Atmospheric Environment 140 (2016) 213–233

Contents lists available at ScienceDirect

Atmospheric Environment

journal homepage: www.elsevier.com/locate/atmosenv

An AERONET-based aerosol classification using the Mahalanobis distance

Patrick Hamill [a, *], Marco Giordano [b], Carolyne Ward [c], David Giles [d], Brent Holben [e]

[a] San Jose State University, San Jose, California and NASA Ames Research Center, Moffett Field, CA, USA
[b] University of Nevada and Desert Research Institute, Reno, NV, USA
[c] California State University, Long Beach, CA, USA
[d] Science Systems and Applications, USA and NASA Goddard Space Flight Center, Lanham, MD, Greenbelt, MD, USA
[e] NASA Goddard Space Flight Center, Greenbelt, MD, USA

HIGHLIGHTS

- Aerosols characterized as Urban-Industrial, Biomass, Dust, Mixed and Maritime.
- Aerosol typing using AERONET retrieved parameters.
- Seasonal variation in aerosol type at AERONET sites.
- Use of Mahalanobis distance to identify type of individual AERONET measurements.

ARTICLE INFO

Article history:
Received 18 December 2015
Received in revised form
30 May 2016
Accepted 3 June 2016
Available online 4 June 2016

Keywords:
Atmospheric aerosols
Aerosol typing
AERONET
Mahalanobis distance
Seasonal aerosol variation
High AOD events

ABSTRACT

We present an aerosol classification based on AERONET aerosol data from 1993 to 2012. We used the AERONET Level 2.0 almucantar aerosol retrieval products to define several reference aerosol clusters which are characteristic of the following general aerosol types: Urban-Industrial, Biomass Burning, Mixed Aerosol, Dust, and Maritime. The classification of a particular aerosol observation as one of these aerosol types is determined by its five-dimensional Mahalanobis distance to each reference cluster. We have calculated the fractional aerosol type distribution at 190 AERONET sites, as well as the monthly variation in aerosol type at those locations. The results are presented on a global map and individually in the supplementary material. Our aerosol typing is based on recognizing that different geographic regions exhibit characteristic aerosol types. To generate reference clusters we only keep data points that lie within a Mahalanobis distance of 2 from the centroid. Our aerosol characterization is based on the AERONET retrieved quantities, therefore it does not include low optical depth values. The analysis is based on "point sources" (the AERONET sites) rather than globally distributed values. The classifications obtained will be useful in interpreting aerosol retrievals from satellite borne instruments.

1. Introduction

Studies of the Earth's climate and its temporal variations require having a detailed knowledge of gases and aerosol particles in the atmosphere, including their optical properties, concentrations and distributions. Aerosols are short lived, highly heterogeneous, and difficult to characterize on a global or regional scale. It is well known that the effect of aerosols is one of the greatest uncertainties in carrying out climate studies. For example the IPCC Assessment (IPCC AR5 (2013)) characterize aerosol-cloud interaction as having a low confidence level. The aerosol radiation interaction confidence level is characterized as high to medium, but both aerosol effects have large uncertainty ranges. Since in situ aerosol measurements are limited in time and space, a great deal of effort has gone into measuring aerosols from orbiting spacecraft. Space-borne instruments that give information on aerosols include CALIPSO (Omar et al. (2009)), MODIS (Remer et al. (2008)), MISR (Levy et al. (2007)), and POLDER (Kacenelenbogen et al. (2006)). These systems yield extensive information about the geographic and temporal aerosol distribution (usually in terms of optical depth). Some satellite systems also measure or derive various optical parameters

* Corresponding author.
E-mail address: patrick.hamill@sjsu.edu (P. Hamill).

http://dx.doi.org/10.1016/j.atmosenv.2016.06.002
1352-2310/

such as polarization or backscatter ratio that might be useful in determining aerosol type. But the determination of aerosol type by remote sensing is a difficult problem. One of the purposes of this paper is to show that the aerosol type can be reasonably determined using a limited amount of information on the aerosol's optical properties. We based our analysis on data from the ground-based network of sun photometers called AERONET (for AErosol RObotic NETwork) Holben et al. (1998). Previous studies which also used the AERONET optical properties to identify aerosol types generally concentrated on the single scattering albedo (Leahy et al. (2007), Johnson et al. (2009), Muller et al. (2010), Toledano et al. (2011)) and/or the extinction and absorption angstrom exponents (Bergstrom et al. (2007), Russell et al. (2010)).

In this paper we used five retrieved AERONET optical properties and a Mahalanobis distance analysis to identify aerosol types and their seasonal variation for nearly 200 AERONET sites. We also present regional overviews of aerosol types by continent and globally. A modified version of the methodology used here was applied by Russell et al. (2014) who focussed primarily on the classification of aerosol types using the Mahalanobis distances (as also done in this study) and the Wilks' lambda distribution (not done in this study) and compared the aerosol types thus obtained with those determined by POLDER. Our methodology is similar but we use a different set of reference clusters and apply it to a much larger population of the AERONET data set to determine (a) the distribution of aerosol types at each location, (b) the seasonal dependence of aerosol type, and (c) the world-wide distribution of aerosol types. We also consider problems associated with this technique, as in the difficulty in identifying aerosol type in Bonanza Creek (Alaska), Mexico City, and some South-East Asian cities. Furthermore, we incorporated maritime aerosol in our analysis by using the relative humidity (obtained from meteorological data) and calculating the expected optical properties of salt solution droplets. Our results quantify the relative percentage of aerosol types at AERONET sites on a climatological basis, which can potentially be used as initial estimates to compare with or be used by satellite retrievals and modeling platforms.

Particular mention should be made of three studies that generated global aerosol climatologies, namely, the papers of Kahn et al. (2001), Omar et al. (2005) and Taylor et al. (2015). These papers differ from the present study in that all of them used k-means clustering to determine the appropriate aerosol clusters for their climatology. Kahn et al. (2001) used six models that evaluated the transport of one or more of the four components: (1) sulfate (2) carbonaceous (3) dust and (4) sea salt. Kahn et al. (2001) used the output of these models to generate a global grid of monthly estimates of optical depth and the fraction of each component at each point of the grid. They found that it was possible to represent the aerosol using 13 different fractional combinations of the four components. They used a k-means clustering algorithm to locate the centroids of the clusters of the 13 "air mass types." They then generated global maps for January, April, July and October, showing the global aerosol distribution of the 13 representative air mass types.

Omar et al. (2005) used the AERONET data set (level 1) up to the end of 2002, and carried out a k-means cluster analysis. This analysis led them to conclude that the AERONET aerosol products formed 6 distinct clusters that they denoted as: (1) dust (2) biomass burning (3) rural or background (4) industrial pollution (5) polluted marine and (6) dirty pollution. They discussed the properties of each category and described the location and seasonal dependence of the various clusters. For example, the cluster denoted "dirty pollution" has optical properties that suggest that this cluster likely represents "flaming" combustion, while the "biomass burning" cluster has higher single scattering albedo and real refractive index and lower imaginary refractive index and likely represents "smoldering" combustion. They show on a global map the locations of the AERONET sites that are most representative of the each of the 6 categories.

Taylor et al. (2015) in an interesting and important paper used the aerosol optical depths generated by the GOCART model over a period of 7 years. They used a k-means clustering algorithm and found that the optimal number of clusters was 10. They assumed each cluster was composed of various proportions of the following aerosol types: (1) biomass burning (2) sulfate (3) dust (4) marine. For example, Cluster 1, denoted "Sulfurous dusty SMOKE" is composed of 27% sulfate, 31% dust and 36% biomass burning (i.e., smoke). Taylor et al. (2015) then plotted multiyear maps showing the geographic and seasonal distribution of the various clusters. Given these maps, they then identified AERONET sites that are located in a region dominated by a particular cluster. This allowed them to determine the optical properties associated with each cluster. Thus, for example, Cluster 1 at 440 nm has single scattering albedo of 0.93, real refractive index of 1.449 and imaginary refractive index of 0.01.

Note that the papers by Kahn et al., Omar et al., and Taylor et al. all used k-means clustering to define the clusters they used to generate their climatologies. In our study we defined 5 clusters in a significantly different way as described in Section 3.1.1 below. In a general sense, our clusters are defined geographically rather than by composition or as a consequence of clustering analysis.

In Section 2 we briefly describe our methodology. In Section 3 we consider the reference clusters used in our study. Section 4 gives an overview of our results for a number of specific AERONET sites, presenting pie charts and month by month histograms of the aerosol types at various locations as determined by our technique (The plots presented for selected sites in Section 4 are given for all 190 sites in the supplementary material). Section 5 presents an AERONET based aerosol climatology on a continental and global basis.

2. Methodology

2.1. The AERONET network

The data generated by the AERONET network of sun photometers span several decades in time and are spread out geographically around the Earth. Measurements are performed by CIMEL spectral radiometers that measure both the solar and sky radiances. The web page for this federated network lists over 650 globally distributed instrumented sites for which level 2 data can be downloaded. The AERONET instruments perform measurements several times a day in mostly cloud-free conditions. The basic physical quantities obtained by the CIMEL instruments are aerosol optical depth (AOD) and sky radiances in several wavelength bands between 340 and 1640 nm (Holben et al. (1998)). The spectral AOD and almucantar sky radiances are inverted to determine additional aerosol properties including volume distribution, complex index of refraction, and single scattering albedo at the four nominal wavelengths of 440, 675, 870 and 1020 nm. AERONET utilizes the inversion algorithm developed by Dubovik and King (2000) that was expanded upon by Dubovik et al. (2000, 2002, 2006). The algorithm fits the measured radiances at the four wavelengths to a radiative transfer model. The aerosol properties are derived with a minimum of assumptions.

This long-term database of high quality optical, microphysical and radiative properties of the aerosol column at numerous sites around the Earth is a particularly attractive factor in performing aerosol classification analysis. In this study the AERONET version 2 Level 2.0 AOD and almucantar retrieval data were obtained from

the AERONET web site (http://aeronet.gsfc.nasa.gov). These data sets are cloud screened and quality assured (Smirnov et al. (2000), Holben et al. (2006)). The uncertainty in AOD varies from s to ±0.02, increasing with wavelength (Holben et al. (1998)).

2.2. AERONET derived aerosol properties used in this study

We now define the aerosol optical properties obtained from the AERONET data sets that were used in this study.

The Extinction Angstrom Exponent (EAE) is the slope of the extinction optical thickness curve as a function of wavelength. (Extinction is defined as absorption plus scattering). EAE is calculated using the aerosol extinction optical thickness (EOT) at 870 and 440 nm, thus:

$$EAE = -\frac{\log(EOT(870)) - \log(EOT(440))}{\log(870) - \log(440)}.$$

The Absorption Angstrom Exponent (AAE) is the slope of the aerosol absorption optical thickness curve as a function of wavelength. It is calculated using the aerosol absorption optical thickness (AAOT) at 870 and 440 nm, thus:

$$AAE = -\frac{\log(AAOT(870)) - \log(AAOT(440))}{\log(870) - \log(440)}.$$

Single Scattering Albedo (SSA) is the ratio of scattering to total extinction:

$$SSA = \frac{\text{Scattering}}{\text{Scattering} + \text{Absorption}}.$$

The uncertainty in SSA is ±0.03 (Dubovik et al. (2002)). A low value of SSA implies the aerosol is highly absorbing. The AERONET inversions yield values for SSA at the four standard wavelengths: 440, 675, 870 and 1020 nm. Values of the 440 nm SSA obtained for optical depths less than 0.4 are not routinely included in the Level 2 (high quality) AERONET data set. In this study we used AERONET values for SSA at 440 nm.

Index of Refraction. The real index of refraction (RRI) and the imaginary index of refraction (IRI) are also AERONET derived quantities; these are given at the four standard wavelengths mentioned above. In this study we routinely used values for RRI and IRI at 440 nm.

As noted by Schuster et al. (2005) the real part of the index of refraction (RRI) decreases as the water content of the aerosol increases. The imaginary part (IRI) is independent of the real part and increases with the absorptivity of the aerosol.

The uncertainty in real index of refraction is 0.04 and the imaginary index of refraction uncertainty is 30–50%. (Dubovik et al. (2000, 2002)). These indices as well as the single scattering albedo are not differentiated according to mode (coarse mode and fine mode), however, studies by Leahy et al. (2007), Johnson et al. (2009), Muller et al. (2010) and Toledano et al. (2011) all show good agreement (within instrument uncertainty) between in situ SSA measurements and AERONET retrieved values.

Other Parameters. The five parameters described above (EAE, AAE, SSA, RRI, IRI) comprise the fundamental set of aerosol properties that we use in our analysis. However, we have also considered using the sphericity and the ratio of fine mode volume density to total (fine plus coarse) volume density where the volume concentration ($\mu m^3/\mu m^2$) is defined by

$$C_V = \int_{r_{min}}^{r_{max}} \frac{dV(r)}{d\ln r} d(\ln r).$$

Taylor et al. (2015) point out that fine mode fraction and percent sphericity can be used to determine aerosol type. The AERONET inversions generate a value for the percent sphericity of the particles (Dubovik et al. (2006)) and this can be used to distinguish between dust (sand) particles that are expected to have a very low sphericity, and, for example, maritime particles that are expected to be highly spherical. We found that this parameter is not always helpful because the AERONET V2 archive does not include values for the sphericity if the optical depth (τ_{440}) is less than 0.2. It might be noted that the sphericity parameter is generally only valid for EAE <1 because as shown by Giles et al. (2011) there is a high variability in the parameter for EAE >1. This is not a problem, because both dust and maritime aerosols tend to have EAE <1. However, if we were to use the sphericity as a standard parameter in our analysis, measurements for which sphericity is not calculated would be unusable and we would lose many values at interesting sites. Consequently, we normally carry out our calculations using the five fundamental parameters mentioned above, but in checking for maritime aerosol we require that if the sphericity parameter is given, it must be greater than 80%. We decided not to use the volume ratio for reasons presented in Section 2.4.

2.3. The Mahalanobis distance

The technique we use to determine the aerosol type that best fits a particular measurement is based on a statistical quantity developed by Mahalanobis (1936). The dimensionless "Mahalanobis distance" is a measure of how far a particular measurement is from the centroid of a reference cluster. A particularly useful property of the Mahalanobis distance (MD) is that there is no limit to the number of variables that can be used to evaluate it. In the analysis presented here we used the five variables EAE, AAE, SSA, RRI and IRI to evaluate the MD. We experimented with using more variables, including sphericity and volume ratio as well as indices of refraction and single scattering albedo at different wavelengths. (We normally obtain 5-dimensional Mahalanobis distances, but if we use, say, 7 parameters, then the comparison has to be made with reference clusters containing 7 parameters and the Mahalanobis distance is 7-dimensional). In general, using more parameters meant that the number of usable data sets decreased, due to missing data points, and did not have significant effects on the aerosol types (with the notable exception of volume ratio and the "mixed aerosol" category).

The Mahalanobis distance is calculated as follows. Let $\mathbf{x} = (x_1, x_2, \cdots, x_N)^{\mathbf{T}}$ represent an N-dimensional vector whose components are the values of N parameters of a "test point" $\mathbf{x}$. Consider a cluster of values with means given by the vector $\mathbf{m} = (m_1, m_2, \cdots, m_N)^{\mathbf{T}}$.

Then the Mahalanobis distance from the test point to the cluster is defined by

$$D_M = \left[(\mathbf{x} - \mathbf{m})^{\mathbf{T}} \mathbf{S}^{-1} (\mathbf{x} - \mathbf{m})\right]^{1/2},$$

where $\mathbf{S} = \mathbf{cov}(\mathbf{x}_i, \mathbf{x}_j)$ is the covariance matrix whose elements are defined by

$$\mathbf{S} = E[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})]^{\mathbf{T}}.$$

Here E is the "expectation" which in our case is just the mean value. The covariance matrix allows one to account for the diffusiveness of a cluster (related to the standard deviations that are the diagonal elements of $\mathbf{S}$) and cross-correlations among different dimensions of a cluster (the off-diagonal elements of $\mathbf{S}$).

2.4. Sensitivity analysis

We carried out a number of tests to determine the effect of using different parameters in our Mahalanobis calculations. For example, instead of using the 440 nm values for SSA, RRI and IRI we used the 675 nm values. This caused very little change in the aerosol identification (changes were generally less than 2%). We associate the small changes with the fact that data points that are near the "border" between two clusters may get reclassified when the clusters are redefined. That is, it is reasonable to expect that changing the reference clusters will have an effect on data points that lie near the boundary of the region of influence of one cluster and another. In another test, we replace the imaginary index of refraction with the sphericity. Once again, the change was minimal, however the AERONET data does not always retrieve the sphericity, so the number of usable data points was smaller. However when we used the volume ratio as a parameter, there was a significant effect on the "mixed" aerosol classification, and for many sites, the dominant aerosol type was changed from mixed to urban industrial. Recall that the AERONET retrieval assumes a bi-normal size distribution and uses the same index of refraction for both modes. Our results suggest that this may not be appropriate for mixed aerosols. As a consequence of our sensitivity tests we decided to use EAE, AAE, SSA, RRI and IRI as our standard set of parameters.

