 2002, (e) Mauna Loa between June 1994 and December 2002, (f) An Myon between September 1999 and March 2002, (g) GSFC between May 1993 and December 2002, (h) Lanai between November 1995 and December 2002, and (i) Skukuza between July 1998 and December 2002.

category 3 site (clean continental) are shown in Figure 6e. These are characterized by low optical depths (<0.05) and though the retrievals of the microphysical properties may be uncertain, the seasonal variation and the locations of the sites belonging to this category are consistent with relatively clean air. The monthly trend shows that the atmosphere is clean most of the year (category 3 aerosol types are predominant) except in January when nearly 35% percent of the records show category 5 (polluted marine) aerosols. Though Mauna Loa is oceanic, the site is at an elevation of 3397 m above mean sea level, which means measurements are above the marine boundary layer most of the time and explains why there is an insignificant frequency of marine-type aerosols most of the year except January.

[26] Figures 6f and 6g show the aerosol type fractions at An Myon (36N, 126E) and GSFC (39N, 76W), respectively. Both sites are category 4 (polluted continental) sites. At An Myon, there is an influx of desert dust between February and May corresponding to months when Asian dust episodes are normally observed. The figure shows the frequency of

aerosol types in records between September 1999 and March 2002. At both sites there is a peak in the fractional frequency of polluted continental type aerosols during the summer months. This is an expected trend since both sites are near industrial and urban sites. The An Myon site is also characterized by small fractions of biomass burning (possibly from southeast Asian sources) while the GSFC sites experiences periods of significant clean continental aerosol during the winter months from November to February. The GSFC data used for this classification are measurements between May 1993 and December 2002.

[27] Figure 6i shows the seasonal variation of aerosol types at Lanai (20N, 156W), an oceanic site, using data measured between November 1995 and December 2002. There is a substantial amount of dust (category 1) in March–May and biomass burning (category 2) in June–August. These events are most likely due to long-range transport of desert dust and smoke from biomass burning. Though polluted marine is dominant throughout the year, there are several records classified in the biomass burning

