



Aerosol Classification Using Multiparameter Retrievals from Remote Measurements on Space and Other Platforms

P. Russell¹, M. Kacenelenbogen², P. Hamill³, J. Livingston⁴, S. Burton⁵, G. Schuster⁵, O. Hasekamp⁶, K. Knobelspiesse^{7,1}, Y. Shinzuka^{1,2}, J. Redemann¹, S. Ramachandran⁸, B. Holben⁹

¹NASA Ames Research Center, Moffett Field, CA (Philip.B.Russell@nasa.gov), ²Bay Area Environmental Research Institute, Sonoma, CA, ³San Jose State University, San Jose, CA, ⁴SRI International, Menlo Park, CA, ⁵NASA Langley Research Center, Hampton, VA, ⁶SRON Netherlands Institute for Space Research, Utrecht, Netherlands, ⁷NASA Goddard Institute for Space Studies, New York, NY, ⁸Physical Research Laboratory, Ahmedabad, India, ⁹NASA Goddard Space Flight Center, Greenbelt, MD

Abstract

Over the past several decades, since the development of truly global aerosol measurements by satellites and AERONET, it has proven useful to classify observed aerosols into several major types (e.g., urban-industrial, biomass burning, mineral dust, maritime, and various subtypes or mixtures of these). Such classification can help to understand aerosol sources, effects, and feedback mechanisms, to improve accuracy of satellite retrievals and to quantify assessments of aerosol radiative impacts on climate. With ongoing improvements in measurement capability, the number of aerosol parameters retrieved from spaceborne and suborbital remote sensors has been growing, from the initial aerosol optical depth at one or a few wavelengths to an extensive list from a variety of sensors such as the AERONET surface-based sun-sky photometers, the PARASOL spaceborne polarimeter, the RSP airborne polarimeter, the new 4STAR airborne sun-sky spectrometer, and combinations of data from these sensors with data from the OMI spaceborne spectrograph, or vertically resolved data from the CALIOP spaceborne lidar and the HSRL airborne lidar. Examples of the various data products available from these sensors include complex refractive index, single scattering albedo (SSA), and depolarization of backscatter, each at several wavelengths; wavelength dependences of extinction, scattering, absorption, SSA, and backscatter; and several particle size and shape parameters.

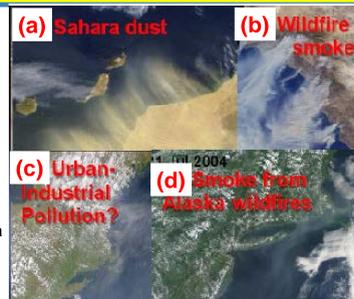
Making optimal use of these varied data products requires objective, multi-dimensional analysis methods. Our methods, analogous to those used with HSRL data, use Mahalanobis distance to assign any given multidimensional (i.e., multiparameter) observation to the pre-specified cluster (aerosol type) to which it is most similar. The pre-specified clusters are defined using AERONET-retrieved parameters for sites and months where a specific type of aerosol is known to dominate. These aerosol types include urban industrial (for both developed and developing economies), biomass burning (for both Amazonian and African Savanna smoke), mineral dust, and marine. Dimensions we currently use include extinction angstrom exponent, absorption angstrom exponent, single scattering albedo and its wavelength dependence, real and imaginary refractive index, and indicators of particle size and sphericity. We show example results including seasonal changes of aerosol type at AERONET stations and identification of aerosol type from PARASOL-retrieved aerosol parameters.

1. Background and goal

In some conditions aerosol type can be identified in imagery from space by tracing the aerosol back to its source (e.g., the individual plumes in Fig. 1a,c). In other cases (e.g., Fig. 1c) it is tempting to guess aerosol type based on aerosol location. However, this can lead to errors, as exemplified by Fig. 1d, in which Alaskan wildfire smoke, carried down the Mississippi Valley, along the Gulf Coast and up the Atlantic seaboard, caused a haze layer off New England, an area typically impacted by urban-industrial pollution.