As described below, we use a large number of appropriate sites to define each reference cluster. After doing this, we discard all points that lie further than a Mahalanobis distance of 2 from the centroid. Consequently, the reference clusters have very little dependence on any particular site.

3. Reference aerosol clusters

3.1. Various reference Aerosol types

As mentioned, our technique involves determining the Mahalanobis distance of a point from the centroid of a reference aerosol cluster. Each reference cluster should be characteristic of a particular type of aerosol. To obtain a set of reference clusters, we associated a specific type of aerosol with a given location (as was done by Cattrall et al. (2005)). Thus, one expects urban industrial aerosols in locations such as Washington, D.C., Hamburg, Lille, Moscow, etc. Similarly, aerosols from biomass burning are expected at sites in the Amazon and Southern Africa. Dust is prevalent in Saudi Arabia, and maritime aerosols are found on islands and coastal locations. However, the aerosols of many cities in Asia do not fit into any of these categories. Cattrall et al. defined a category called "South-East Asian Aerosol," but we found that this type of aerosol, although prevalent in Asia, is also present at other locations. Consequently, we classified it as "Mixed Aerosol," following the nomenclature of Giles et al. (2012). In this work we used five reference clusters denoted Urban Industrial, Biomass Burning, Mixed Aerosol, Dust, and Maritime.

It should be noted that these reference clusters are based on the locations where they are found, and not on their composition. It is perhaps more logical to generate a set of clusters based on composition, such as sulfate, nitrate, black carbon, brown carbon, sea salts and mineral dusts such as Montmorillonite. In fact, this is to some degree the approach of Taylor et al. (2015) who assumed aerosols were different combinations of four substances, biomass burning, sulfate, dust and sea salt, based on the GOCART transport model. However, we have chosen to base our study on the AERONET retrievals that make no *a priori* assumption on the composition of the particles.

There have been numerous attempts to classify aerosols, each using a different definition of aerosol types. For example, the CALIPSO analysis assumes six aerosol types, denoted clean continental, clean marine, dust, polluted continental, polluted dust, and smoke (Omar et al., 2013). The MODIS system (version C5) uses five aerosol models, which are described in detail by Levy et al. (2007), Mielonen et al. (2010), and in the MODIS ATDB by Remer et al. (2004). The models are: continental, dust, non-absorbing, neutral and absorbing. They were modified from the results of Dubovik et al. (2002). Curier et al. (2008) indicate that OMI uses 24 different aerosol models of which 10 are classified as "weakly absorbing," 9 are carbonaceous aerosol models, 4 are dust models and 1 is a volcanic ash model. These are differentiated from one another by the values of the geometric mean radii, the standard deviations of the two modes, the number fraction of the second mode, and the indices of refraction. In their analysis of AERONET aerosol measurements, J. Lee et al. (2010) categorized aerosols as dust, black carbon, non-absorbing anthropogenic aerosol (sulfates in the fine mode and sea salt in coarse mode) and mixed aerosol (defined as aerosols whose fine mode fraction is between 0.4 and 0.6). Russell et al. (2010), Giles et al. (2011, 2012) used AAE and EAE to determine the dominant absorbing aerosol type in the aerosol mixture. Giles et al. (2012) expanded upon Russell et al. (2010) and J. Lee et al. (2010) by using AERONET Version 2, Level 2.0 data to classify dominant aerosol types (i.e., urban industrial, biomass burning, dust and mixed) based on typical aerosol source regions using SSA and AAE as a function of EAE and fine mode fraction of the AOD. Giles et al. (2012) showed that SSA and EAE provide the most distinct reference aerosol type clusters of the 2-D combinations analyzed using Voronoi clustering. They presented scatter plots of absorption angstrom exponent vs. extinction angstrom exponent and of absorption angstrom exponent vs. fine mode aerosol fraction, showing the distinctions between "Mostly Dust", "Mostly B.C." and "Mixed B.C. and Dust." Burton et al. (2012) used values obtained with an airborne high spectral resolution lidar to define eight aerosol types. The AeroCom model intercomparison (Kinne et al. (2006)) used five aerosol types denoted sulfate, organic carbon, black carbon, mineral dust and sea-salt.

Thus, each attempt to identify aerosol type tends to use a somewhat different way of defining the aerosol type. The choice of parameters, as is only natural, is based on the measurement characteristics of the particular instrument under consideration. Nevertheless, we can appreciate that our classifications are reasonably in line with what has been chosen by other investigators.

3.1.1. Choice of reference aerosol clusters

As mentioned, we use the Mahalanobis distance from a specific AERONET measurement to a set of reference aerosol clusters and classify the aerosol according to its "distance" from the reference clusters. The technique yields a quantifiable metric that indicates how "close" we are to a particular aerosol type. The procedure depends on having a set of reference clusters that are characteristic of a particular aerosol type.

In this section we define a set of reference clusters using the five aerosol types mentioned above. However, we will also briefly mention alternative reference clusters in which some of the categories were split into two. For example, we considered including two types of biomass burning aerosols, one based on sites that primarily had flaming fires and the other on sites that primarily had smoldering fires (e.g., Giles et al. (2012)). Similarly, we considered the differences between Arabian dust and African dust and between Urban Industrial aerosols in North America and Europe. (In the analysis presented here we did not include results obtained with the alternative reference clusters because we feel that the statistical tools we are using do not differentiate in a meaningful way between, say, desert dust from Arabia and desert dust from

Africa).

The basis for our aerosol classification is the fact that aerosols in different regions tend to have different optical properties. However, we should keep in mind that different kinds of aerosols are actually differentiated by their microphysical and optical properties. As stated by Twomey (1977), "To describe fully even a single particle … would be difficult: one would need to specify geometry, dimensions, composition and spatial distribution of several components … It would be a totally impossible task to categorize fully these details for a population of aerosol particles …" Although we agree with Twomey's assessment, we believe that it is possible to make a reasonable identification of aerosol types using a small set of optical parameters, such as those generated by AERONET.

As an illustration of our procedure, let us consider just two of these optical properties, namely the Single Scattering Albedo (at 440 nm) and the Extinction Angstrom Exponent (870/440 nm/nm). Fig. 1 is a scatter plot of the values of SSA vs. EAE for the following locations: Goddard Space Flight Center (GSFC) near Washington D.C., Mongu (Zambia), Solar Village (Saudi Arabia), and Bac Giang, a city in northern Vietnam. Note that the points are individual AERONET retrieved values of SSA and EAE .

These four sites were selected because they are in geographic locations where one might expect certain characteristic aerosols. The separation of the clusters in Fig. 1 leads to the reasonable assumption that the different clusters represent different kinds of aerosols. The black dots from GSFC presumably represent urban industrial pollution, the Mongu aerosols are almost certainly the products of biomass burning, and the Solar Village aerosol is primarily dust. Bac Giang (as well as many other sites in Asia) present a type of aerosol that is different from the others, as can be appreciated from the figure. It has been suggested that this is "polluted dust" (a CALIPSO aerosol category, Omar et al. (2013)) but the values of SSA extend to lower values than either dust (orange) or urban pollution (black). Cattrall et al. (2005) classified such aerosols as "S.E. Asia", probably because this type of aerosol is commonly found in Asia. Nevertheless, it is also found in other regions, so the term "Asian Aerosol" is not appropriate. We use the nomenclature of Giles et al. (2012) and refer to this aerosol type as "mixed." Leaving aside the question of the actual composition of the mixed aerosol, we note that it has optical properties that differ from those of the other clusters and deserves to be categorized separately. (This difference is also observed in plots of EAE vs. RRI, EAE vs. IRI and EAE vs. AAE).

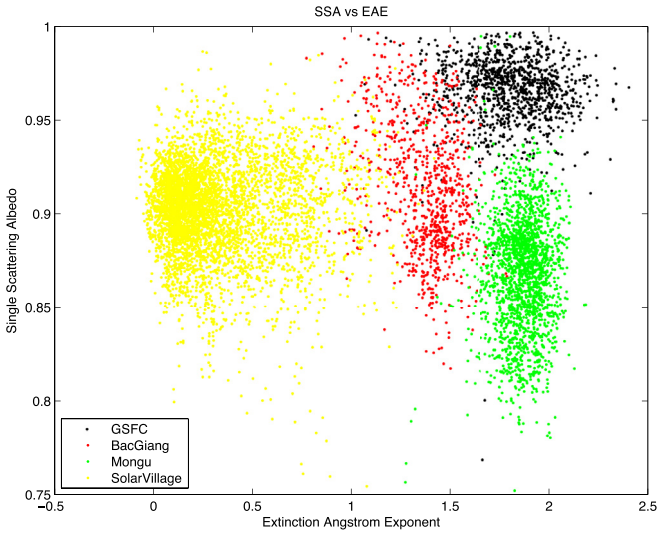

Fig. 1. Scatter plots of Single Scattering Albedo at 440 nm vs. Extinction Angstrom Exponent (870/440) for the following sites: Goddard Space Flight Center (black dots), Mongu (green dots), Solar Village (orange dots) and Bac Giang (dak red dots). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Although it is apparent from Fig. 1 that plots of SSA vs. EAE place the aerosols from different sites into reasonably separated clusters, there is, nevertheless, considerable overlap between the green dots (Biomass Burning) and the red dots (Mixed). Also note that the clusters have a significant number of "outliers", that is, points from one site that appear in one of the other clusters.

Among the considerations to be kept in mind when selecting a set of reference clusters, are the following: First of all, it is important to have clusters that do not overlap. Secondly, it is not reasonable to base the reference cluster on a single specific location since the cluster should be representative of a particular aerosol type, whereas the data from any one location may be contaminated by intrusions of other types of aerosols. Finally, it is desirable to have a maritime cluster, something not directly obtainable from the AERONET data set.

For example, although the GSFC measurements are primarily urban industrial in nature, a glance at Fig. 1 shows that the black dots also extend into the red and green dots, and vice versa. Although it is true that occasionally the aerosol at, say, GSFC might be a product of biomass burning, it is not appropriate to use such points in defining a reference aerosol cluster. Therefore, our reference clusters are defined by combining measurements from several sites and only keeping points that lie within a Mahalanobis distance of 2 from the centroid. The overlaps are then minimized. (Note that Fig. 1 only shows two dimensions (EAE and SSA) but our calculations are carried out with all five dimensions).

We now address these questions and develop a reasonable set of reference clusters.

3.2. Urban industrial sites

To determine the characteristics of an urban industrial aerosol (as measured by AERONET) we considered the inversion results from a number of locations where such aerosols are prevalent. Specifically, we considered AERONET results for various locations in North America and in Europe.

The sites selected to represent urban industrial aerosol are presented in Table 1. The values used were taken primarily from the months June–September. The choice of locations and time periods was made by an inspection of scatterplots (primarily but not exclusively of SSA vs. EAE) as well as preliminary pie and bar charts similar to those described below in Section 4. We also relied on the site information given on the AERONET web site, information gleaned from a literature search, and descriptions of the environmental conditions as found in Wikipedia and other on-line

Table 1
Sites used to generate the reference clusters.

Urban industrial	GSFC, GISS, Columbia, Brookhaven, Billerica, MD Science Center, Stennis, UMBC, Wallops, Athens, Belsk, Hamburg, Leipzig, Lille, Mainz, Minsk, Moldova, Moscow, Paris.
Biomass	Abracos Hill, Alta Floresta, Cuiaba, Concepcion, Los Fieros, Mongu, Zambezi, Senanga, Etosha Pan, Skukuza.
Dust	Agoufou, Banizoumbou, Blida, Dahkla, Dakar, Djougou, Ouagadougou, Bahrain, Mezaira, Mussafa, Solar Village.
Mixed	Beijing, Pokhara, Pune, Silpakorn, XiangHe, Anmyon, Chen-Kung, Gosan, Gwangju, Osaka, Shirahama, Taipei.
Maritime	Lanai (and theoretical considerations).

resources.

Fig. 2 shows a scatter plot of all the North American sites (in red) and all the European sites (in blue). Since overlapping points cannot be seen, we plotted the data twice. The left panel was generated by first plotting the North American sites and then the European sites. The right panel reversed the order. One can appreciate that the European sites tend towards slightly lower values of EAE and SSA. The question arises whether we should consider the European and the American clusters as a single cluster or as two distinct clusters. Calculating the cluster to cluster five dimensional Mahalanobis distance between the two clusters yields the (dimensionless) distance of 1.50. As we shall see later when we consider cluster to cluster Mahalanobis distances, this separation is significantly less than the separation between our final reference clusters (Table 3). Consequently, in the following, we have assumed that European and North American urban industrial aerosols are sufficiently similar that they can be considered a single cluster.

3.3. *Biomass burning sites*

We now consider the biomass burning sites. In biomass burning, the emissions of gases and particles depends strongly on the combustion conditions, which can be broadly classified as flaming or smoldering (T. Lee et al., 2010). Flaming fires produce black smoke, whereas smoldering fires produce white smoke.

Flaming fires are usually due to the burning of grass and are characteristic of the African Savanna in Zambia (the regions around Mongu, Senanga and Zambezi). Similar conditions are found at Etosha Pan (in Namibia) and Skukuza (South Africa). Flaming fires (grass, etc.) have a high black carbon content and are strongly absorbing, thus yielding a lower value of single scattering albedo. On the other hand, fires burning wood (i.e., trees) tend to smolder, last longer, and are less strongly absorbing, leading to higher SSA values and white smoke. This type of fire is prevalent during the biomass burning period in South America, although there are also pasture burns in the regions of Alta Floresta and Abracos Hill. The Cuiaba site is in a woody savanna environment (Eck et al. (2013)). We considered the possibility of separating the two types of burning on the basis of the geographical location of the site, assuming that the South American biomass aerosols are due to smoldering fires and that African biomass aerosols are due to flaming fires. A problem with this assumption is the fact that some South American sites (e.g., Rio Branco, Brazil) exhibit a significant amount of urban pollution even during the biomass burning season (August–October). On the other hand, the African sites are characterized by aerosols that are almost entirely a result of biomass burning. Thus, for example, a scatter plot of SSA vs. EAE would tend to show that the South American sites have larger values of SSA than the African sites, even if one does not account for the difference in type of fire.

To test the hypothesis that there are two types of biomass burning, we generated two "reference clusters" for biomass burning, one based on the African sites and another based on the South American sites. The South American sites were: Abracos Hill, Alta Floresta, Cuiaba, Concepcion, and Los Fieros. The African Sites were: Mongu, Zambezi, Senanga, Etosha Pan, and Skukuza.

To generate the reference clusters we used the usual 5 parameters. Compared to the South American (smoldering?) cluster, the African (flaming?) cluster has nearly the same EAE, and lower values of AAE and SSA. The real index of refraction for Africa is slightly larger and the imaginary index of refraction is nearly twice as great. It is not obvious that these differences are due to the difference of flaming vs. smoldering, as they could be due to urban industrial pollution at the South American sites.

As shown by the scatterplots of Fig. 3, there is a great deal of overlap between the two types of biomass aerosol in the sense that observations from the African sites often reach into the smoldering region and observations from the South American sites often

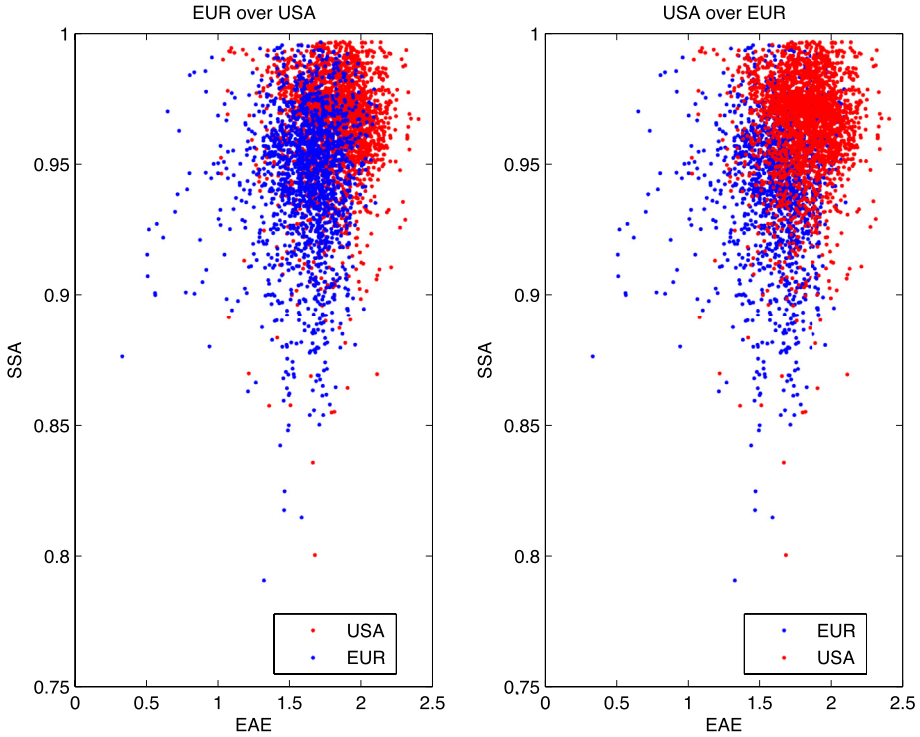

Fig. 2. A scatter plot of SSA (at 440 nm) vs. EAE (870/440) for USA urban industrial sites (red) and European urban industrial sites (blue). In the left panel, the blue dots "mask" the red dots (because the red dots were printed first). The order in which the points are printed is reversed in the second panel. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

appear to be flaming. As before, in producing Fig. 3, we plotted the points twice. We show the plots with African points plotted over the South American points in the left panel and the reverse order of plotting in the right panel. It was noted by Eck et al. (2013) that the value of IRI changes as the nature of the fires changes from flaming to smoldering, so we also plotted SSA vs. IRI (plot not shown) but the two sets of points overlapped almost completely. One reason why there is so much overlap is that the nature of the aerosols at a given site might be flaming at one time and smoldering at another.