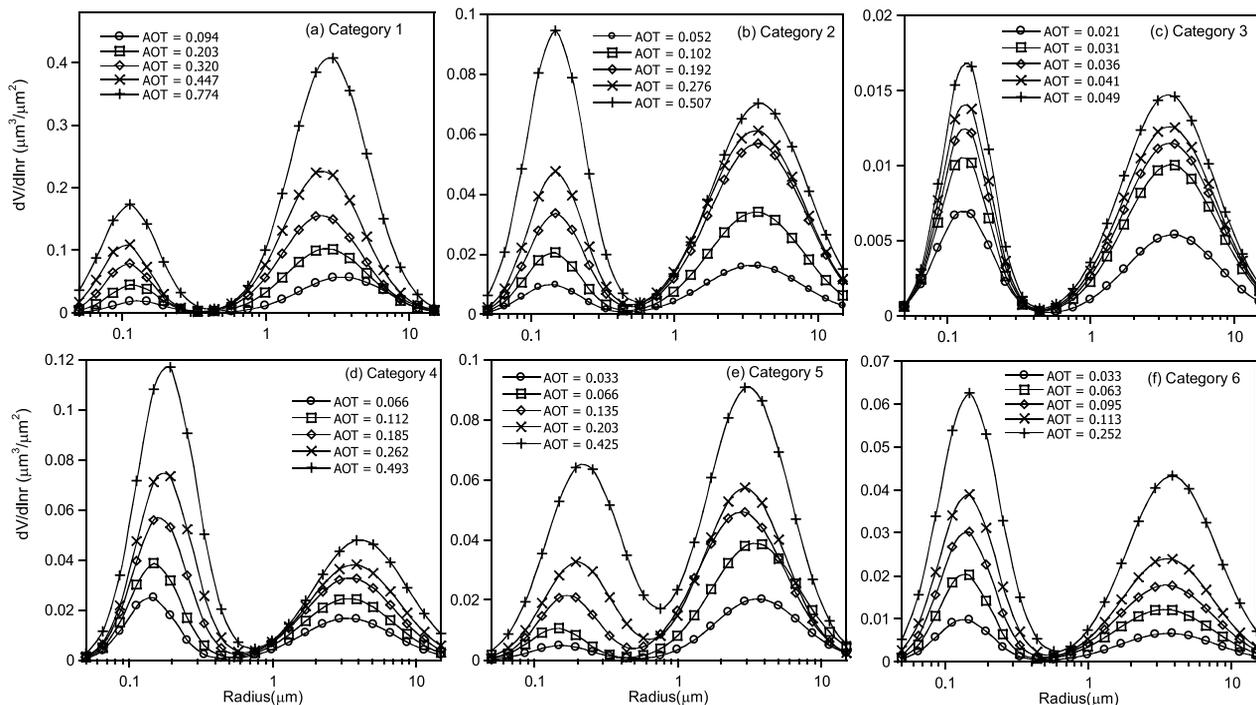


Figure 7. Size distributions of (a) dust, (b) biomass burning, (c) background/rural, (d) polluted continental, (e) marine, and (f) dirty pollution aerosol, i.e., categories 1, 2, 3, 4, 5, and 6, respectively. The data in each category were partitioned into five aerosol optical thickness (AOT) classes.

cluster throughout the year. The aerosol type at Lanai is therefore likely mixed with significant amounts of both organic and elemental carbon since the imaginary part of the refractive index, shown in Table 2, is quite elevated at 0.0044 compared to 0.001 for a pure marine aerosol measured at clean oceanic sites [Smirnov *et al.*, 2003].

[28] The aerosol type frequencies at Skukuza (24S, 31E) are shown in Figure 6i. Skukuza is also a southern African site in the proximity of biomass burning activities in May to October. The figure shows this trend in addition to a peak in the dirty pollution frequencies in January and February. Category 6 aerosols are characterized by thin layers (mean optical depth of 0.1) and a high absorption. The imaginary part of the refractive index of 0.033 resulting in a single scattering albedo of 0.72 at 673 nm for the center of the cluster shown in Table 2. Three of the sites belonging to this cluster (Skukuza, IMS Metu Erdeml, Dalanzadgad) also belong to the biomass burning cluster, suggesting that the smoke may have similar sources and the differences between the two clusters may be due to smoke aging. The plot shows averaged monthly frequencies of records in categories for measurements made between July 1998 and December 2002.

6. Consistency of Aerosol Microphysical Properties

[29] To check for the consistency of the individual categories, we divided the measurements in each category into five optical depth classes and plotted the size distributions of each class within a category. The results are shown in Figure 7. Note that for each category, the magnitude of the fine and coarse mode amplitudes of the size distribution

denotes fine and coarse loading, respectively. As shown, the mode amplitudes increase with the optical depth for all types. The optical depth is an extensive property, i.e., a property that depends on the amount of aerosol, as is the mode amplitude and this behavior is expected for a given aerosol type. The fine and coarse mean radii and geometric standard deviations, on the other hand, are intensive properties, i.e., properties that depend on the type of aerosol. These are relatively constant across optical depth classes within categories. This means that despite changes in the optical depth, the aerosol size distribution is consistently the same in each category. While this is not a sufficient validation of the clustering method, it is an indication of the ability of the algorithm to group similar data sets using the prescribed variables.

7. Discussion

[30] The AERONET classification yields six groups of aerosols shown in Table 3. The table also shows the likely constituent species of each type inferred from the cluster

Table 3. Likely Composition of Aerosol Types Determined by the Cluster Analysis

Cluster Category	Aerosol Type
1	desert dust (mineral dust)
2	biomass burning (soot + OC)
3	background/rural (SO_4^{2-} , NO_3^- , OC, NH_4^+)
4	polluted continental (SO_4^{2-} , NO_3^- , OC, NH_4^+ + soot)
5	polluted marine (NaCl + OC + soot)
6	dirty pollution (SO_4^{2-} , NO_3^- , OC, NH_4^+ + soot)

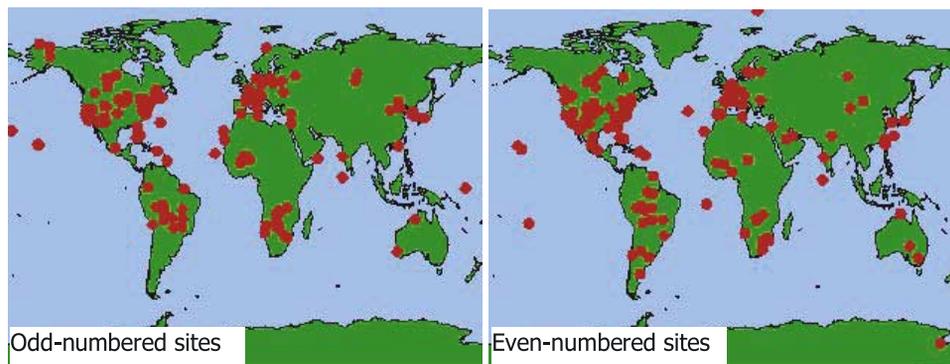


Figure 8. Locations of the odd- and even-numbered sites used for the error analysis.