The goal of this research is to develop robust methods for identifying aerosol type from the opto-physical information retrievable from an individual image pixel or group of pixels used in a retrieval (see, e.g., the pixel groupings used by the PARASOL retrieval of Dubovik et al. (2011)).

Fig. 1. (a, b) Plumes of Sahara dust and wildfire smoke in MODIS imagery. (c) Image of a large-scale haze over the eastern US and western Atlantic. (d) Image of a large-scale haze over the same area as (c), which was traceable back to Alaska wildfires.

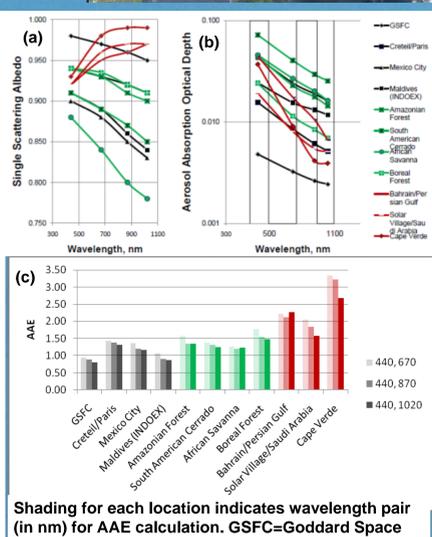


2. Previous results

In Russell et al. (2010) we showed that correlations between aerosol type and aerosol optical parameters, which had previously been noted via radiometric measurements of aerosol layers (e.g., Bergstrom et al., 2007), and via in situ measurements of aerosol volumes (e.g., Shinzuka et al., 2009), were also present in AERONET-retrieved parameters describing full aerosol vertical columns (as represented by the AERONET pre-Version 1 data in Dubovik et al. (2002)). In particular, as illustrated in Fig. 2,

- SSA spectra from three desert dust sites (red curves in Fig. 2a) have slopes opposite to those for four urban-industrial and four biomass burning sites (black and green curves).
- Despite the variety of SSA spectral shapes in Fig. 2a, the corresponding curves in Fig. 2b of aerosol absorption optical depth (AOD) are all nearly straight lines in the log-log plot. In other words, they have nearly constant absorption Angstrom exponent (AAE)
- AAE is correlated with aerosol type (Fig. 2c), with AAE values near 1 (the theoretical value for black carbon) for AERONET-measured aerosol columns dominated by urban-industrial aerosol (black bars in Fig. 2c), larger AAE values for biomass burning aerosols (green bars in Fig. 2c), and the largest AAE values for Sahara dust aerosols (red bars).

Fig. 2. (a) Spectra of AERONET-derived Single Scattering Albedo (SSA) from specific sites and seasons identified by Dubovik et al. (2002) as particular types. Black: Urban/Industrial or Mixed; Green: Biomass Burning; Red-Brown: Desert Dust. (b) Corresponding Aerosol Absorption Optical Depth (AOD) spectra. (c) Absorption Angstrom Exponent (AAE) values for the AOD spectra in (b).



In Russell et al. (2010) we also noted the overlap in AAE for some urban-industrial and biomass-burning sites (black and green bars in Fig. 2c), indicating that aerosol classification using AAE alone could lead to ambiguity. We showed that using a two-dimensional plot, of AAE vs EAE, reduced the ambiguity but did not eliminate it. And we suggested the use of other retrieved aerosol parameters in multidimensional analyses as a potential way to reduce remaining ambiguity.