In an attempt to define two separate clusters we added an additional condition to our definitions of the reference clusters, by requiring that the smoldering reference cluster be generated from South American measurements made during the burning season and having a single scattering albedo greater than 0.85, and the flaming reference cluster be generated from African measurements made during the burning season and having a single scattering albedo less than 0.9. These values of SSA are roughly equivalent to the values cited by Eck et al. (2013). This does achieve a separation of the two clusters. If we only keep points within two MD of the centroid, we find the cluster to cluster Mahalanobis distance between these South American and African Biomass clusters to be 4.89, a significant separation between the clusters. Nevertheless, we did not feel comfortable in trying to separate the two kinds of smoke on a geographic basis, since a given site can produce both white and black smoke. Furthermore, our technique is probably not sophisticated enough to actually distinguish between the two types of burning. Consequently, we decided to be conservative and use a single biomass burning cluster. This "general" biomass reference cluster was generated by combining values from various biomass burning sites and only keeping points that are within two MD from the centroid. The biomass burning reference cluster thus generated is illustrated below in Fig. 6.

Biomass burning is notoriously difficult to distinguish from Urban Industrial and we had difficulties in categorizing the Mahalanobis distance results for two locations, namely Bonanza Creek, Alaska and Mexico City. We discuss these in Section 4.

3.4. Dust sites

As was the case for Urban Industrial and Biomass Burning clusters, there seem to be two types of dust aerosol that we shall denote "African" and "Arabian" (although the latter has the same optical characteristics as dust from Central Asia).

The African dust sites selected were: Agoufou (Mali), Banizoumbou (Niger), Blida (Algeria), Dahkla (Morocco), Dakar (Senegal), Djougou (Benin), and Ouagadougou (Burkina Faso). The Arabian sites were: Bahrain, Mezaira (UAE), Mussafa (UAE), and Solar Village (Saudi Arabia).

The different geographic regions seem to have different types of dust. In Fig. 4 we show the difference between dust from the Sahara ("African Dust") and dust from sites on the Arabian Peninsula ("Arabian Dust"). From the figure we note that at larger values of EAE the African dust tends towards lower values of SSA, whereas the Arabian does not. However, as noted by Giles et al. (2012), for values of EAE <0.2 the SSA for African and Arabian sites had very similar values. This similarity was also observed by Kim et al. (2011).

The large number of points in Fig. 4 with EAE greater than 0.8 suggest a significant amount of fine mode aerosols is included in our sample (Giles et al. (2012) show SSA vs. EAE with many fewer points in the "tail." Their cluster is centered at 0.2 EAE and 0.91 SSA in agreement with our dust reference cluster shown in Fig. 6).

African sites like Dakar and Ouagadougou can have some episodic plumes of biomass burning smoke as well. This may partly explain why SSA is decreasing with increasing EAE for African sites. In the supplementary material, plots of aerosol types at Dakar, Ouagadougou, Capo Verde, Banizoumbou and Djougou all show small amounts of biomass aerosol in the time frame November,

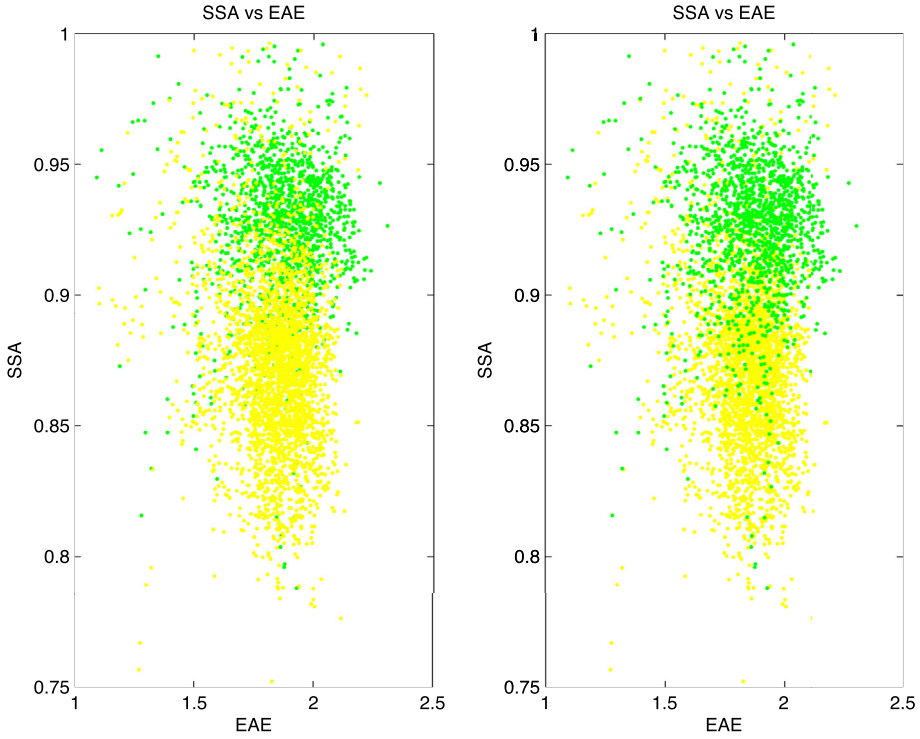

Fig. 3. SSA at 440 nm vs. EAE (870/440) scatter plot for biomass clusters for South America (Green) and Africa (Yellow), illustrating the large degree of overlap between the two clusters. (The order in which the points were plotted is reversed in the second panel). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

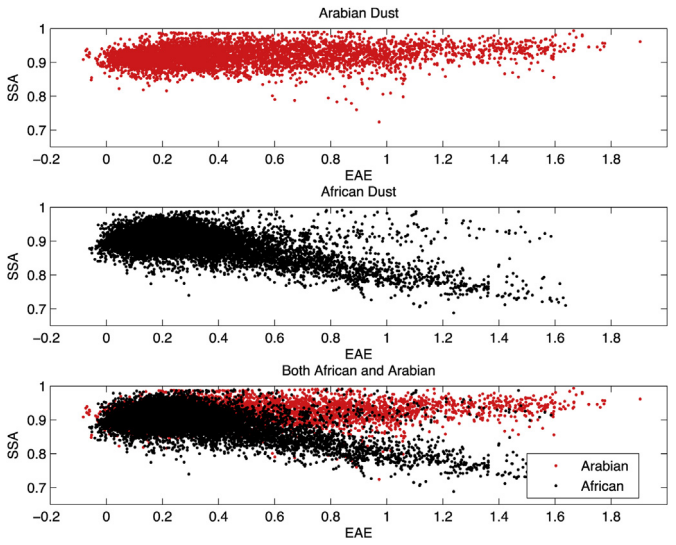

Fig. 4. SSA (440 nm) vs. EAE (870/440) scatter plot for the two types of dust. The decrease in SSA with increasing EAE in the African case, may be due to admixtures of smoke from biomass burning.

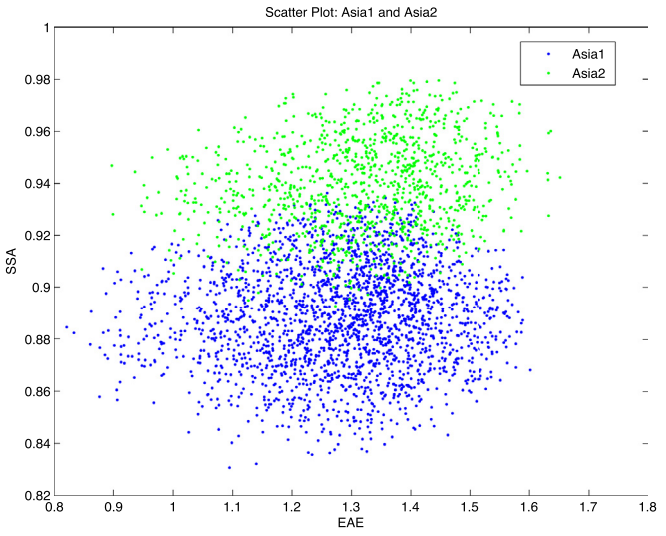

Fig. 5. Scatter plot of SSA (440 nm) vs. EAE (870/440) for two types of mixed aerosol. The blue dots are denoted Asia1 (typified by Beijing) and the green dots are Asia2 (typified by Korean and Japanese aerosol sites). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

December, January.

The cluster to cluster Mahalanobis distance between Arabian Dust and African Dust is 1.87 which is not significant, as can be appreciated from Table 3. Consequently, we defined the dust reference cluster by combining the African and the Arabian dust and keeping points that lie within 2 MD of the centroid, as shown in Fig. 6.

3.5. Mixed Aerosol sites

In discussing Fig. 1 we noted that the aerosol in Bac Giang occupies a different region of SSA/EAE space than the aerosols from other cities. We are assuming that each of these is characteristic of a given type of aerosol. It is difficult to give an appropriate name to this aerosol. It seems to be particularly prevalent in Asia, and was denoted "South East Asian Aerosol" by Cattrall et al. (2005), but as noted previously, it is also found at other locations. Russell et al. (2012) called it "Developing Urban" (Russell et al. (2012)) but this terminology is not applicable to some of the highly developed industrial sites in Japan, Korea and elsewhere in Asia. Giles et al. (2012) suggested that this aerosol is a mixture of dust and black carbon and following the terminology in Giles, we shall refer to it as "Mixed" aerosol.

In generating a mixed aerosol reference cluster based on AERONET sites located in Asia, we find (once again) that there appear to be two kinds of aerosol. This can be appreciated from an inspection of the extinction angstrom exponent and single scattering albedo illustrated in the scatterplot of Fig. 5.

In the figure the blue dots represent SSA vs. EAE for "Asia1" which consists of the data from the following sites: Beijing (China), Pokhara (Nepal), Pune (India), Silpakorn (Thailand), XiangHe (China). The Green dots are values for "Asia2" which consists of data from Anmyon (Korea), Chen-Kung (Taiwan), Gosan (Korea), Gwangju (Korea), Osaka (Japan), Shirahama (Japan), Taipei (Taiwan).

Note that the "Asia2" sites are generally in Eastern Asia whereas the "Asia1" sites stretch from China into Central and South Asia. The Asia 2 sites are similar to the urban industrial cluster defined by the North American and European sites. In fact, we find that many North American and European sites exhibit aerosol that are closer (in Mahalanobis distance) to the mixed cluster than to the urban industrial cluster. The cluster to cluster Mahalanobis distance between Asia1 and Asia2 is 2.81. This is significantly less than inter-cluster distances for our defined "reference clusters" (Table 3) and we decided to incorporate them into a single "mixed" aerosol cluster.

3.6. Maritime Aerosol

The most difficult reference cluster to define is the maritime reference cluster because the AERONET inversions (Version 2) rarely yield inverted values of absorption and sphericity for maritime aerosol. The reason is that the optical depth for this type of aerosol is frequently less than 0.4 and the inverted values of the complex index of refraction, single scattering albedo, and sphericity may not be reliable (Dubovik and King (2000), Holben et al. (2006)).

To define a maritime reference cluster we used the criteria given by Sayer et al. (2012), namely, an optical depth less than 0.2 and EAE between 0.1 and 1.0. We used the AERONET Lanai data set and selected all measurements that met these conditions. We then obtained the relative humidity at Lanai from the National Climatic Data Center of NOAA, and determined the composition of a salt solution droplet in equilibrium with the environment. This was used to determine the real refractive index. The imaginary refractive index was generated by adding a small random number to 0.001 so that IRI $= 0.001 \pm 0.0002($ rnd). Next, we used the AERONET derived size distribution and our indices of refraction in a Mie code (Bohren and Huffman (1983)) and calculated the extinction and absorption optical depths at 870 nm and 440 nm, as well as the single scattering albedo at 440 nm. This allowed us to determine the extinction and absorption angstrom exponents. We assumed the sphericity was 100%. The result of this calculation is the maritime cluster shown in Fig. 6. Note that unlike the other clusters, the maritime cluster is a theoretically derived quantity, but it is based on the AERONET inverted size distributions.

3.7. Conclusions regarding reference clusters

In Table 2 we give the centroids of our reference clusters for the five parameters. It might be noted that any one or even any two of

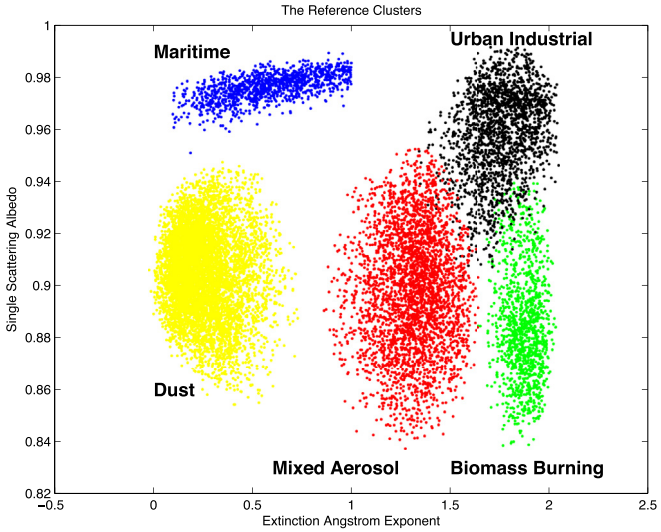

Fig. 6. The Aerosol Reference Clusters used in this study, represented as a scatter plot of SSA (440 nm) vs. EAE (870/440).

Table 2
Centroids of the five reference clusters.

	EAE	AAE	SSA	RRI	IRI
Urb Ind	1.76	1.15	0.96	1.40	0.005
Biomass	1.87	1.30	0.89	1.48	0.020
Mixed	1.32	1.20	0.92	1.45	0.011
Dust	0.28	1.75	0.91	1.47	0.004
Maritime	0.59	0.93	0.97	1.40	0.001

the parameters will not give a very good separation of the clusters. This is, of course, the benefit of using the Mahalanobis distance which utilizes all five parameters.

In Table 3 we present the cluster to cluster Mahalanobis distances. The table is symmetric across the diagonal because we are using the "pooled covariances" to calculate these Mahalanobis distances. The MD values given in the table are quite a bit larger than the distances between the two clusters for urban industrial (USA and EUR: MD = 1.50), dust (African and Arabian: MD = 1.87) and mixed (Eastern Asia and Western Asia: MD = 2.81). The biomass clusters (smoldering and flaming: MD = 4.89) are separated by a fairly large MD and it could be argued that the two types of aerosol should be treated separately. In fact, Giles et al. (2012) considered the possibly of sub-clusters within the biomass burning category. However, for simplicity (and because the division between the two clusters cannot be made solely on geographic grounds) we are treating them as a single cluster.

4. Results: aerosol classification using multi-dimensional Mahalanobis distances

Armed with the reference clusters described above we applied the Mahalanobis distance calculations to the AERONET inverted quantities to identify the aerosol type for specific sun photometer measurements. (In the analysis presented here we are using the basic five parameter set of EAE, AAE, SSA, RRI, IRI).

Table 3
Cluster to cluster 5-D Mahalanobis distances between the reference clusters.