center values of the imaginary refractive indices, single scattering albedos, and size distributions. Desert dust is assumed to be mostly mineral soil. Biomass burning is an aged smoke aerosol consisting primarily of soot and organic carbon (OC). Rural background, also referred to as clean continental aerosol is a lightly loaded aerosol consisting of sulfates (SO_4^{2-}), nitrates (NO_3^-), OC, and ammonium (NH_4^+). Polluted marine aerosol consists primarily of seasalt (NaCl) with traces of polluted continental species. Both polluted continental and dirty pollution consist of the same species (OC, soot, SO_4^{2-} , NH_4^+ , NO_3^-) but the large imaginary part of the refractive index of dirty pollution suggests that this type of aerosol contains a significantly larger (than polluted continental) fraction of soot.

[31] Since some of these records span a period of nearly a decade, the most conclusive validation of the clustering algorithm would be to examine a climatology of in situ measurements at the individual sites. While such in situ data may become available for more sites in future, there are not many in situ measurements of aerosol types at the AERONET sites. Even if such a climatology were available, it would only be useful in locations where a given type of aerosol (say polluted continental) was dominant almost to the exclusion of all other types. For sites that experience several different types of aerosols, aerosol optical properties based on time averages longer than a few hours would be misleading.

8. Error Analysis of the Clusters

[32] To check for the consistency of the clustering algorithm, we divided the entire data set from 283 stations into two nearly equal groups and repeated the clustering analysis. These sites were indexed alphabetically (by site name) with the first site (An Myon referred to as site 1) and the last site (Wits University referred to as site 283). We then divided the sites into odd-numbered (sites 1, 3, 5, ..., 283) and even-numbered sites (sites 2, 4, 6, ..., 282). Figure 8 shows locations of the odd and even-numbered sites. The distribution of the sites, shown in Figure 8, is such that there is little or no geographic bias in so far as both are nearly equally well distributed globally. Therefore applying the clustering algorithm to each data set individually should produce the nearly the same result as using the total data set. Partitioning the data into these two subsets resulted in 72,728 records from the odd-numbered stations and 71,185 records from the even-numbered stations. These two data sets were individually

clustered and compared to the cluster results obtained by using the entire data set of 143,913 records. The assessment of the clustering is deemed “correct” if the classification of a record using a subset of the data is the same as that obtained by clustering the entire data set. The results of the clustering analysis are shown in Table 4. Here the rows and columns denote the number of records of category x in the odd- (or even-) numbered subset that are classified as category y in the entire data set. The diagonal (boldface values) contains the number of records that are grouped in the same category when the subset of data (odd- or even-numbered sites) and the entire data set are independently analyzed. The last column (% correct) refers to the percentage of records in the subset of the data that is classified correctly. The robustness of the clustering algorithm is illustrated by the large fraction of records that are classified in the same category when three data sets are independently categorized.

[33] In the first row of Table 4, 10,069 (out of a total of 10,130 or more than 99%) records are classified as category 1 aerosols in both the subset and the entire data set. One record is classified as category 1 in the subset

Table 4. Results of Clustering Odd- and Even-Numbered Subsets of the AERONET Data Compared to Clustering the Entire Data Set^a

Category	1	2	3	4	5	6	Total	Percent Correct
<i>Odd-Numbered Sites</i>								
1	10,069	1	0	0	60	0	10,130	99.4
2	0	13,373	0	569	0	45	13,987	95.6
3	0	0	10,377	0	0	0	10,377	100.0
4	0	811	0	26,061	0	207	27,079	96.2
5	971	0	0	0	3969	0	4940	80.3
6	0	195	0	174	0	5846	6215	94.1
Total	11,040	14,380	10,377	26,804	4029	6098	72,728	
<i>Even-Numbered Sites</i>								
1	11,870	0	0	0	202	0	12,072	98.3
2	0	12,127	0	461	0	87	12,675	95.7
3	0	0	9930	0	0	0	9930	100.0
4	0	1420	0	27,163	0	5	28,588	95.0
5	13	0	0	0	1574	0	1587	99.2
6	0	2116	0	258	0	3959	6333	62.5
Total	11,883	15,663	9930	27,882	1776	4051	71,185	