References

Bergstrom, R. W., et al., Spectral absorption properties of atmospheric aerosols, *Atmos. Chem. Phys.*, 7, 5937-5943, 2007.
 Burton, S. P., et al., Aerosol classification using airborne High Spectral Resolution Lidar measurements – methodology and examples, *Atmos. Meas. Tech.*, 5, 73-98, 2012.
 Catrall, C., et al., Variability of ... aerosol types..., *J. Geophys. Res.*, 110, D10511, doi:10.1029/2004JD005124, 2005.
 Hasekamp, O. P., et al. (2011), Aerosol properties over the ocean from PARASOL multiple photopolarimetric measurements, *J. Geophys. Res.*, 116, D14204, doi:10.1029/2010JD015469.
 Dubovik, O., et al.: Variability of absorption and optical properties of key aerosol types observed in worldwide locations, *J. Atmos. Sci.*, 59, 590-608, 2002.
 Dubovik, O., et al. Statistically optimized inversion algorithm for enhanced retrieval of aerosol properties..., *Atmos. Meas. Tech.*, 4, 975-1018, 2011.
 Giles, D. M., et al. (2012), An analysis of AERONET aerosol absorption properties and classifications..., *J. Geophys. Res.*, 117, D17203, doi:10.1029/2012JD018127.
 Hostetler, C. A., et al., Airborne Multi-wavelength High Spectral Resolution Lidar..., A13K-0336, Fall Meeting, Amer. Geophys. Union, San Francisco, 3 December 2012.
 Mahalanobis, P. C. (1936), On the Generalized Distance in Statistics, *Proceedings of the National Institute of Sciences in India*, 2, 49-55.
 Mousiadis, L., and A. Vakali, Benchmark graphs for the evaluation of clustering algorithms, 3rd Int'l Conf. Research Challenges in Information Science, Fez, Morocco, DOI 10.1109/RCIS.2009.5089283, 22-24 April 2009.
 Remer, L. A., et al. (2005), The MODIS aerosol algorithm, products, and validation, *J. Atmos. Sci.*, 62, 947-973, doi:10.1175/JAS3385.1
 Russell, P. B., et al., Absorption Angstrom Exponent in AERONET and related data as an indicator of aerosol composition, *Atmos. Chem. Phys.*, 10, 1155-1169, 2010.
 Sayer, A. M., et al. (2012), A pure marine aerosol model, for use in remote sensing applications, *J. Geophys. Res.*, 117, D05213, doi:10.1029/2011JD016689.
 Shinzuka, Y., et al., Aerosol optical properties relevant to regional remote sensing of CCN activity and links to their organic mass fraction: airborne observations over Central Mexico and the US West Coast during MILAGRO/INTEX-B, *Atmos. Chem. Phys.*, 9, 6727-6742, 2009.
 Wikipedia, Mahalanobis distance, http://en.wikipedia.org/wiki/Mahalanobis_distance, 2010.

3. Examples of parameters available

To illustrate the variety of parameters available, Table 1 lists examples of aerosol data products produced by selected spaceborne, airborne, and surface-based sensors. To save space, Table 1 focuses on sensors or combinations that produce or promise more aerosol parameters than the MODIS or MISR operational sets, although MODIS and MISR have supported very useful aerosol classification studies with their extensive, well-documented, and validated data sets.

Aerosol Property	Passive Sensors				Active Sensors			
	AERONET Version 2 or 4STAR	PARASOL polarimeter (Dubovik et al. (2011))	Hasekamp et al. (2011)	Glory APS	RSP	OMI	HSRL-2 Hostetler et al. (2012)	CALIOP
AOD	7 λ, 340-1020 nm	6 λ, 440-1020 nm	3 λ, 490-865 nm	≥3 λ, typical 410-865 nm	9 λ, 410-2250 nm	388 nm	355, 532 nm	532 nm
SSA	4 λ, 440-1020 nm	6 λ, 440-1020 nm	3 λ, 490-865 nm	≥3 λ, typical 410-865 nm	From RRI, IRI, size dist.	388 nm		
RRI	4 λ, 440-1020 nm	6 λ, 440-1020 nm	λ-indep, 490-865 nm	≥3 λ, typical 410-865 nm	λ-indep, 410-1590 nm			
IRI	4 λ, 440-1020 nm	6 λ, 440-1020 nm	λ-indep, 490-865 nm	≥3 λ, typical 410-865 nm	λ-indep, 410-1590 nm			
Size parameters	r _v , σ _v ^{1,2} r(λ)	r _v , σ _v ^{1,3}	r _v , V _v ¹	r _v , V _v ¹	r _v , V _v ¹			
Column amount	C _v (μm ³ /μm ²)	C _v (μm ³ /μm ²)	N (cm ⁻²)		N(μm ³)			
Shape indicator	% spheres	% spheres	% spheres ²				Depolarization (355, 532, 1064 nm)	Depolarization (532 nm)
Layer height								
Backscatter coef. B							355, 532, 1064 nm	532, 1064 nm
Extinction coef. E							355, 532 nm	
S _a							355, 532 nm	