	Urb Ind	Biomass	Mixed	Dust	Maritime
Urb Ind	0	4.93	4.84	11.31	7.70
Biomass		0	7.65	13.90	11.54
Mixed			0	9.38	6.98
Dust				0	5.70
Maritime					0

Consider an AERONET measurement at an arbitrary location. If values for the five basic parameters are given in the AERONET data set, we can calculate the MD to each of the five reference clusters. The measured aerosol is assumed to correspond to the aerosol type that is closest in MD. If one or more of the five parameters have not been inverted (usually because the optical depth is too small) we need to determine if the measurement corresponds to a maritime aerosol. This is done by first determining if the optical depth and EAE satisfy the Sayer criteria (AOD < 0.2, 0.1 ≤ EAE ≤ 1.0). If so, we then consider the sphericity. If the sphericity is given by AERONET and is greater than 80%, we tentatively assume the measurement corresponds to maritime aerosol. For each such measurement we checked the meteorological data for the site, using data from NCDC (the National Climatic Data Center), ECAD (the European Climate Assessment Dataset) and, in a few cases, from other sources such as Weather online. co.uk. If the relative humidity at the time of the AERONET observation exceeds 70%, we categorize the aerosol as probable maritime. (The value of 70% for relative humidity was chosen because it is well known that hygroscopic aerosols tend to absorb environmental water vapor for RH >70%). If all these conditions are met, we incorporate the size distribution and other parameters into a Mie code and calculate the single scattering albedo and angstrom exponents. Finally, we evaluate the Mahalanobis distance to the various reference clusters and classify the aerosol as belonging to the type specified by the nearest cluster.

As mentioned above, the sphericity is not given as an AERONET inverted quantity when AOD at 440 nm is less than 0.2. For maritime locations (such as islands or coastal regions), if we accept measurements for which sphericity has not been inverted, we often find a large number of measurements that satisfy the Sayer criteria. For example, the La Jolla, California site is located at Scripps Institute of Oceanography about 500 m from the shore of the Pacific Ocean. If we only consider those measurements for which sphericity is given, we obtain 20 "complete" sets of values of which 25% are classified as maritime, whereas, if we relax the requirement that sphericity is given, we obtain nearly 1000 values of which 93% are classified as maritime. It might be noted, however, that relaxing the sphericity condition may result in dust aerosols being classified as maritime. Therefore, the decision whether or not to include sphericity is somewhat subjective and depends on the location of the site relative to oceans and/or deserts. On occasion it is difficult to know which condition to use. For example, Lampedusa is an Italian island near the north coast of Africa. According to an analysis by Pace et al. (2006), both desert dust and maritime aerosol are prevalent at Lampedusa. In our analysis, if we only consider measurements for which the sphericity is inverted by AERONET, we obtain an aerosol that is primarily dust (81%), whereas if we relax the sphericity constraint, we obtain an aerosol that is primarily maritime (56%). Comparing our results with those given by Pace et al. (2006), the seasonal variation in the aerosol type agrees better if we use the results with the relaxed constraint.

In defining the reference clusters, we only kept points that lie within a Mahalanobis distance of 2 from the centroid. This is roughly equivalent to keeping 45% of the points. The result of this operation is presented in Fig. 6 as a scatter plot of SSA vs. EAE.

It should be kept in mind that a reference cluster can have as many "dimensions" as desired. Applying the 2 MD filter we generated reference clusters having 5 parameters: EAE, AAE, (at 870 nm/440 nm), SSA, RRI and IRI (at 440 nm). We also generated reference clusters with seven and with ten parameters (selected from SSA at 675, 870, 1020 nm, the ratio of fine to coarse mode volumes, and the sphericity). We found that using more parameters

did not significantly change the results obtained but decreased the total number of data points in our analysis mainly due to sphericity not being retrieved.

We have presented the reference clusters in plots of SSA vs EAE because in a two-dimensional representation, such a plot shows the best separation between the various clusters. However, it might be of interest to note that other 2-D plots show various degrees of separation between other variables. As an example, we show in Fig. 7 four plots in which the cluster separation can be seen to a greater or lesser degree. Plots not shown have more overlap between the clusters.

In summary, we can calculate the aerosol type for any AERONET measurement for which the inverted values of EAE, AAE, SSA, RRI and IRI are given. We can represent the aerosol types determined both as a fraction of the various types of aerosols (as a pie chart) and as the number of each type determined as a function of month of the year (as a histogram). An example of such a "pie and bar" plot for GSFC (the NASA Goddard Space Flight Center, near Washington D.C.) is shown below in Fig. 8.

Fig. 8 is an example of the basic result generated in this study. The pie chart presents the fraction of each type of aerosol from the GSFC inversions given by AERONET. In this particular example, the number of measurements that were closest to the urban industrial cluster centroid was 1005. There were 72 data points categorized as biomass burning aerosol and 48 categorized as mixed aerosol. At least some of the biomass burning events can be explained. For example, Russell et al. (2012) describe a specific situation in which a smoke plume from a Canadian forest fire was carried into the Southeastern United States. We classified two measurements as dust and four as maritime. The four maritime characterizations might be considered erroneous, however, it is interesting to note that HYSPLIT trajectories indicated that two of the maritime observations were made during a period of onshore flow from the

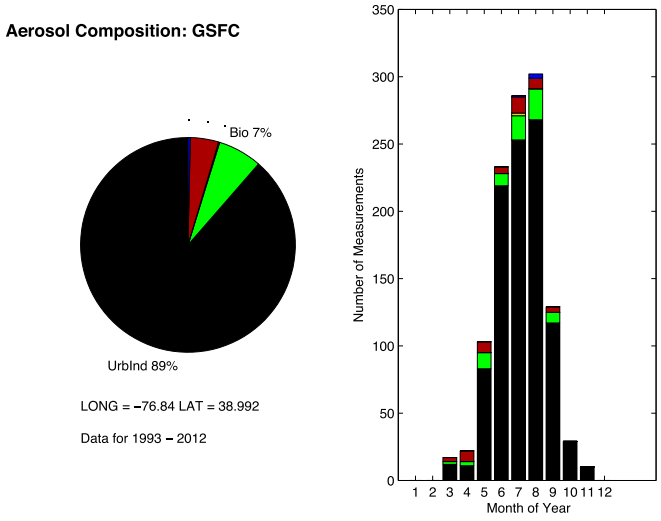

Fig. 8. Pie and Bar Chart for GSFC (Goddard Space Flight Center, near Washington D.C.). During the winter months the sky is often overcast, preventing AERONET measurements.

Northeast and the third was made at a time when Hurricane Bonnie (August 1988) was making landfall to the south of Washington D. C. and there was a strong onshore maritime flow to the north of the hurricane. The remaining maritime observation is not justified by HYSPLIT trajectories which indicate continental air. Similarly, one of the two dust characterizations occurred during a period of long range transport of North African dust to the continental United States (Perry et al. (1997)). This suggests, that our aerosol characterization scheme is fairly robust.

The bar chart on the right in Fig. 8 shows the seasonal

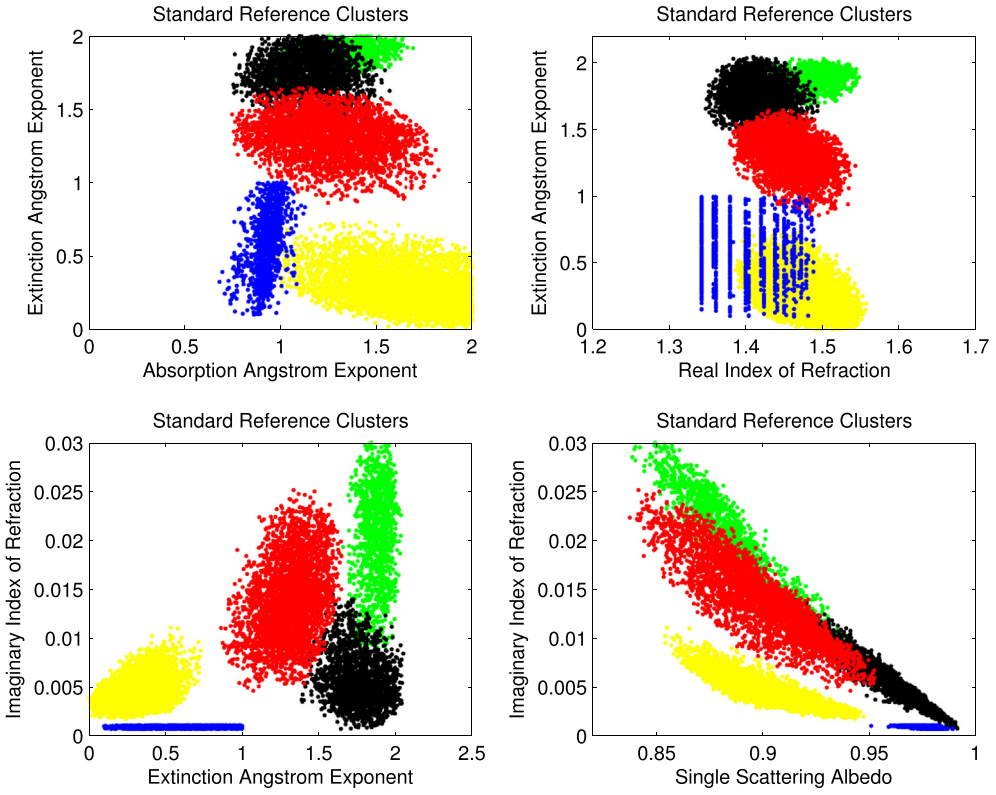

Fig. 7. Four 2-D plots showing the separation of the reference clusters in terms of other parameters. Cluster colors are the same as in Fig. 6.

dependence of the aerosol type. The most obvious feature of the bar chart is the fact that the number of measurements increases dramatically in the June to September time frame and falls to zero during the depth of winter. This is, of course, due to the fact that the AERONET measurements are clear sky measurements. Also, the AOD is not as large in winter as in summer, reducing the number of retrieved parameters. The total number of points plotted is 1131. This is much smaller than the number of points in the AERONET data set that we downloaded (9812 points) because we only kept the results of measurements for which there are inversions for all five of our basic parameters, meaning that AOD at 440 nm was greater than 0.4.

We generated pie and bar plots for all of the AERONET inversions which contained at least 100 observations having inverted values of the five basic parameters. This gave us a collection of aerosol properties for nearly 200 sites. (A few exceptions were made for interesting sites, such as Munich whose file contained 94 "complete"inversions). The pie and bar plots for all of these AERONET stations and short descriptions of each site are included in the Supplementary Material as a "Compendium of Aerosol Types."

We now consider a few representative samples of these plots. For example, the measurements made at Midway Island yield the aerosol type shown in Fig. 9. It comes as no surprise that essentially 100% of the aerosols are classified as maritime.

The Pie and Bar chart for Beijing is shown in Fig. 10. This is a more complex and interesting plot than the previous two which were nearly monodisperse. About 80% of the observations at Beijing lead to a mixed aerosol, but dust is prevalent in the spring. The amount of urban industrial aerosol peaks in January and during the summer months. The Beijing aerosol has been extensively studied by Zhao et al. (2011), Sun et al. (2013), Wang et al. (2005), Cheng et al. (2013) and many others. Cheng et al. (2013) suggest that 50% of the OC and EC in Beijing aerosols is due to biomass burning, specifically the combustion of crop residuals. Wang et al. (2005) considered PM 2.5 aerosol in Beijing and noted that Beijing is subjected to "invaded dust" throughout the year. The fine aerosol in Beijing contains SO_4, NO_3 and NH_4.

The mixed aerosol type is found to some extent at nearly all locations, but it is particularly prevalent in Asia, from India to Japan and from Northern China to South-East Asia. Among many other examples of this type of aerosol, we might mention the AERONET values of single scattering albedo and extinction angstrom exponent at the Met Station at Chiang Mai (Thailand). These are similar

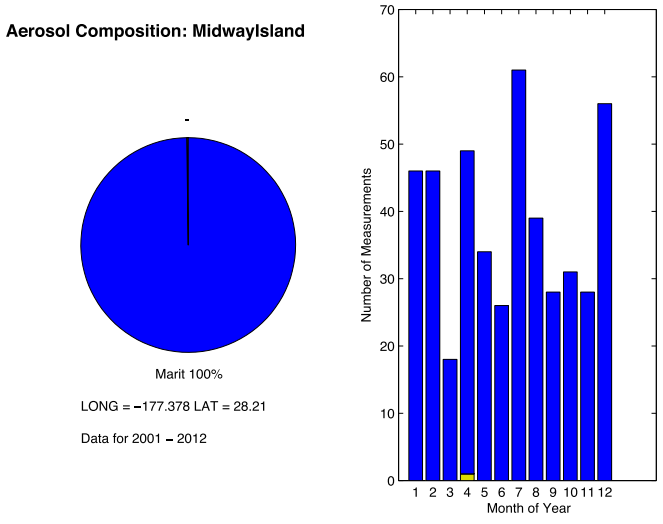

Fig. 9. Pie and bar chart for Midway island.

to those for Bac Giang shown in Fig. 1, but with even lower values of single scattering albedo. It is known that the Chiang Mai aerosol is influenced by urban pollution, seasonal mass burning and transport of aerosols from India, as illustrated in the presentation by Eck et al., 2004. The region is also affected by optically thin cirrus clouds that could result in lower values of EAE (Chew et al. (2011), and Huang et al. (2012)). Although we classify these aerosols primarily as "mixed," J. Lee et al. (2010) state that they are primarily black carbon.

See et al. (2006) considered the chemical and physical properties of South East Asian aerosols and Reid et al. (2013) give an in-depth review of aerosol properties in South East Asia, describing the various components that comprise the Asian aerosol. Many studies suggest that the aerosol commonly found in Asia contains dust, black carbon, sulfates and nitrates.

Giles et al. (2012) showed by electron microscopy that the aerosol of Kanpur (India) is sometimes composed of large dust particles mixed with open clusters of black carbon spherules. The interpretation of the image was consistent with particles identified by Martins et al. (1998). Giles et al. (2012) refer to this as "Mixed BC and Dust," referring to the dominant aerosol types. It might be noted that sulfates and nitrates are not the dominant absorbing type in regions influenced by black carbon and dust, because black carbon and dust absorb more strongly than sulfates or nitrates.

Fig. 11 through 16 give the pie and bar charts for a number of different sites that were selected to illustrate the variety of aerosol types found in different locations.

Fig. 11 gives our results for Bahrain, a desert site situated on the Persian Gulf. Its aerosol is primarily dust with a small amount of mixed aerosol. Although the AERONET site is near the shore, we only find a tiny fraction of the aerosol characterized as maritime. This result is not surprising when one considers that out of over 90,000 hourly measurements of relative humidity at Bahrain between 1996 and 2006, only on 9 days on which AERONET measurements were made was the relative humidity greater than 70%.

The aerosol at Bahrain shows a somewhat higher proportion of mixed aerosol than that of Kuwait, Solar Village or Eilat (in Israel). Based on the discussion in Sokolik and Toon (1999) we would expect the dust aerosol at these sites to be primarily Kaolinite, Palygorskite, and Montmorillonite. Kaolinite and Montmorillonite have single scattering albedos at 440 nm between 0.92 and 1.0 (Sokolik and Toon (1999)). The average values of single scattering albedo at the three sites mentioned are all equal to 0.91, somewhat lower than might be expected, perhaps due to the influence of mixed aerosols, or, as noted by Giles et al. (2012), instrument artifacts may decrease the calculated values of SSA. It might be noted that in considering dust aerosols, size is an important parameter, but it has not been included in our analysis.

Fig. 12 presents the aerosol in Crete (at the FORTH Institute at Heraklion). This aerosol is primarily dust and urban industrial but it has a significant amount of maritime aerosol, as might be expected at a site on a Mediterranean island. The urban industrial aerosol is generally observed during the spring and summer. Russell et al. (2014) compared AERONET derived aerosol types with POLDER derived aerosol types at Crete, obtaining 94 coincidences between AERONET measurements and satellite overpasses. It is difficult to compare our results with those of Russell et al. (2014) because they used different aerosol categories than we do. However, in terms of our aerosol categories, Russell et al. (2014) found 38% dust, 21% maritime and 33% combined urban industrial and mixed. Except for the dust component, this compares reasonably with our results given in Fig. 12 which included a much larger collection of data points. A further comparison of the aerosol type at Crete with the aerosol types estimated by the CALIPSO lidar system showed that of five "coincidences" four were in agreement with our evaluations.