^aValues in boldface are the number of records that are grouped in the same category when the subset of data (odd- or even-numbered sites) and the entire data set are independently analyzed.

and category 2 in the entire data set and 60 of these records are classified as category 1 in the subset and category 5 in the entire data. Note that dust and marine aerosol share the characteristics of a large coarse fraction and a small imaginary part of the refractive index and therefore a small percentage of the desert dust aerosol (6%) is likely to be misclassified as marine aerosol. In the second row, 13,373 (more than 95%) records of the odd-numbered sites data subset are correctly classified as category 2 aerosols. A small number of category 2 (biomass burning) records (569 ~4%) are mischaracterized as category 4 (polluted continental) because of the similarity (of refractive indices, size distributions) of the two aerosol types. The third row shows using the odd-numbered subset and the entire data set has the same result, i.e., all category 3 records are grouped in the same class. Category 3 (clean continental) records have a low optical depth and the retrieved microphysical parameters are quite uncertain. This is somewhat encouraging in that it shows the clustering algorithm is insensitive to uncertainties if these are associated with only a few parameters. More than 96% of the category 4 records are classified correctly. A few misclassifications of 3% and 1% are grouped in category 2 (biomass burning) and category 6 (dirty pollution) respectively, i.e., aerosols are misclassified into groups that have similar microphysical and optical characteristics. The next row shows that 3969 (~80%) of the category 5 (marine) records are classified correctly and the remaining 20% are misclassified as category 1 (desert dust). The dirty pollution records (category 6) are correctly classified in more than 94% of the records and the misclassified records are grouped in category 2 (biomass burning).

[34] The analysis of the even-numbered sites yielded results that are quite similar to the odd-numbered sites described above. Of the records the even-numbered sites subset of data, categories 1, 2, 3, 4, and 5 aerosols were classified correctly 98%, 96%, 100%, 95%, and 99%, respectively, of the records in the even-numbered sites subset of data. A significant percentage (37.5%) of the category 6 (dirty pollution) is misclassified as category 2 (biomass burning) aerosol. As in the odd sites subset, all the category 3 records in the even sites subset are correctly classified.

[35] The subdivision of data sets by the arbitrarily chosen odd/even site indices is akin to random sampling of the data to produce two unrelated data subsets and forms an objective test of the algorithm. Furthermore, in cases where data is miscategorized, the misclassified records are placed in groups that have similar microphysical and optical characteristics. The misclassified desert dust records are likely to be grouped with marine aerosols and vice versa, biomass burning aerosols are likely to be misclassified as continental pollution and vice versa; misclassified dirty pollution records likely to be categorized as biomass burning aerosols and vice versa. In summary, the clustering algorithm is quite robust. The algorithm reproduced the correct classification in 94.7% of all the records tested, i.e., the algorithm error rate is less than 6%.

9. Error Propagation

[36] A small subset of the data was used to test the effects of the uncertainty in the retrieved microphysical parameters

Table 5. Uncertainty Analysis of the Categorization Using a 10% Random Error Propagated Through the Microphysical Properties to the Optical Properties and Clustering Algorithm^a

Test	Category						Correct Classification, %
	1	2	3	4	5	6	
1	0.042	0.068	0.050	0.047	0.137	0.204	91.3
2	0.049	0.065	0.044	0.026	0.080	0.127	92.1
3	0.043	0.097	0.024	0.032	0.077	0.129	92.2
4	0.027	0.061	0.034	0.044	0.084	0.170	91.8
5	0.022	0.079	0.026	0.023	0.056	0.143	91.6
6	0.043	0.071	0.047	0.034	0.068	0.140	91.2
7	0.030	0.078	0.036	0.022	0.067	0.127	91.9
8	0.032	0.061	0.031	0.025	0.078	0.150	91.7
9	0.065	0.079	0.034	0.026	0.089	0.145	91.6
10	0.040	0.089	0.050	0.026	0.074	0.134	94.7
Average	0.039	0.075	0.038	0.031	0.081	0.147	92.0

^aThe table shows the normalized Euclidean distances between the free and perturbed cluster centers for each category found in 10 independent tests.