From the basic parameters in Table 1, other parameters can be derived, some of which can be very useful for aerosol classification. Examples, for wavelengths λ₁, include:
 Extinction Angstrom Exponent: $EAE_{\lambda_1, \lambda_2} = -\ln(AOD_{\lambda_1}/AOD_{\lambda_2})/\ln(\lambda_1/\lambda_2)$, (1)
 Aerosol Absorption Optical Depth: $AAOD_{\lambda} = [1-SSA_{\lambda}] AOD_{\lambda}$, (2)
 Absorption Angstrom Exponent: $AAE_{\lambda_1, \lambda_2} = -\ln(AAOD_{\lambda_1}/AAOD_{\lambda_2})/\ln(\lambda_1/\lambda_2)$, (3)

4. Multidimensional specified clustering and Mahalanobis classification

Specified clustering (e.g., Mousiadis and Vakali, 2009) uses "a priori" information in a "reference" data set to assign points to clusters. This a priori information can include information (e.g., trajectory or chemical analyses or previous studies) beyond the optical parameters that will be available to the classification method in the general case. In Fig. 3, a 2-dimensional scatterplot of AAE_{440,870} vs EAE_{440,870}, we have assigned AERONET Version 2 Level 2.0 retrieved data points to clusters (symbol colors) using the aerosol type designations of Dubovik et al. (2002) or Catrall et al. (2005), which specify months during which certain aerosol types tend to dominate at certain sites. This is an example of specified clustering.

Mahalanobis classification (Mahalanobis, 1936; Wikipedia, 2010; Burton et al., 2012) assigns any given N-dimensional point (x₁, x₂, ..., x_N) to the cluster that has minimum Mahalanobis distance, D_M, from that point. For purposes of defining D_M, a cluster is defined by its mean (μ₁, μ₂, ..., μ_N) and its covariance matrix S. In N dimensions, where the elements of S are given by

$$D_M(x) = \sqrt{(x - \mu)^T S^{-1} (x - \mu)} \quad (4)$$

where the elements of S are given by

$$S_{ij} = \rho_{ij} \sigma_i \sigma_j \quad (5)$$

σ_i is the standard deviation of x_i for all points in the cluster, and ρ_{ij} is the correlation coefficient of x_i and x_j. In 2 dimensions, curves of constant D_M are ellipses, several of which are shown in Fig. 3. When points in a class are multi-normally distributed, D_M² belongs to a chi-square distribution; hence, the probability P(D_M) of a point being closer than D_M to the cluster mean can be easily obtained from standard chi-square tables. The legend of Fig. 3 shows the correspondence between D_M and P(D_M) for points and clusters with 2 dimensions.

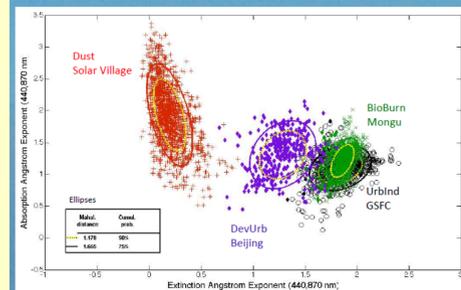


Fig. 3. A 2-dimensional scatterplot of data from AERONET Version 2 retrievals at sites/months designated by Dubovik et al. (2002) or Catrall et al. (2005) to be dominated by certain aerosol types. Abbreviated names of classes/ specified clusters (e.g., "DevUrb") are defined in Table 3.