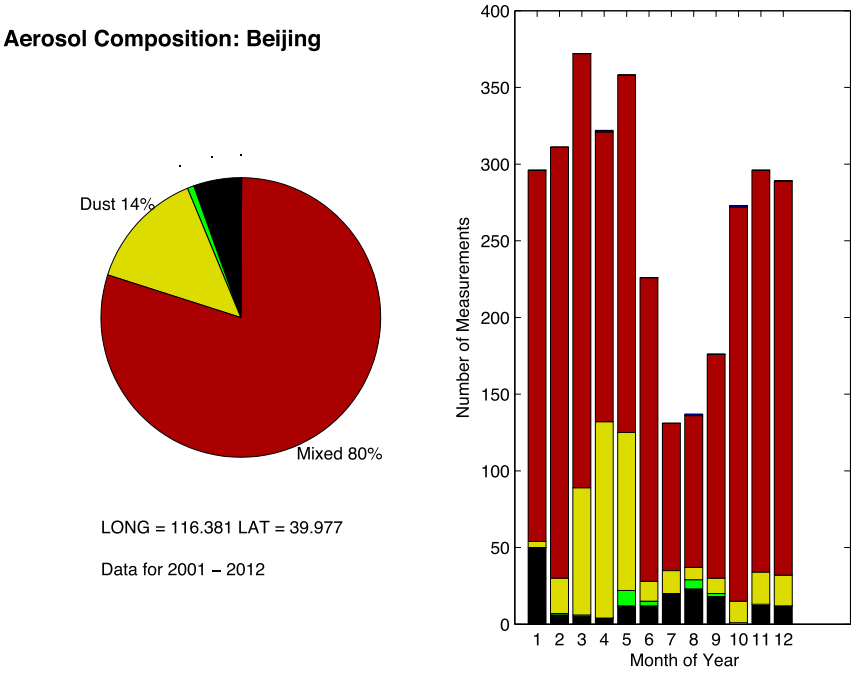

Fig. 10. Pie and Bar chart for Beijing. The aerosol is primarily mixed, but note significant amounts of dust during the Spring.

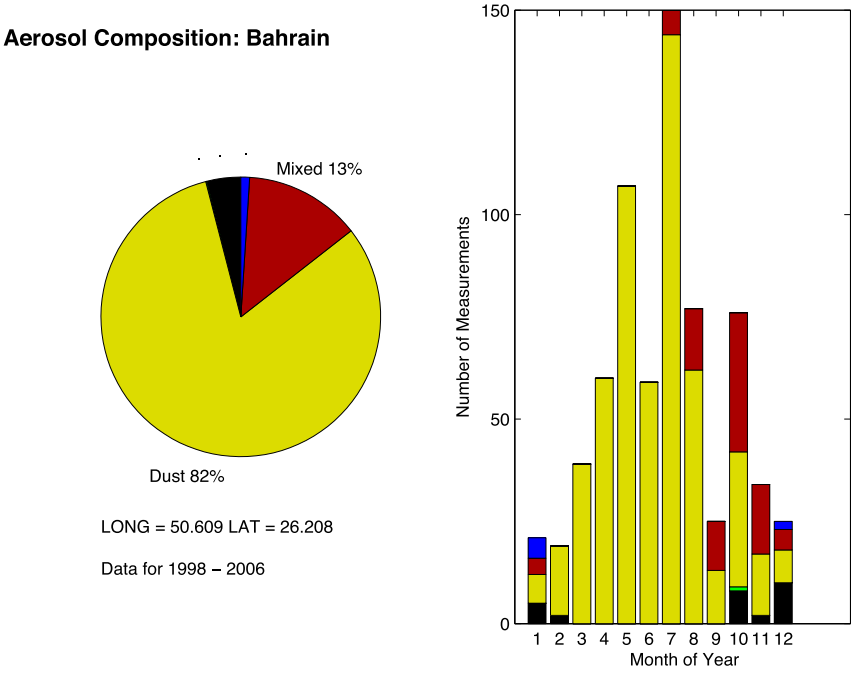

Fig. 11. Pie and Bar chart for Bahrain.

Once again, the comparison is difficult because of the different conventions on aerosol type. On one occasion (23 May 2009) both CALIPSO and our evaluation agreed that the aerosol was maritime, and on 27 April 2011 the two calculations led to dust. On 2 July 2006 CALIPSO categorized the aerosol as polluted dust, and we described it simply as dust. On 3 August 2006 CALIPSO categorized the aerosol as "smoke" and we described it as urban industrial which is, presumably, the same thing since CALIPSO does not distinguish between urban industrial and biomass burning aerosols. However, on 10 August 2009 the CALIPSO category was dust and our calculation led to maritime. It might be mentioned that the CALIPSO tracks lie about 80 km to either side of the AERONET site at Heraklion.

Fig. 13 shows the aerosol type at Bonanza Creek (Alaska). This site led to a difficulty with our Mahalanobis distance aerosol typing because our results indicated that the aerosol at Bonanza Creek, Alaska was primarily urban industrial during the month of August for 2004 and 2005. Bonanza Creek lies about 30 km downwind of Anchorage, Alaska and perhaps some urban industrial aerosol was carried down the valley joining the two locations, but our

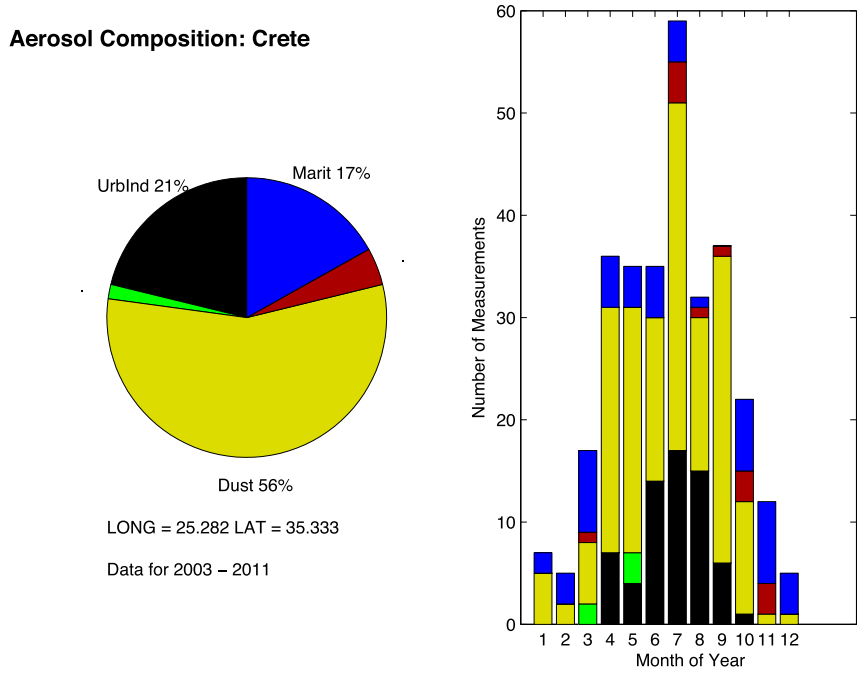

Fig. 12. Pie and Bar chart for Crete.

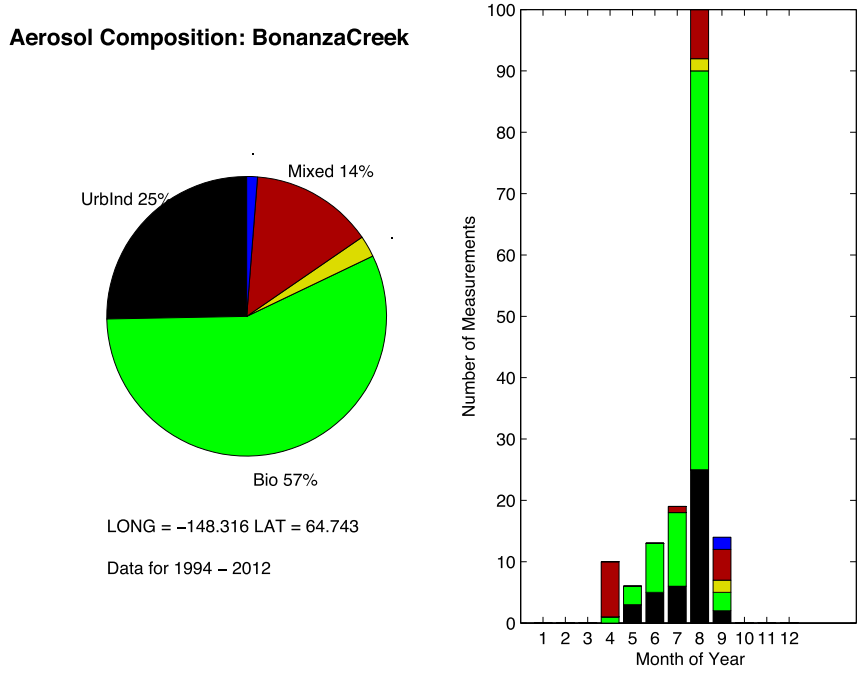

Fig. 13. Pie and Bar chart for Bonanza Creek, Alaska.

categorization was not consistent with the fact that August 2004 and 2005 were periods of intense forest fire activity in Alaska. Satellite views show the region around Bonanza Creek to be blanketed with smoke. Considering the Bonanza Creek retrievals more carefully, we noted that they exhibited anomalously large values of AAE. Consequently, we included in our analysis the condition that if a particular measurement was categorized as urban industrial but had a value of AAE $\geq$1.6, it would be re-categorized as biomass burning. This "ansatz" led to more reasonable values for the Bonanza Creek aerosol (as shown in Fig. 13) and did not lead to any unexpected changes in the aerosol at other locations. For example, for data obtained at Goddard Space Flight Center (GSFC) the ansatz led to an increase in aerosol measurements characterized as biomass by less than 2%.

The Bonanza Creek results shown in Fig. 13 indicate that the aerosol is primarily biomass burning, but also shows urban industrial and mixed aerosol. Note that biomass aerosol dominates in August. It is probable that some of the aerosols we have characterized as urban industrial are actually biomass burning. (It is well known that biomass burning aerosols are difficult to distinguish

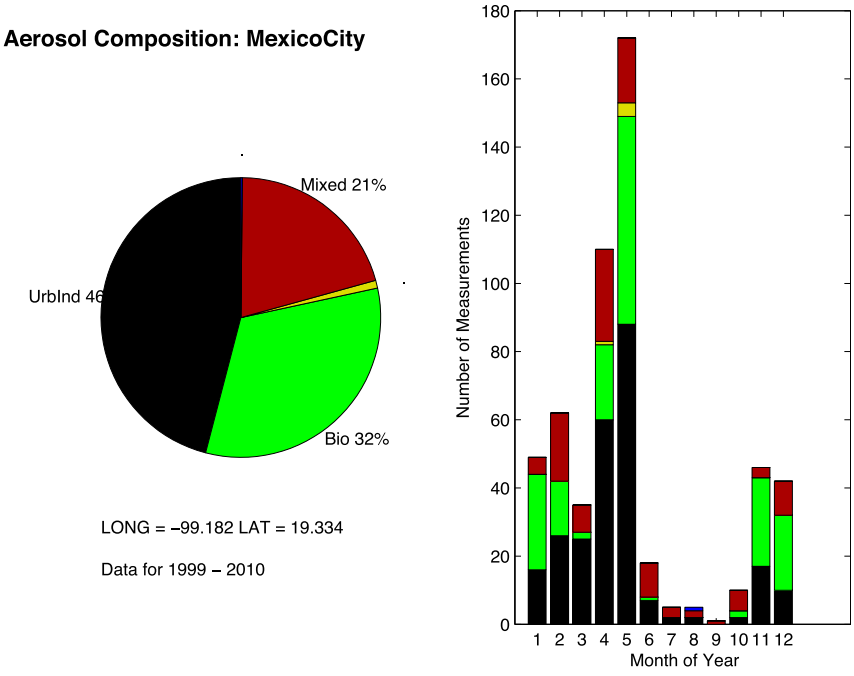

Fig. 14. Pie and Bar chart for Mexico City.

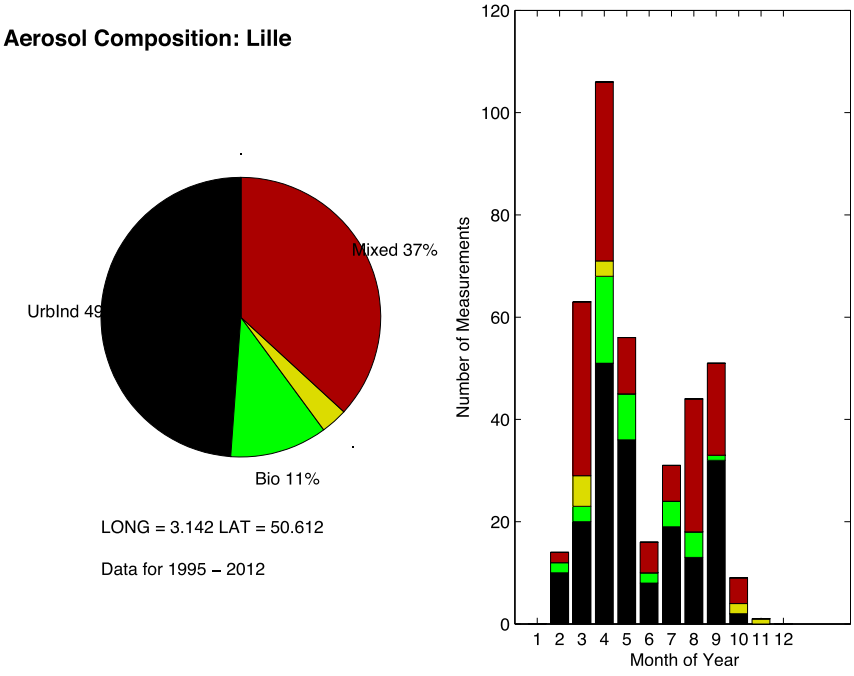

Fig. 15. Pie and Bar chart for Lille.

from urban industrial aerosols).

Fig. 14 presents the aerosol typing and seasonal variation in the aerosols of Mexico City. The classification of the Mexico City aerosol led to a second ansatz. Our Mahalanobis calculations indicated that the Mexico City aerosol was primarily due to biomass burning. There is, undoubtedly, a significant amount of biomass aerosol in the Mexico City atmosphere. Yokelson et al. (2007) point out that Mexico City is surrounded by forested mountains and there are numerous fires in the dry months from November to May (This agrees with our month-by-month bar chart of Fig. 14 which shows significant amounts of biomass burning aerosol during these months). Yokelson et al. estimate that biomass burning contributes 70–80% of the particulate mass in an area that includes metropolitan Mexico City and adjacent mountains, and point out that much of the emissions from fires could pass above ground level monitors. Nevertheless, it is generally accepted that the Mexico City aerosol is primarily urban industrial. Consequently, when an aerosol was classified as biomass burning, but the extinction angstrom exponent was less than 1.7 and the single scattering albedo was greater than 0.85, we reclassified it as urban industrial.

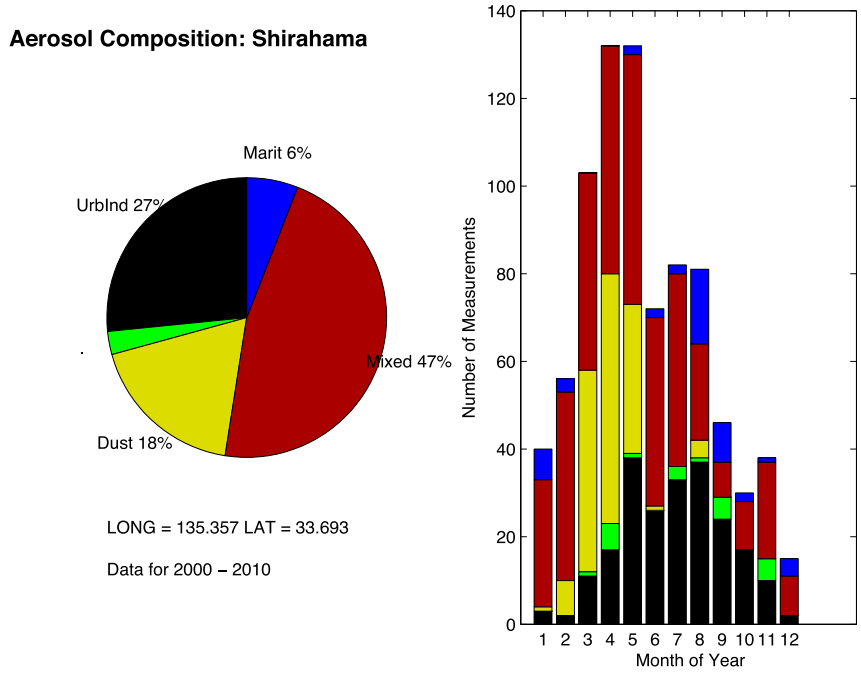

Fig. 16. Pie and Bar chart for Shirahama.