on the cluster results. We selected 5% of the total data set and applied the cluster analysis to partition the data into six categories. As shown above, clustering the subsets of the data yields classifications similar to those obtained using the whole data set. We then added a 10% normally distributed random error to the complex refractive index and size distribution parameters to obtain a new set of “noisy” data. These noisy records were used as inputs to the scattering calculations to yield perturbed optical parameters. The error-laden optical and microphysical parameters are then used as variables in the clustering algorithm. An objective measure of the robustness of the clustering algorithm is the percentage of records classified in the same category using the error-free and error-laden data. A quantitative measure of the effect of noise on the cluster analysis is the Euclidean distance, in 26-parameter space, between the centers of the new clusters (C_n , obtained using the perturbed data set) and old clusters (C_o , obtained using the error free data set) in each corresponding category. This distance Δd_e is defined by

$$\Delta d_e = \frac{|C_n - C_o|}{C_{std}}, \quad (3)$$

where we have normalized the distance by the standard deviation C_{std} of the distributions of the individual record distances from the unperturbed center.

[37] The test was repeated for ten realizations of the randomly generated optical and microphysical properties. The test checks whether say, a record classified as biomass burning in the error-free data set is classified as a biomass burning record in the error-laden data and by how much does the new center shift. Table 5 below shows the results of the error propagation tests. The shift in the centers of the clusters of the perturbed data set are presented as normalized Euclidean distances. The table shows that the largest shift in the centers occurs in category 6 (dirty pollution), an expected result since this category has a relatively small number of data points. The error propagation tests show that the error in the classification due to a 10% error in the retrieved parameters (size distributions and complex refractive indices) is less than 9%, i.e., given a 10% error in the

retrieved parameters, at least 91% of the records will be correctly classified.

10. Conclusion

[38] A global data set, AERONET, has been used to identify main clusters of aerosol types and to determine microphysical properties of aerosol groups. The clustering algorithm objectively groups all the 143,000+ records examined into six categories. Using the mean values of the optical and microphysical properties together with the geographic locations, we identified these categories as desert dust, biomass burning, urban industrial pollution, rural background, polluted marine, and dirty pollution and presented the mean properties of these aerosol models. The effects of aerosol particle nonsphericity on the cluster analysis are unknown but not negligible. We expect that these models will enhance the available database of the characteristics of aerosol types. Since the cluster analysis assigned a category to each record, it is possible to examine the frequency of occurrence of different types of aerosols at each station. The data showed periods of intense biomass burning activity and desert dust generation consistent with independent observations. The variation of the extensive and intensive size distribution properties within categories showed consistent trends. In particular, when each cluster was subdivided by optical depth class, the trends of the class size distributions show that the extensive properties (mode amplitude and total volume) vary by optical depth while the intensive properties (mean radius and standard deviation) are relatively constant. The uncertainty and sensitivity tests showed that the clustering algorithm is quite robust and reproduces more than 94% of the classification when the data is arbitrarily halved. When random errors of 10% are added to the microphysical properties and propagated through the optical properties and the clustering algorithm, the records are correctly classified at a rate of at least 91%.

[39] **Acknowledgments.** We would like to thank all the AERONET site principal investigators for the data used for this study. Jae-Gwang Won was supported in part by the SRC program of Korea Science and Engineering Foundation and the Brain Korea 21 program.