5. Aerosol Parameters and Classes Used in Our Current Classification Method

We are currently using 8 parameters and 6 aerosol classes (specified clusters), as listed in Tables 2 and 3 and illustrated in Figs. 4 and 5. The 8 parameters were chosen to help improve separation of classes, as measured by values of Wilks' overall lambda statistic, both for all 6 classes and for the UrbInd and two BioBurn classes, which are particularly difficult to separate in many parameters (including the EAE_{440,870} and AAE_{440,870} of Fig. 3).

Parameter name	Wilks' partial lambda ²		For all Table 3 classes excluding Marine		For the UrbInd & 2 BioBurn classes	
	For all	For the				
EAE _{440,870}	0.660	0.963				
AAE _{440,870}	0.768	0.867				
SSA ₄₄₀	0.692	0.789				
RR1 ₆₇₀	0.811	0.705				
dSSA _{440,870} (=SSA ₄₄₀ -SSA ₈₇₀)	0.717	0.797				
Vol/Vol	0.945	0.957				
%Spheres	0.041	0.989				
IRI ₆₇₀	0.906	0.877				

*Smaller values indicate greater resolving power.

Class name	Abbreviations	Included sites, numbers of data points, periods
Urban-Industrial, developed economy	UrbInd	GSFC, MD, US (N=836); GISS, NY, US (N=51); Creteil, FR (N=29); Aire Adour, FR (N=9); Lille, FR (N=136). All Jun-Sep 1992-2009; US group & FR group given equal weight.
Urban-Industrial, developing economy	DevUrb	Beijing, CN with %Sphere>90 (to filter out cases with dust present) (N=406, Jun-Feb, 2000-2012)
Biomass Burning, Amazon Basin	BioBurn-Ama, BB-Ama	Alta Floresta, BR (N=501), Los Fieros, BO (N=137). All Aug-Oct 1995-2008
Biomass Burning, African Savanna	BioBurn-Afr, BB-Afr	Mongu, ZM (N=1513), Senanga, ZM (N=164), Zambezi, ZM (N=213) All Aug-Nov 1995-2008.
Mineral Dust	Dust	Solar Village, SA (N=1573), Mar-Jul 1992-2008.
Pure Marine	Marine, Mari	Lana'i, HI, US (N=615, 20 Mar 1999 - 13 Mar 2010), filtered with Sayer et al. (2012) criteria.

Summary and Conclusions

- Specified clustering and Mahalanobis classification together provide a useful way of combining several dimensions of opto-physical information to assign aerosols to classes.
- The Wilks' lambda statistics (overall and partial) are useful in judging the separability of aerosol classes, and which dimensions (parameters) contribute most to class separability.
- We have applied these methods to AERONET Version 2 data sets using the parameters: extinction angstrom exponent (EAE), absorption angstrom exponent (AAE), single scattering albedo (SSA), its wavelength dependence (dSSA), real and imaginary refractive index (RRI, IRI), and indicators of particle size (Vol_v/Vol_l) and shape (%Spheres).
- Using these 8 AERONET-retrieved parameters, we assign aerosols to 6 classes: UrbInd, DevUrb (urban-industrial for developed and developing economies), BioBurn-Ama, BioBurn-Afr (biomass burning for Amazonian (less absorbing) and African Savanna (more absorbing) smoke), mineral Dust, and Marine. This classification is useful in showing how aerosol type varies from month to month and site to site.
- In applying the method to PARASOL-retrieved parameters, retrieval uncertainty is critical in choosing which parameters and data points to include. We have used the method with PARASOL-retrieved EAE, SSA, RRI, and dSSA to assign aerosols at FORTH-Crete to the classes UrbInd, DevUrb, Dust, BioBurn-Ama, and Marine.