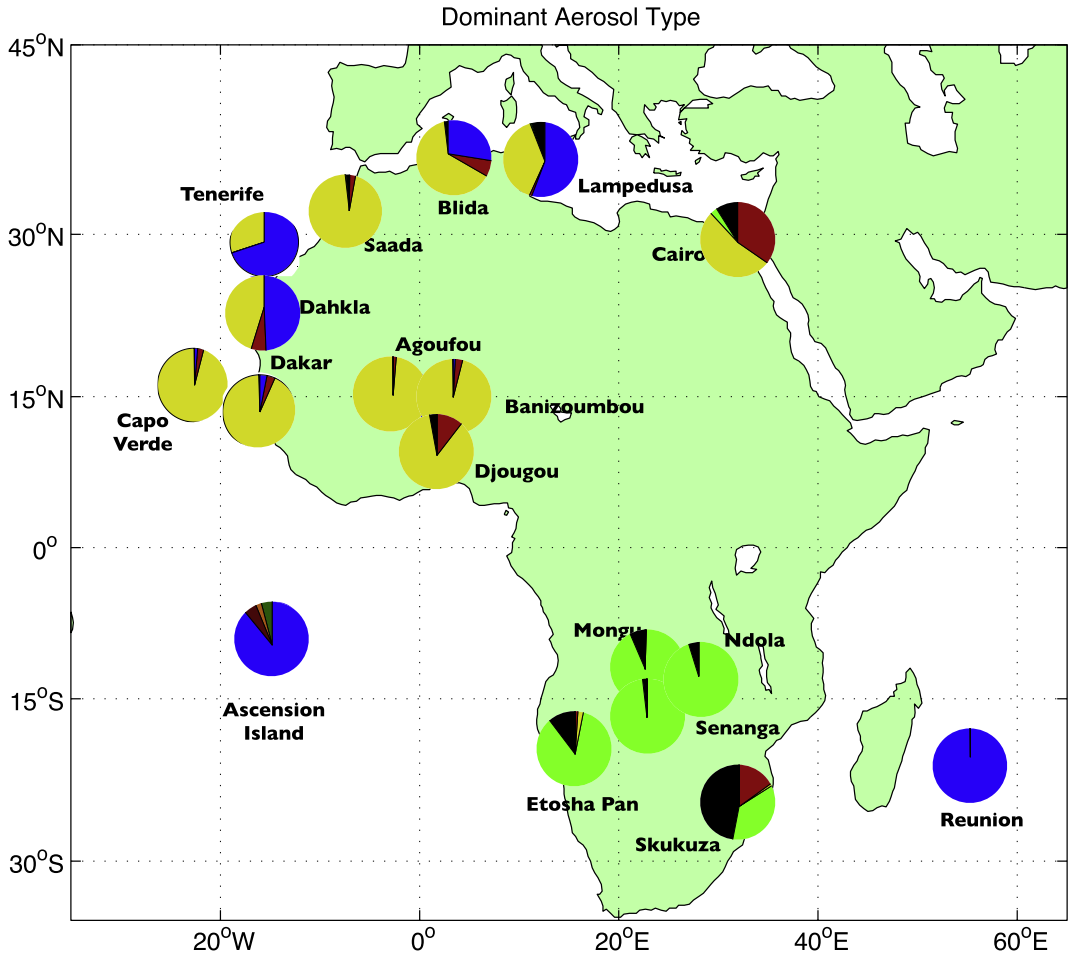

Fig. 17. Pie plots showing aerosol types present at various AERONET sites in Africa. Note prevalence of dust in Northern Africa and biomass in Southern Africa. In this and subsequent figures the size of the pie is constant and not indicative of the optical depths at the location.

This ansatz resulted in the Mexico City aerosol being classified as predominantly urban industrial, but had essentially no effect on the aerosol classification of any of the other sites. We do not have any theoretical justification for introducing this condition. As shown in Fig. 14, the ansatz led to results indicating that biomass burning accounts for 32% of the Mexico City aerosol whereas urban industrial and mixed aerosol combined account for 67%. It might be mentioned that the pie chart for Mexico may be somewhat misleading because there are so few retrievals during the months of June through October. This is the rainy season, when urban industrial aerosol would dominate over biomass burning aerosol. Consequently, an aerosol measuring system that did not depend on clear skies might give a significantly different picture of aerosol typing in Mexico City. The single measurement characterized as maritime aerosol in the Mexico City aerosol is probably a dust aerosol that has been misclassified as maritime.

The aerosol measured at Lille (Fig. 15) is predominantly urban industrial, as would be expected at this industrial French city near the Belgian border. Urban industrial and mixed aerosol account for fully 86% of the AERONET derived aerosols. About 11% of the aerosol is characterized as biomass burning and is present throughout the year. It is not clear whether this aerosol is of local origin or transported from elsewhere, but it might be noted that such a fraction of biomass burning aerosol is typical of most Northern European cities.

The aerosol of Shirahama, Japan shown in Fig. 16 is mostly mixed, but includes a significant amount of urban industrial as well as maritime aerosol. The urban industrial and mixed aerosols together account for 74% of the Shirahama aerosol. The dust occurs mainly in the spring and probably originates in mainland Asia. The maritime aerosol can be accounted for by the fact that Shirahama is a resort town well known for its beaches.

The pie and bar charts in Fig. 8 through 16 are intended to give an overview of the aerosol types and their seasonal variation at several representative locations. These are just a sampling of the total number of such plots which we have generated. Similar plots for all of the locations for which we calculated the aerosol types, are presented in the Supplementary Material, along with brief descriptions of the AERONET site where the aerosol was observed.

5. The aerosol climatology

We can use the tools we developed for aerosol typing to generate an aerosol climatology. To present this climatology we place a pie plot on each of the AERONET sites, as shown in Fig. 17 for the African continent. It should be kept in mind that the pie charts are “point source” evaluations of the aerosol type and may not be representative of the entire region. (The size of the pie is constant and not indicative of the optical depths at the location).

The figure shows the unsurprising result that sites in Northern Africa are dominated by dust aerosols and sites in Southern Africa are dominated by biomass burning. The islands off the coast of Africa (Lampedusa, Ascension Island, Reunion) are primarily maritime, but the Capo Verde aerosol is predominantly dust. The aerosol at Tenerife is mostly maritime but it has a significant amount of dust. As expected, Cairo, on the very edge of the Sahara desert, has a large fraction of dust aerosol, but being an industrial city with a metropolitan population of about 20 million, it is not surprising that fully 37% of the aerosol is mixed. It might be mentioned that there are two AERONET sites in Cairo, and our Mahalanobis calculations lead to the same aerosol typing at both locations, giving us additional confidence in our technique.

A similar plot for Europe is shown in Fig. 18. It is interesting to note that the European aerosol is mainly urban industrial in the north, and mainly dust in the south. As expected, sites on islands or on the coast have maritime aerosol and northeastern Europe exhibits a significant amount of biomass aerosol. It is possible that some of this is a consequence of the fact that urban industrial

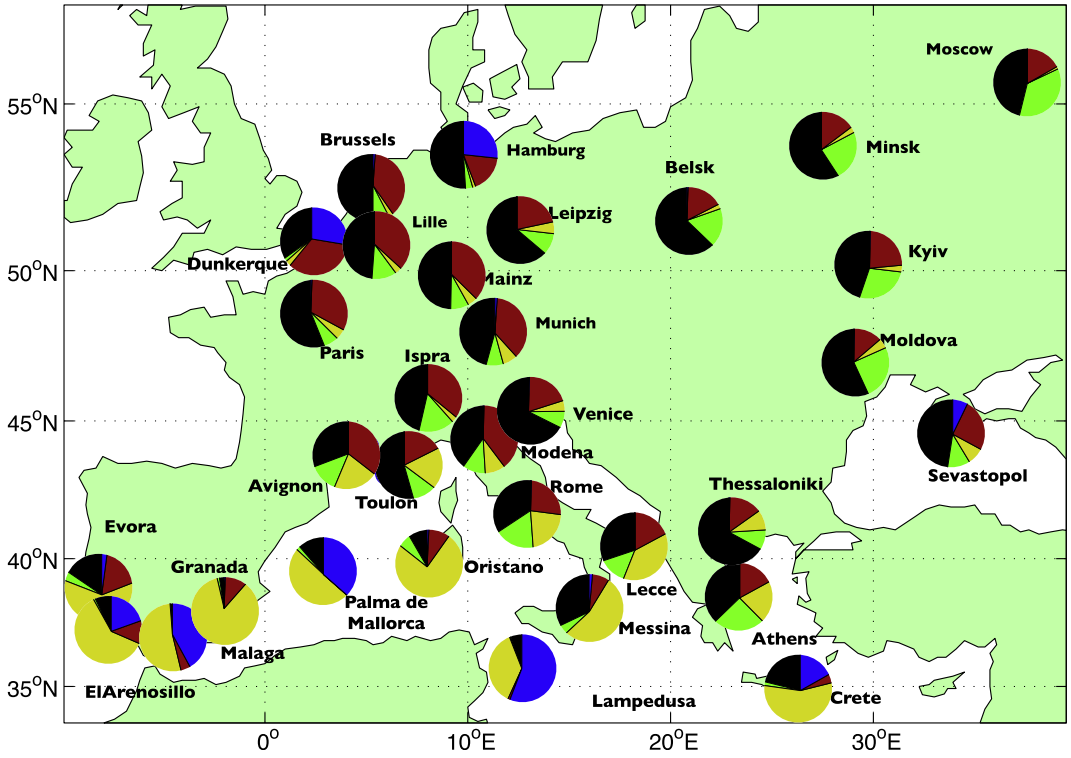

Fig. 18. Pie plots showing aerosol types present at various AERONET sites in Europe. Southern Europe shows a prevalence of dust aerosol, Northwestern Europe has a significant amount of urban industrial aerosol, and Eastern Europe has approximately equal amounts of biomass and urban industrial aerosol.

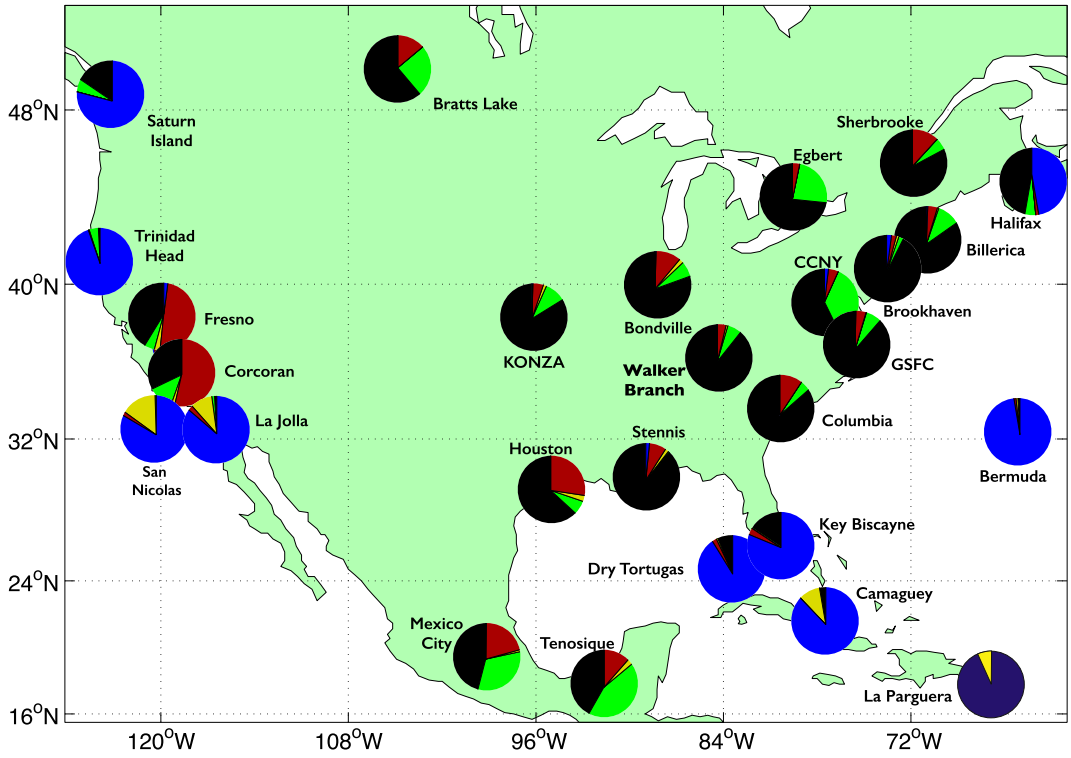

Fig. 19. Pie plots showing aerosol types present at various AERONET sites in North America. Coastal locations indicate significant amounts of maritime aerosol, but the principal aerosol type in continental regions is urban industrial.

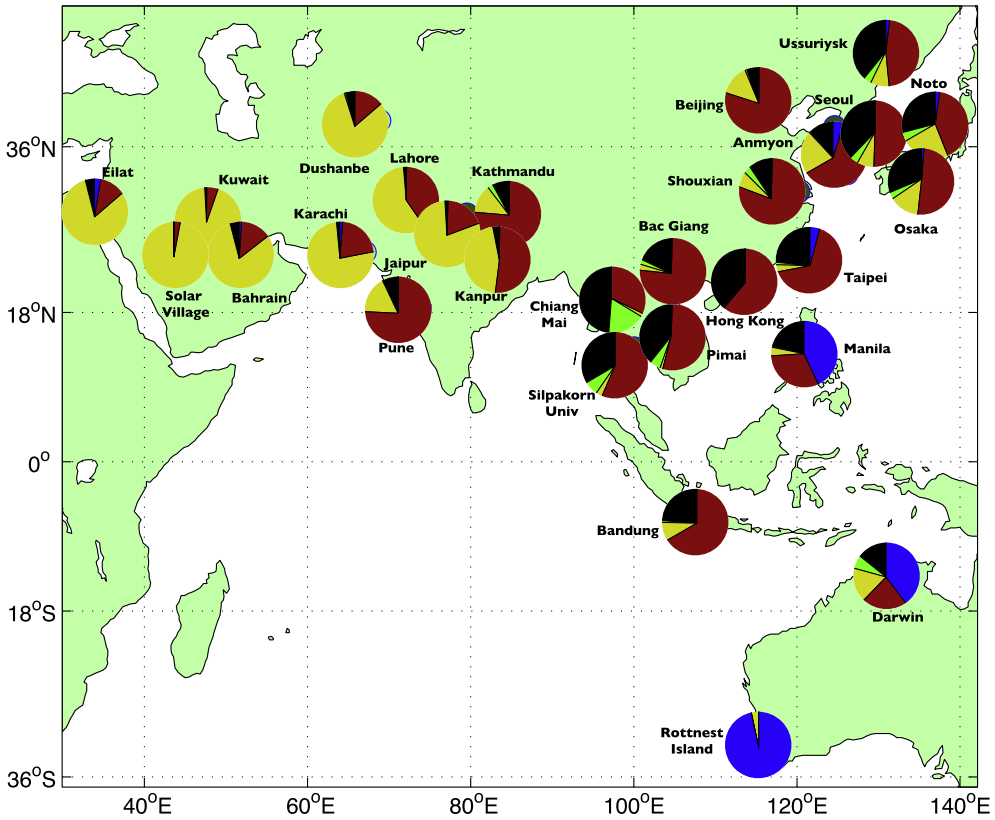

Fig. 20. Pie plots showing aerosol types present at various AERONET sites in Asia. The predominant aerosol is mixed. Japanese and Korean sites have a significant amount of urban industrial aerosol. In the East the aerosol is primarily dust, as expected from the deserts in that region.

aerosols at low relative humidity may yield biomass type characteristics. The aerosols in the larger cities appear to be both urban industrial and mixed.

In Fig. 19 we present the aerosol types in North America. Note the maritime aerosol along the coasts and on islands, particularly in the Caribbean. The eastern seaboard is primarily urban industrial with some mixed aerosol and small amounts of biomass aerosol. The mid-west has more biomass but is still dominated by urban industrial aerosol. Fresno, California, shows a large amount of mixed aerosol. The Caribbean sites, Guadeloupe, Camaguey in Cuba, Cape San Juan and La Parguera in Puerto Rico, all show significant amounts of dust during the summer months. This transport of Saharan dust across the Atlantic during this time period is a well known phenomenon. The west-coast sites at La Jolla, San Nicolas, Trinidad Head and Saturna Island are located very near the sea shore and exhibit maritime aerosol.

Fig. 20 presents pie charts for Asian AERONET sites. Note the predominance of mixed aerosol in the East and dust aerosol in the West. Southeastern Asia exhibits a significant amount of biomass burning. The rather large fraction of maritime aerosol in Manila is surprising since the AERONET site is in the heart of the city and not very near the shore of Manila Bay. Rottnest Island and Darwin in Australia are significantly impacted by maritime aerosol. The aerosol of Japan and Korea is primarily urban industrial or mixed.