References

- Ansmann, A., D. Althausen, U. Wandinger, K. Franke, D. Müller, F. Wagner, and J. Heintzenberg (2000), Vertical profiling of the Indian aerosol plume with six-wavelength lidar during INDOEX: A first case study, *Geophys. Res. Lett.*, *27*(7), 963–966.
- Bates, T. S., B. J. Heuebert, J. L. Gras, F. B. Griffiths, and P. A. Durkee (1998), International Global Atmospheric Chemistry (IGAC) Project's First Aerosol Characterization Experiment (ACE 1): Overview, *J. Geophys. Res.*, *103*, 16,297–16,318.
- Chiapello, I., G. Bergametti, B. Chatenet, P. Bousquet, F. Dulac, and E. Soares (1997), Origins of African dust transported over the northeastern tropical Atlantic, *J. Geophys. Res.*, *102*(D12), 13,701–13,710.
- Colarco, P. R., O. B. Toon, and B. N. Holben (2003), Saharan dust transport to the Caribbean during PRIDE: 1. Influence of dust sources and removal mechanisms on the timing and magnitude of downwind aerosol optical depth events from simulations of in situ and remote sensing observations, *J. Geophys. Res.*, *108*(D19), 8589, doi:10.1029/2002JD002658.
- d'Almeida, G. A., P. Koepke, and E. P. Shettle (1991), *Atmospheric Aerosols: Global Climatology and Radiative Characteristics*, A. Deepak, Hampton, Va.
- Di Iorio, T., A. di Sarra, W. Junkermann, M. Cacciani, G. Fiocco, and D. Fua (2003), Tropospheric aerosols in the Mediterranean: 1. Microphysical and optical properties, *J. Geophys. Res.*, *108*(D10), 4316, doi:10.1029/2002JD002815.
- Dubovik, O., and M. D. King (2000), A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements, *J. Geophys. Res.*, *105*, 20,673–20,696.
- Dubovik, O., A. Smirnov, B. N. Holben, M. D. King, Y. J. Kaufman, T. F. Eck, and I. Slutsker (2000), Accuracy assessments of aerosol optical properties retrieved from Aerosol Robotic Network (AERONET) Sun and sky radiance measurements, *J. Geophys. Res.*, *105*, 9791–9806.
- Dubovik, O., B. N. Holben, T. F. Eck, A. Smirnov, Y. J. Kaufman, M. D. King, D. Tanre, and I. Slutsker (2002a), Variability of absorption and optical properties of key aerosol types observed in worldwide locations, *J. Atmos. Sci.*, *59*, 590–608.
- Dubovik, O., B. N. Holben, T. Lapyonok, A. Sinyuk, M. I. Mishchenko, P. Yang, and I. Slutsker (2002b), Non-spherical aerosol retrieval method employing light scattering by spheroids, *Geophys. Res. Lett.*, *29*(10), 1415, doi:10.1029/2001GL014506.
- Holben, B. N., et al. (1998), AERONET—A federated instrument network and data archive for aerosol characterization, *Remote Sens. Environ.*, *66*, 1–16.
- Holben, B. N., et al. (2001), An emerging ground-based aerosol climatology: Aerosol optical depth from AERONET, *J. Geophys. Res.*, *106*, 12,067–12,098.
- Intergovernmental Panel on Climate Change (IPCC) (2001), *Climate Change 2001: The Scientific Basis, Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change*, 944 pp., Cambridge Univ. Press, New York.
- Kaufman, L., and P. J. Rousseeuw (1990) *Finding Groups in Data*, John Wiley, Hoboken, N. J.
- Kaufman, Y. J., et al. (1998), Smoke, Clouds, and Radiation-Brazil (SCAR-B) experiment, *J. Geophys. Res.*, *103*(D24), 31,783–31,808.
- Kiehl, J. T., and B. P. Briegleb (1993), The relative roles of sulfate aerosols and greenhouse gases in climate forcing, *Science*, *260*, 311–314.
- Kiehl, J. T., and H. Rodhe (1995), Modeling geographical and seasonal forcing due to aerosols, in *Dahlem Workshop on Aerosol Forcing of Climate*, edited by R. J. Charlson and J. Heintzenberg, John Wiley, Hoboken, N. J.
- Kubilay, N., T. Cokacar, and T. Oguz (2003), Optical properties of mineral dust outbreaks over the northeastern Mediterranean, *J. Geophys. Res.*, *108*(D21), 4666, doi:10.1029/2003JD003798.
- Mie, G. (1908), Beigrade zur optik trüber medien, speziell kolloidaler metallösungen, *Ann. Phys.*, *25*(4), 377–445.
- Mishchenko, M. I., A. A. Lacis, B. E. Carlson, and L. D. Travis (1995), Nonsphericity of dust-like tropospheric aerosols: Implications for aerosol remote sensing and climate modeling, *Geophys. Res. Lett.*, *22*(9), 1077–1080.
- Müller, D., F. Wagner, D. Althausen, U. Wandinger, and A. Ansmann (2000), Physical properties of the Indian aerosol plume derived from six-wavelength lidar observations on 25 March 1999 of the Indian Ocean Experiment, *Geophys. Res. Lett.*, *27*(9), 1403–1406.
- National Research Council (NRC) (1996), A plan for a research program on aerosol radiative forcing and climate change, Natl. Acad. of Sci., Washington, D. C.
- Nemesure, S., R. Wagener, and S. E. Schwartz (1995), Direct shortwave forcing of climate by anthropogenic sulfate: Sensitivity to particle size, composition, and relative humidity, *J. Geophys. Res.*, *100*, 26,105–26,116.
- Penner, J. E., R. E. Dickinson, and C. A. O'Neill (1992), Effects of aerosol from biomass burning on the global radiation budget, *Science*, *256*, 1432–1434.
- Raes, F., T. Bates, F. McGovern, and M. V. Liedekerke (1999), The 2nd Aerosol Characterization Experiment (ACE-2), General overview and main results, *Tellus, Ser. B*, *52*(2), 111–125.
- Ramanathan, V., et al. (2001), Indian Ocean Experiment: An integrated analysis of the climate forcing and effects of the great Indo-Asian haze, *J. Geophys. Res.*, *106*(22), 28,371–28,398.
- Reid, J. S., P. V. Hobbs, R. J. Ferek, D. R. Blake, J. V. Martins, M. R. Dunlap, and C. Liousse (1998), Physical, chemical, and optical properties of regional hazes dominated by smoke in Brazil, *J. Geophys. Res.*, *103*(D24), 32,059–32,080.
- Russell, P. B., P. V. Hobbs, and L. L. Stowe (1999), Aerosol properties and radiative effects in the United States East Coast haze plume: An overview of the Tropospheric Aerosol Radiative Forcing Observational Experiment (TARFOX), *J. Geophys. Res.*, *104*, 2213–2222.
- Sheridan, P. J., D. J. Delene, and J. A. Ogren (2001), Four years of continuous surface aerosol measurements from the Department of Energy's Atmospheric Radiation Measurement Program Southern Great Plains Cloud and Radiation Testbed site, *J. Geophys. Res.*, *106*(D18), 20,735–20,748.
- Smirnov, A., B. N. Holben, O. Dubovik, R. Frouin, T. F. Eck, and I. Slutsker (2003), Maritime component in aerosol optical models derived from Aero-