Next Steps

- Include other AERONET sites in the Dust cluster, to make the class more generally representative. Investigate this for the UrbInd, DevUrb, and other classes, too.
- Investigate use of fine-mode fraction of AOD(λ) as an input parameter.
- Apply the classification to a larger PARASOL data set, including larger AOD values, and hence smaller retrieval uncertainties, possibly allowing use of more parameters.

5. Aerosol Parameters and Classes Used in Our Current Classification Method (cont'd)

Fig. 4, patterned on Fig. 10 of Burton et al. (2012), shows a way of presenting data for all eight parameters (dimensions) in Table 2 and all 6 aerosol classes (specified clusters) in Table 3. It illustrates, for example, that:
 - UrbInd and DevUrb are somewhat distinct in EAE_{440,870}, but not in AAE_{440,870}.
 - UrbInd is not distinct from BioBurn-Ama and BioBurn-Afr in EAE_{440,870} or AAE_{440,870}, but is somewhat distinct from them in RRI₆₇₀ (in accord with the Wilks' partial lambda values in the right column of Table 2).
 Giles et al. (2012) show AERONET-derived plots of AAE vs EAE and SSA vs EAE that are quite similar to our Figs. 3 and 5.

Fig. 4. Left axes: Box-and-whisker plots for the 8 parameters in Table 2 and the 6 aerosol classes (specified clusters) in Table 3. Central dot: median; Box: 25th to 75th %ile; Whiskers: 5th and 95th %ile. Right axes: Number of data points in each specified cluster (class), shown as gray columns.

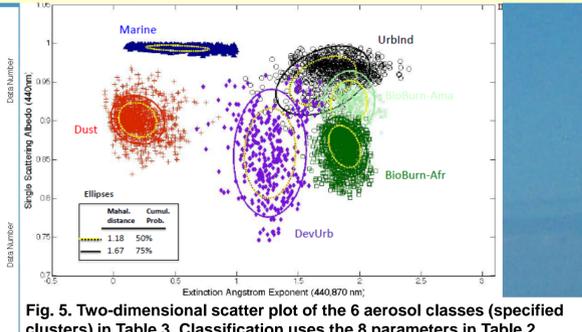
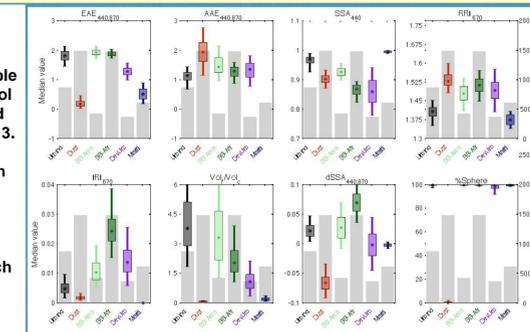


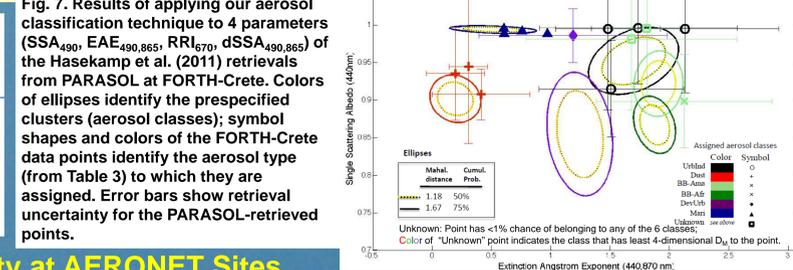
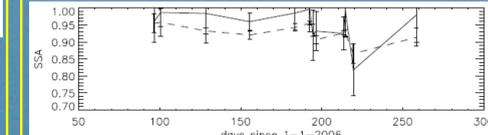
Fig. 5. Two-dimensional scatter plot of the 6 aerosol classes (specified clusters) in Table 3. Classification uses the 8 parameters in Table 2.