The aerosol classifications in some Asian locations do not agree with expected aerosol types. In general, the mixed aerosol category may be overestimated. For example, one would expect Kanpur to have more urban industrial aerosol. Bandung would be expected to show mainly urban industrial with episodic biomass burning. Our Southeast Asia results may be affected by cloud contamination. This problem was discussed by Omar et al. (2013) in their comparison of AERONET and CALIPSO results. This may be alleviated in the AERONET Version 3 cloud screening algorithm.

As shown in Fig. 21, South America has relatively few sites that meet our criterion of data sets having at least 100 records of retrieved values. The cities (Buenos Aires, Cordoba, Sao Paulo, Brasilia and Arica) show mixed and urban industrial aerosols with significant amounts of biomass burning whereas the sites in the Amazonian region show primarily biomass burning.

Finally, Fig. 22 shows the geographic distribution of the

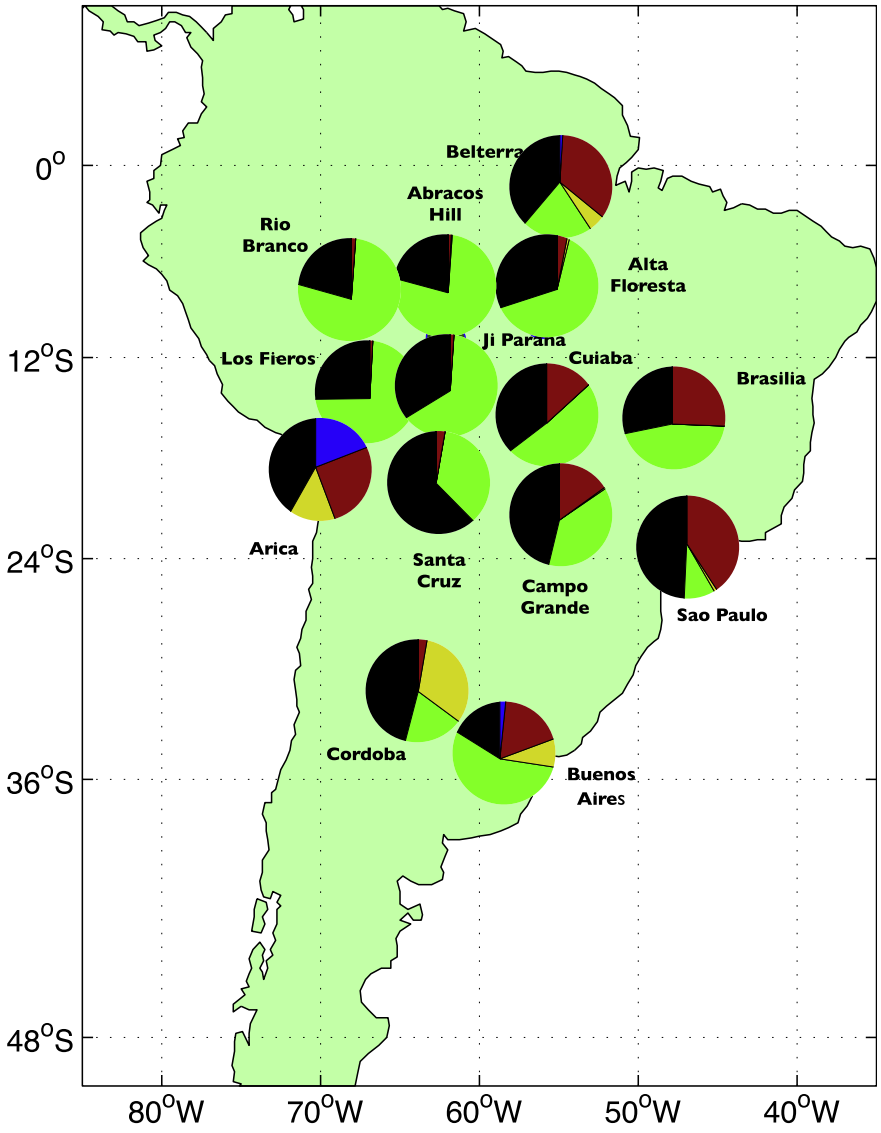

Fig. 21. Pie plots showing aerosol types present at various AERONET sites in South America. The cities have mixed aerosol as well as biomass, whereas the amazonian sites are primarily composed of biomass aerosol.

dominant aerosol type, based on inversions and MD calculations for some 200 AERONET sites. The figure shows that coastal cities and islands tend to have a predominance of maritime aerosol. The aerosol of the United States, Canada and Northern Europe are mainly urban industrial. There is a dust dominated swath stretching from North Africa across the Arabian Peninsula and into Asia. The Eastern Asian aerosol is mainly of the mixed type. A band of biomass burning aerosol stretches globally in the southern hemisphere, across the Amazon and Southern Africa. There is a comparable biomass dominated band in the Northern Hemisphere, but it is obscured by the regions of urban industrial aerosol.

6. Discussion and conclusions

We have shown that the Mahalanobis distance calculation using a set of reference clusters can be used to obtain reasonable and consistent identifications of aerosol types from the AERONET inverted aerosol properties.

The aerosol identification scheme described in this paper may not be flawless, yet it generates very reasonable results. To quantify the accuracy of our results would require knowing the column averaged aerosol type at the particular time and location of an AERONET observation. In theory, one could rely on the identifications made by other systems such as CALIPSO, MODIS, ground based lidar systems, etc. However, comparisons between AERONET and other systems are not easily carried out and often yield ambiguous results (Omar et al. (2013) compared CALIPSO optical depths with 1081 coincident AERONET values; Sawamura et al. (2014) compared 11 lidar and in-situ measurements with AERONET). Although we have made preliminary comparisons of the Mahalanobis distance identifications to other methods, we have not carried out an in depth study, and we leave it as the subject of planned future studies. It might be noted, however, that such a comparison is hindered by the fact that nearly every system has a different way of describing the various aerosol types.

An obvious source of error is occasionally identifying maritime aerosol where we do not expect it. It is easy to identify some fairly obvious mis-identifications. For example, one of the Mexico City observations is classified as maritime aerosol, which is almost certainly wrong. However, in other cases, such as the aerosol on the island of Lampedusa, it is difficult to know whether the aerosol should be classified as maritime or dust.

A more serious question is the validity of our identification of 32% of the Mexico City aerosol as due to biomass burning, although as mentioned above, studies by Yokelson et al. (2007), as well as Crounse et al. (2009), Fast et al. (2007), Patadia et al. (2013), and Baumgardner et al. (2009) indicate that a significant portion of the Mexico City aerosol is due to biomass burning. Similarly, we categorized 58% of the aerosol in New York City (at CCNY) as urban industrial, but our identification of 34% as biomass is troubling. Thus, our analysis of the aerosol at both New York City and Mexico City suggest that urban industrial aerosol is being classified as biomass. The opposite effect occurs in our analysis of Bonanza Creek where 22% of the aerosol is classified as urban industrial. Although there may be some urban industrial aerosol at this site, 22% is probably too high. This could be due to biomass aerosols with a low value of AAE, in which case our scheme would not be able to distinguish them from urban industrial aerosol. (A lower value of AAE would result from a more significant contribution of black carbon, as for example, fresh smoke from flaming combustion).

Another criticism of our results might be that we are overestimating the amount of mixed aerosol, especially in South East Asia. Problems of over or underestimating aerosol types can be addressed by manipulating the reference clusters, but we have not done so in the present study.

A problem in comparing our aerosol identification technique with the CALIPSO method arose when we compared our results for the AERONET site at Camaguey, Cuba with the CALIPSO aerosol types. Out of 12 coincidences, our calculations consistently showed the aerosol to be maritime whereas CALIPSO described it to as “polluted continental” except for one measurement which it categorized as “polluted dust.” (It should be mentioned that “coincidences” may be more than 40 km from the AERONET site). Plots for the coincident measurements at Camaguey show the values of SSA and EAE to be located in the heart of our Maritime cluster, making us suspicious of the CALIPSO identification in these cases, especially since a preliminary comparison at other locations shows that our results generally do agree the CALIPSO identifications, particularly in places where the aerosol type is fairly constant, such as at Mongu or Solar Village. Nevertheless, the disagreement at Camaguey is disturbing and clearly worth further study.

Comparing our results with those presented by Omar et al. (2005) for dust, and maritime aerosols (the two categories that

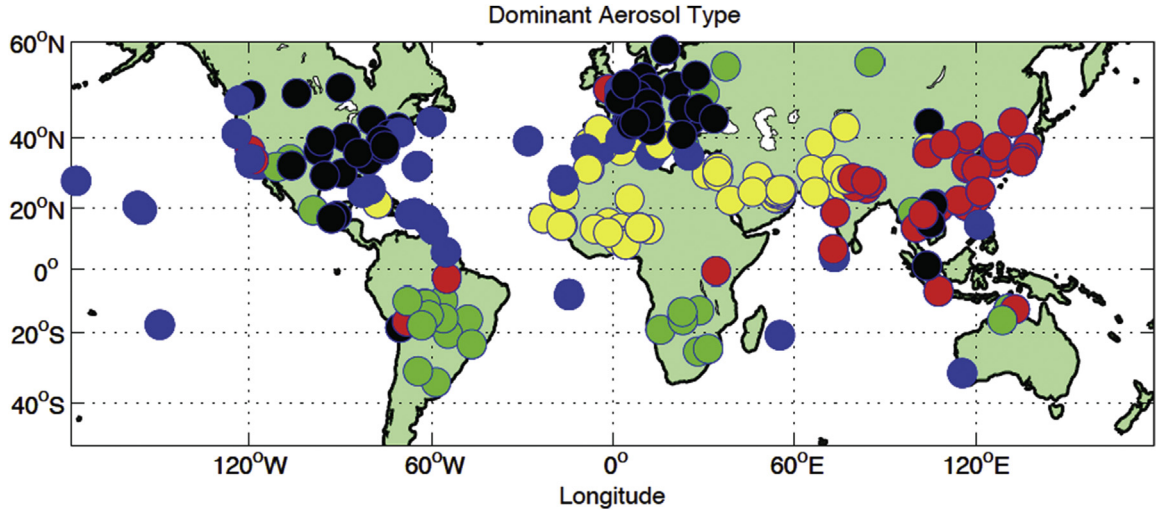

Fig. 22. The dominant aerosol type as a function of global distribution. Black circles represent urban industrial aerosol, green circles represent biomass burning aerosol, orange circles represent dust, dark red circles represent mixed aerosol and blue circles represent maritime aerosol. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are in both studies) we find very good agreement, particularly since a number sites appear in more than one of the Omar et al. (2005) categories. For example, La Parguera (in Puerto Rico) is characterized both as dust and as maritime in Omar et al. (2005) whereas we find that aerosol is 90% maritime and 10% dust. Similarly, Omar et al. (2005) categorize Kanpur (India) as dust whereas we find this site is 45% dust and 52% mixed. Omar et al. (2005) give the percentage of the various categories at 6 sites, Cape Verde, Banizoumbou, Solar Village, Mongu, Mauna Loa, Anmyon, GSFC, Lanai and Skukuza. Our results are quite similar. For example, we find that at Anmyon the aerosol is 62% mixed, 22% dust and 11% urban industrial. Omar et al. (2005) indicate that the aerosol is primarily "polluted continental" with significant amounts of dust in March April May. We also find that the dust observed occurs primarily in March, April and May. Our main disagreement with the results of Omar et al. (2005) are for Mongu (Zambia) which we find to be nearly 100% biomass burning but they find significant amounts of "polluted continental" aerosols. This points out once again the difficulty in distinguishing between biomass burning and urban industrial aerosols.

We find that the global climatology of Taylor et al. (2015) agrees in general with most of our characterizations. It should be mentioned once again that ours is a "point source" analysis whereas Taylor et al. (2015) generated a global climatology on a $2.5 \times 2^{\circ}$ grid. Of course, it is difficult to compare their 10 cluster analysis with our 5 clusters, but it appears that the greatest difference is the swath of "sulfurous dust" across Southern Europe where we find dust and mixed aerosols, and some maritime aerosol on Mediterranean islands and coastal sites.

In spite of the few difficulties we have just described, we believe that the scheme we have developed generally leads to reasonable aerosol identifications. This is clearly the situation for sites where there is little question as to the aerosol type, such as desert locations or regions where biomass burning is prevalent. It is not so obvious for locations where there are several different aerosol types, such as in Crete (Fig. 12). Nevertheless, comparisons with other classification schemes (Omar et al. (2005), Russell et al. (2014), etc.) gives us a degree of confidence in our results. For the sake of simplicity we used 5 aerosol types, but as described in Section 3 it would be possible to split some of our clusters into two, thus generating a number of clusters more closely approximating those of Taylor et al. (2015) and others. We suggest that a satellite-borne system able to determine the optical parameters used in this study would be helpful in identifying aerosol types from space.

Acknowledgements

We thank the AERONET principal investigators and their staff for establishing and maintaining the 190 sites used in this investigation. We also thank the AERONET team for calibration and maintenance of instrumentation and quality assuring and processing these data.We acknowledge interesting conversations with Dr. Philip B. Russell and Dr. Meloe Kacenelenbogen. The idea of using the Mahalanobis distance was suggested to us by Dr. Sharon Burton. We would like to acknowledge the insightful and helpful comments of the two anonymous reviewers. One of us (MG) wishes to acknowledge Mr. Joshua Schuler of Lemelson Lab, MIT.

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.atmosenv.2016.06.002.

References

Baumgardner, D., Grutter, M., Allan, J., Ochoa, C., Rappenglueck, B., Russell, L.M., Arnott, P., 2009. Physical and chemical properties of the regional mixed layer of Mexico's Megapolis. Atmos. Chem. Phys. 9, 5711–5727.

Bergstrom, R.W., Pilewskie, P., Russell, P.B., Redemann, J., Bond, T.C., Quinn, P.K., Sierau, B., 2007. Spectral absorption properties of atmospheric aerosols. Atmos. Chem. Phys. 7, 5937–5943. http://dx.doi.org/10.5194/acp-7-5937-2007.

Bohren, C.F., Huffman, D.R., 1983. Absorption and Scattering of Light by Small Particles. John Wiley & Sons, New York, p. 530.

Burton, S.P., Ferrare, R.A., Hostetler, C.A., Hair, J.W., Rogers, R.R., Obland, M.A., Butler, C.F., Cook, A.L., Harper, D.B., Froyd, K.D., 2012. Aerosol classification using airborne High Spectral Resolution Lidar measurements - methodology and examples. Atmos. Meas. Tech. 5, 73–98. http://dx.doi.org/10.5194/amt-5-73-2012.

Cattrall, C., Reagan, J., Thome, K., Dubovik, O., 2005. Variability of aerosol and spectral lidar and backscatter and extinction ratios of key aerosol types derived from selected aerosol robotic network locations. J. Geophys. Res. 110 http://dx.doi.org/10.1029/2004JD005124.

Cheng, C.Y., Engling, G., He, K.B., Duan, F.K., Ma, Y.L., Du, Z.Y., Liu, J.M., Zheng, M., Weber, R.J., 2013. Biomass burning contribution to beijing aerosol. Atmos. Chem. Phys. Discuss. 13, 8387–8434. http://dx.doi.org/10.5194/acpd-13-8387-2013.

Chew, B.N., Campbell, J.R., Reid, J.S., Giles, D.M., Welton, E.J., Salinas, S., Liew, S.C., 2011. Tropical cloud contamination in sun photometer data. Atmos. Environ. 45, 6724–6731.

Crounse, J.D., DeCarlo, P.F., Blake, D.R., Emmons, L.K., Campos, T.L., Apel, E.C., Clarke, A.D., Weinheimer, A.J., McCabe, D.C., Yokelson, R.J., Jimenez, J.L., Wennberg, P.O., 2009. Biomass burning and urban air pollution over the central Mexican Plateau. Atmos. Chem. Phys. 9, 4929–4944.

Curier, R.L., Veefkind, J.P., Braak, R., Veihelmann, B., Torres, O., de Leeuw, G., 2008. Retrieval of aerosol optical properties from OMI radiances using a multiwavelength algorithm: application to Western Europe. J. Geophys. Res. 1134, 2008. http://dx.doi.org/10.1029/2007JD008738.

Dubovik, O., King, M.D., 2000. A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements. J. Geophys. Res. 105, 20673–20696.

Dubovik, O., Smirnov, A., Holben, B.N., King, M.D., Kaufman, Y.J., Eck, T.F., Slutsker, I., 2000. Accuracy assessment of aerosol optical properties retrieved from AERONET sun and sky radiance measurements. J. Geophys. Res. 103, 865.

Dubovik, O., Holben, B.N., Eck, T.F., Smirnov, A., Kaufman, Y.J., King, M.D., Tanre, D., Slutsker, I., 2002. Variability of absorption and optical properties of key aerosol types observed in worldwide locations. J. Atmos. Sci. 59, 590–608.