- sol Robotic Network data, *J. Geophys. Res.*, *108*(D1), 4033, doi:10.1029/2002JD002701.
- Sokolik, I. N., and O. B. Toon (1999), Incorporation of mineralogical composition into models of the radiative properties of mineral aerosol from UV to IR wavelengths, *J. Geophys. Res.*, *104*(D8), 9423–9444.
- Sokolik, I. N., D. M. Winker, G. Bergametti, D. A. Gillette, G. Carmichael, Y. J. Kaufman, L. Gomes, L. Schuetz, and J. E. Penner (2001), Introduction to special section: Outstanding problems in quantifying the radiative impacts of mineral dust, *J. Geophys. Res.*, *106*(D16), 18,015–18,028.
- Tanré, D., L. A. Remer, Y. J. Kaufman, S. Mattoo, P. V. Hobbs, J. M. Livingstone, P. B. Russell, and A. Smirnov (1999), Retrieval of aerosol optical thickness and size distribution over ocean from the MODIS airborne simulator during TARFOX, *J. Geophys. Res.*, *104*(D2), 2261–2278.
- Tanré, D., Y. J. Kaufmann, B. N. Holben, B. Chatenet, A. Karnieli, F. Lavenu, L. Blarel, L. A. Remer, and A. Smirnov (2001), Climatology of dust aerosol size distribution and optical properties derived from remotely sensed data in the solar spectrum, *J. Geophys. Res.*, *106*(D16), 18,205–18,218.
- Tanré, D., J. Haywood, J. Pelon, J. F. Léon, B. Chatenet, P. Formenti, P. Francis, P. Goloub, E. J. Highwood, and G. Myhre (2003), Measurement and modeling of the Saharan dust radiative impact: Overview of the Saharan Dust Experiment (SHADE), *J. Geophys. Res.*, *108*(D18), 8574, doi:10.1029/2002JD003273.
- Yang, P., and K. N. Liou (1996), Geometric-optics-integral-equation method for light scattering by nonspherical ice crystals, *App. Opt.*, *35*, 6568–6584.
-
- O. Dubovik, Goddard Earth Sciences and Technology Center, University of Maryland, Baltimore County, Baltimore, MD 21228, USA.
- M. P. McCormick, Center for Atmospheric Sciences, Hampton University, Hampton, VA 23668, USA.
- A. H. Omar and D. M. Winker, NASA Langley Research Center, Atmospheric Sciences, MS 401 A, Hampton, VA 23681, USA. (ali.h.omar@nasa.gov)
- J.-G. Won and S.-C. Yoon, School of Earth and Environmental Sciences, Seoul National University, Seoul 151-747, Korea.