6. Application to PARASOL Retrieved Aerosol Parameters

Hasekamp et al. (2011) describe retrievals of aerosol properties from pixels viewed by the POLDER-3 polarimeter on the PARASOL spacecraft. Table 1 lists the properties retrieved by the Hasekamp algorithm, and Fig. 6 shows an example time series of retrieved SSA₆₇₀ values for the location FORTH-Crete, compared to measurements by AERONET. Because of its location in the Eastern Mediterranean, FORTH-Crete can experience different aerosol types (including marine, desert dust, urban-industrial, and biomass burning) at different times.

Fig. 7 shows results of applying our aerosol classification technique to a PARASOL data set from FORTH-Crete. First, points that satisfy the Sayer et al. (2012) criteria (0.1 ≤ EAE₁, AOD₂₀₀ < 0.2), are classified as Marine. Then, remaining points that pass filters on retrieval uncertainty (δSSA₄₉₀ ≤ 0.15, δEAE_{490,865} ≤ 1) are Mahalanobis-classified using 4 parameters (SSA₄₉₀, EAE_{490,865}, RRI₆₇₀, dSSA_{490,865}).

Fig. 6. Example time series of SSA₆₇₀ retrieved from PARASOL polarimeter measurements at FORTH-Crete (solid line), compared to AERONET measurements (dashed line). [From Fig. 4 of Hasekamp et al., 2011]



7. Application to Aerosol Variability at AERONET Sites

We have also applied the procedures described above to showing how aerosol type varies with time at particular AERONET sites. More specifically, for each AERONET Version 2 Level 2.0 observation, we use the 8 parameters in Table 2 to compute D_M to each of the prespecified classes in Table 3, and we assign the observation to the class from which it has least D_M. (In certain cases we exclude certain parameters as being ill-defined. For example, at island sites, for points (observations) satisfying the Sayer et al. (2012) "pure marine" filters AOD₂₀₀ < 0.2 and 0.1 ≤ EAE_{440,870} ≤ 1, we exclude AAE because absorption is too small for this parameter to be well defined.) Fig. 8 shows example results for FORTH-Crete, illustrating how aerosol types (including Marine, Dust, UrbInd, and BioBurn) are distributed over the full period 2003-2009 (in the pie chart) and by month within that period (in the column chart). These results show some commonality with our classification results from PARASOL data in Fig. 7 for FORTH-Crete, although the data periods differ (April-September 2006 for the PARASOL data in Figs. 6-7, vs. all months 2003-2009 for the AERONET data in Fig. 8).

Fig. 9 shows further examples of application of our classification procedure to a variety of AERONET sites at different global locations.

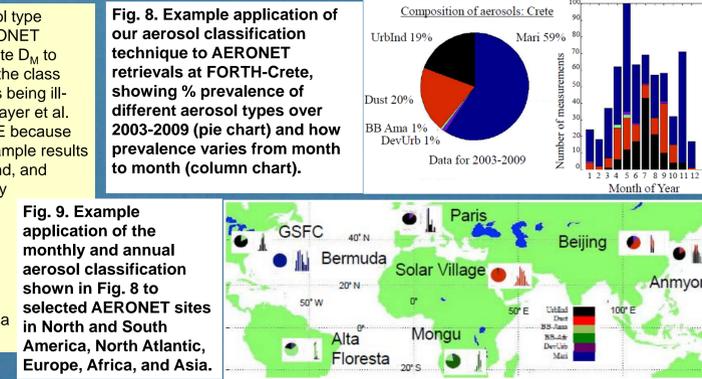


Fig. 9. Example application of the monthly and annual aerosol classification shown in Fig. 8 to selected AERONET sites in North and South America, North Atlantic, Europe, Africa, and Asia.