Dubovik, O., Sinyuk, A., Lapyonuk, T., Holben, B.N., Mishchenko, M., Ping, Y., Eck, T.F., Volten, H., Mufioz, O., Velhelmann, B., van der Zande, W.J., Leon, J.F., Sorokin, M., Slutsker, I., 2006. Application of spheroid models to account for aerosol particle nonsphericity in remote sensing of desert dust. J. Geophys. Res. 111, D11208. http://dx.doi.org/10.1029/2005JD006619.

Eck, T.F., Holben, B.N., Boonjawat, J., Snidvongs, A., Le, H.V., Schafer, J.S., Kaewkonga, T., Mongkolnavin, R., Reid, J.S., Dubovik, O., Smirnov, A., 2004. Aerosol Optical Properties in Southeast Asia from AERONET Observations. Powerpoint presentation found at: aeronet.gsfc.nasa.gov/THAILAND/AERONET.pdf.

Eck, T.F., Holben, B.N., Reid, J.S., Mukelabai, M.M., Piketh, S.J., Torres, O., Jethva, H.T., Hyer, E.J., Ward, D.E., Dubovik, O., Sinyuk, A., Schafer, J.S., Giles, D.M., Sorokin, M., Smirnov, A., Slutsker, I., 2013. A seasonal trend of single scattering albedo in southern African biomass-burning particles: implications for satellite products and estimates of emissions for the world's largest biomass-burning source. J. Geophys. Res. 118 http://dx.doi.org/10.1002/jgrd.50500.

Fast, J.D., de Foy, B., Acevedo Rosas, F., Caetano, E., Carmichael, G., Emmons, L., McKenna, D., Mena, M., Skamarock, W., Tie, X., Coulter, R.L., Barnard, J.C., Wiedinmyer, C., Madronich, S., 2007. A meteorological overview of the MILAGRO field campaigns. Atmos. Chem. Phys. 7, 2233–2257.

Giles, D.M., Holben, B.N., Tripathi, S.N., Eck, T.F., Newcomb, W.W., Slutsker, I., Dickerson, R.R., Thompson, A.M., Mattoo, S., Wang, S.-H., Singh, R.P., Sinyuk, A., Schafer, J.S., 2011. Aerosol properties over the Indo-Gangetic plain: a mesoscale perspective from the TIGERZ experiment. J. Geophys. Res. 116, D18203. http://dx.doi.org/10.1029/2011JD015809.

Giles, D.M., Holben, B.N., Eck, T.F., Sinyuk, A., Smirnov, A., Slutsker, I., Dickenson, R.R., Thompson, A.M., Schafer, J.S., 2012. An analysis of AERONET aerosol absorption properties and classifications representative of aerosol source regions. J. Geophys. Res. 117, D17203. http://dx.doi.org/10.1029/2012JD018127.

Holben, B.N., Eck, T.F., Slutsker, I., Tanre, D., Buis, J.P., Setzer, A., Vermote, E., Reagan, J.A., Kaufman, Y., Nakajima, T., Lavenu, F., Jankowiak, I., Smirnov, A., 1998. AERONET – a federated instrument network and data archive for aerosol characterization. Rem. Sens. Environ. 66, 1–16.

Holben, B.N., Eck, T.F., Slutsker, I., Smirnov, A., Sinyuk, A., Schafer, J., Giles, D., Dubovik, O., 2006. AERONET's version 2.0 quality assurance criteria. In: Tsay, S.-C., et al (Eds.), Remote Sensing of the Atmosphere and Clouds, Proc. SPIE, 6408, p. 64080Q. http://dx.doi.org/10.1117/12.706524.

Huang, J., Hsu, N.C., Tsay, S.C., Holben, B.N., Welton, E.J., Smirnov, A., Jeong, M.-J., Hansell, R.A., Berkoff, T.A., Liu, Z., Liu, G.-R., Campbell, J.R., Liew, S.C., Barnes, J.E., 2012. Evaluations of cirrus contamination and screening in ground aerosol

observations using collocated lidar systems. J. Geophys. Res. 117 http://dx.doi.org/10.1029/2012JD017757.
IPCC, 2013. Climate change 2013: the physical science basis. In: Stocker, T.F., Qin, D., Plattner, G.-K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, B., Midgley, P.M. (Eds.), Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, p. 1535. http://dx.doi.org/10.1017/CB09781107415324.
Johnson, B.T., Christopher, S., Hayward, J.M., Osborne, S.R., McFarlane, S., Hsu, C., Salustro, C., Kahn, R., 2009. Measurements of aerosol properties from aircraft and ground based remote sensing: a case study from the Dust and Biomass-burning Experiment (DABEX). Quart. J. R. Meteorol. Soc. 135 http://dx.doi.org/10.1002/qj420.
Kacenelenbogen, M., Leon, J.-F., Chiapello, I., Tanre, D., 2006. Characterization of aerosol pollution events in France using ground based and POLDER-2 satellite data. Atmos. Chem. Phys. 6, 4843–4849.
Kahn, R., Banerjee, P., McDonald, D., 2001. Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean. J. Geophys. Res. 106, 18219–18238.
Kim, D., Chin, M., Yu, H., Eck, T.F., Sinyuk, A., Smirnov, A., Holben, B.N., 2011. Dust optical properties over North Africa and arabian Peninsula derived from the AERONET dataset. Atmos. Chem. Phys. 11, 10733–10741.
Kinne, S., Schultz, M., Textor, C., Guibert, S., Balkanski, Y., Bauer, S.E., Bernsten, T., Berglen, T.F., Boucher, O., Chin, M., Collins, W., Dentener, F., Diehl, T., Easter, R., Feichter, J., Fillmore, D., Ghan, S., Ginoux, P., Gong, S., Grini, A., Hendricks, J., Herzog, M., Horowitz, L., Isaksen, I., Iversen, T., Kirkevag, A., Kloster, S., Koch, D., Kristjansson, J.E., Krol, M., Lauer, A., Lamarque, J.F., Lesins, G., Liu, X., Lohmann, U., Montanaro, V., Myhre, V., Penner, J.E., Pitari, G., Reddy, S., Sesland, O., Stier, P., Takemura, T., Tie, X., 2006. An AeroCom initial assessment – optical properties in aerosol component modules of global models. Atmos. Chem. Phys. 6, 1815–1834.
Leahy, L.V., Anderson, T.L., Eck, T.F., Bergstrom, R.W., 2007. A synthesis of single scattering albedo of biomass burning aerosol over southern Africa during SAFARI 2000. Geophys. Res. Lett. 34, L12814. http://dx.doi.org/10.1029/2007GL029697.
Lee, J., Kim, J., Song, C.H., Kim, S.B., Chun, Y., Sohn, B.J., Holben, B.N., 2010a. Characteristics of aerosol types from AERONET sun photometer measurements. Atmos. Environ. 44, 3110–3117.
Lee, T., Sullivan, A.P., Mack, L., Jimenez, J.L., Kreidenweis, S.M., Onasch, T.B., Worsnop, D.R., Malm, W., Wold, C.E., Hao, W.M., Collett, J.L., 2010b. Chemical smoke marker emissions during flaming and smoldering phases of laboratory open burning of wildland fuels. Aerosol Sci. Technol. 44 http://dx.doi.org/10.1080/02786826.2010.499884.
Levy, R.C., Remer, L.A., Mattoo, S., Vermote, E.F., Kaufman, Y.J., 2007. Second-generation operational algorithm: retrieval of aerosol properties over land from inversion of Moderate Resolution Imaging Spectroradiometer spectral reflectance. J. Geophys. Res. 112, 3428.
Mahalanobis, P.C., 1936. On the generalized distance in statistics. Proc. Natl. Inst. Sci. India 2, 49–55.
Martins, J.V., Artaxo, P., Liousse, C., Reid, J.S., Hobbs, P.V., Kaufman, Y.J., 1998. Effects of black carbon content, particle size, and mixing on light absorption by aerosols from biomass burning in Brazil. J. Geophys. Res. 103, 32,041–32,050. http://dx.doi.org/10.1029/98JD02593.
Mielonen, T., Levy, R.C., Aaltonen, V., Komppula, M., de Leeuw, G., Huttunen, J., Lihavainen, H., Kolmonen, P., Lehtinen, K.E.J., Arola, A., 2010. Evaluating the assumptions of surface reflectance and aerosol type selection within the MODIS aerosol retrieval over land: the problem of dust type selection. Atmos. Meas. Tech. Discuss. 3, 3425–3453.
Muller, D., Weinzierl, B., Petzold, A., Kandler, K., Ansmann, A., Muller, T., Tesche, M., Freudenthaler, V., Esselborn, M., Heese, B., Althausen, D., Schladitz, A., Otto, S., Knippertz, P., 2010. Mineral dust observed with AERONET sun photometer, Raman lidar, and in situ instruments during SAMUM 2006: shape-independent particle properties. J. Geophys. Res. 115, D07202. http://dx.doi.org/10.1029/2009JD012520.
Omar, A.H., Won, J.-G., Winker, D.M., Yoon, S.-C., Dubovick, O., McCormick, M.P., 2005. Development of global aerosol models using cluster analysis of Aeronet Robotic Network (AERONET) measurements. J. Geophys. Res. 110 http://dx.doi.org/10.1029/2004JD004874.
Omar, A.H., Winker, D.M., Kittaka, C., Vaughan, M.A., Liu, Z., Hu, Y., Trepte, C., Rogers, R.R., Ferrare, R.A., Lee, K.-P., Kuehn, R.E., Hostetler, C.A., 2009. The CALIPSO automated aerosol classification and lidar ratio selection algorithm. J. Atmos. Ocean. Technol. 26, 1994–2014.
Omar, A.H., Winker, D.M., Tackett, J.L., Giles, D.M., Kar, J., Liu, Z., Vaughan, M.A., Powell, K.A., Trepte, C.R., 2013. CALIOP and AERONET aerosol optical depth comparisons: one size fits none. J. Geophys. Res. 118, 4748–4766.
Pace, G., diSarra, A., Meloni, D., Piacentino, S., Chamard, P., 2006. Aerosol optical properties at Lampedusa (Central Mediterranean). 1. Influence of transport and identification of different aerosol types. Atmos. Chem. Phys. 6, 697–713.
Patadia, F., Kahn, R.A., Limbacher, J.A., Burton, S.A., Ferrare, R.A., Hostetler, C.A., Hair, J.W., 2013. Aerosol airmass type mapping over Urban Mexico City region from space-based multi-angle imaging. Atmos. Chem. Phys. 13, 9525–9541. http://dx.doi.org/10.5194/acp-13-9525-2013.
Perry, K.D., Cahill, T.A., Eldred, R.A., Dutcher, D.D., 1997. Long-range transport of North African dust to the eastern United States. J. Geophys. Res. 102, 11225–11238.
Reid, J.S., Hyer, E.J., Johnson, R.S., Holben, B.N., Yokelson, R.J., Zhang, J., Campbell, J.R., Christopher, S.A., Di Girolamo, L., Giglio, L., Holz, R.E., Kearney, C., Miettinen, J., Reid, E.A., Turk, F.J., Wang, J., Xian, P., Zhao, G., Balasubramanian, R., Chew, B.N., Janjai, S., Lagrosas, N., Lestari, P., Lin, N.-H., Mahmud, M., Nguyen, A.X., Norris, B., Oanh, N.T.K., Oo, M., Salinas, S.V., Welton, E.J., Liew, S.C., 2013. Observing and understanding the Southeast Asian aerosol system by remote sensing: an initial review and analysis for the Seven Southeast Asian Studies (7SEAS) program. Atmos. Res. 122, 403–468.
Remer, L.A., Didier, T., Kaufman, Y.J., 2004. Algorithm for Remote Sensing of Tropospheric Aerosol from MODIS. Collection 005 (MODIS ATDB) Product ID: MOD04/MYD04.
Remer, L.A., Kleidman, R.G., Levy, R.C., Kaufman, Y.J., Tanre, D., Mattoo, S., Martins, J.V., Ichoku, C., Koren, I., Yu, H., Holben, B.N., 2008. Global aerosol climatology from the MODIS satellite sensors. J. Geophys. Res. 113 http://dx.doi.org/10.1029/2007JD009661.
Russell, P.B., Bergstrom, R.W., Shinozuka, Y., Clarke, A.D., DeCarlo, P.F., Jimenez, J.L., Livingston, J.M., Redemann, J., Dubovik, O., Strawa, A., 2010. Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition. Atmos. Chem. Phys. 10, 1155–1169.
Russell, P., Kacenelenbogen, M., Hamill, P., Livingston, J., Burton, S., Schuster, G., Hasekamp, O., Knobelspiesse, K., Shinozuka, Y., Redemann, J., Ramachandran, S., Holben, B., Dec. 2012. Aerosol Classification Using Multiparameter Retrievals from Remote Measurements on Space and Other Platforms. Poster A13K-0329, Am. Geophys. Union, Fall Meeting, San Francisco.
Russell, P.B., Kacenelenbogen, M., Livingston, J.M., Hasekamp, O.P., Burton, S.P., Schuste, G.L., Johnson, M.S., Knobelspiesse, K.D., Redemann, J., Ramachandran, S., Holben, B., 2014. A multi-parameter aerosol classification method and its application to retrievals from spaceborne polarimetry. J. Geophys. Res. Atmos. 119, 9838–9863. http://dx.doi.org/10.1002/jgrd.v.119.16.
Sawamura, P., Müller, D., Ho, R.M., Hostetler, C.A., Ferrare, R.A., Hair, J.W., Rogers, R.R., Anderson, B.E., Ziemba, L.D., Beyersdorf, A.J., Thornhill, K.L., Winstead, E.L., Holben, B.N., 2014. Aerosol optical and microphysical retrievals from a hybrid multiwavelength lidar dataset DISCOVER-AQ 2011. Atmos. Meas. Tech. Discuss. 7, 3113–3157.
Sayer, A.M., Smirnov, A., Hsu, N.C., Holben, B.N., 2012. A pure marine aerosol model for use in remote sensing applications. J. Geophys. Res. 117 http://dx.doi.org/10.1029/2011JD016689.
Schuster, G.L., Dubovik, O., Holben, B.N., Clothiaux, E.E., 2005. Inferring black carbon content and specific absorption from AERONET aerosol retrievals. J. Geophys. Res. 110 http://dx.doi.org/10.1029/2004JD004548.
See, S.W., Balasubramanian, R., Wang, W., 2006. A study of the physical, chemical, and optical properties of ambient aerosol particles in Southeast Asia during hazy and nonhazy days. J. Geophys. Res. 111 http://dx.doi.org/10.1029/2005JD006180.
Smirnov, A., Holben, B.N., Eck, T.F., Dubovnik, O., Slutsker, I., 2000. Cloud screening and quality control algorithms for the AERONET database. Remote Sens. Environ. 73, 337–349.
Sokolik, I.N., Toon, O.B., 1999. Incorporation of mineralogical composition into models of the radiative properties of mineral aerosol from UV to IR wavelengths. J. Geophys. Res. 104, 9423–9444.
Sun, Y.L., Wang, Z.F., Fu, P.Q., Yang, T., Jiang, Q., Dong, H.B., Li, J., Jia, J.J., 2013. Aerosol composition, sources and processes during wintertime in Beijing, China. Atmos. Chem. Phys. 13, 4577–4592.
Taylor, M., Kazadzis, S., Amiridis, V., Kahn, R.A., 2015. Global aerosol mixtures and their multiyear and seasonal characteristics. Atmos. Environ. 116, 112–129.
Toledano, C., Wiegner, M., Gross, S., Freudenthaler, V., Gasteiger, J., Muller, D., Muller, T., Schladitz, A., Weinzierl, B., Torres, B., O'Neill, N.T., 2011. Optical properties of aerosol mixtures derived from sun-sky radiometry during SAMUM-2. Tellus, Ser. B 63, 635–648. http://dx.doi.org/10.1111/j.1600-0889.2011.00573.
Twomey, S., 1977. Atmospheric Aerosols. Elsevier Scientific Pub. Co, Amsterdam, p. 14.
Wang, Y., Zhuang, G., Tang, A., Yuan, H., Sun, Y., Chen, S., Zheng, A., 2005. The ion chemistry and the source of PM 2.5 aerosol in Beijing. Atmos. Environ. 39, 3771–3784.
Yokelson, R., Urbanski, S., Atlas, E., Toohey, D., Alvarado, E., Crounse, J., Wennberg, P., Fisher, M., Wold, C., Campos, T., Adachi, K., Busek, P.R., Hao, W.M., 2007. Emissions from forest fires near Mexico City. Atmos. Chem. Phys. Discuss. 7, 6687–6718.
Zhao, X., Zhang, X., Pu, W., Meng, W., Xu, X., 2011. Scattering properties of the atmospheric aerosol in Beijing, China. Atmos. Res. 101, 799–808.