A21F-0113

AGU Fall 2013

Classification of Aerosol Retrievals From Spaceborne Polarimetry Using a Multiparameter Algorithm

Philip B. Russell[1], Meloe Kacenelenbogen[2], John M. Livingston[3], Otto P. Hasekamp[4], Sharon P. Burton[5], Gregory L. Schuster[5], Matthew S. Johnson[1], Kirk D. Knobelspiesse[1], Jens Redemann[1], S. Ramachandran[6], Brent Holben[7]

[1]NASA Ames Research Center, Moffett Field, CA (Philip.B.Russell@nasa.gov), [2]B ay Area Environmental Research Institute, Sonoma, CA, [3]SRI International, Menlo Park, CA, [4]SRON Netherlands Institute for Space Research, Utrecht, Netherlands, [5]NASA Langley Research Center, Hampton, VA, [6]Physical Research Laboratory, Ahmedabad, India, [7]NASA Goddard Space Flight Center, Greenbelt, MD

Abstract

Since the development of global aerosol measurements by satellites and AERONET, classification of observed aerosols into several types (e,g., urban-industrial, biomass burning, mineral dust, maritime, and various subtypes or mixtures of these) has proven useful to understanding aerosol sources, transformations, effects, and feedback mechanisms; to improving accuracy of satellite retrievals; and to quantifying assessments of aerosol radiative impacts on climate. With ongoing improvements in satellite measurement capability, the number of aerosol parameters retrieved from spaceborne sensors has been growing, from the initial aerosol optical depth at one or a few wavelengths to a list that now includes 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. We describe such a method, which uses a modified Mahalanobis distance to quantify how far a data point described by N aerosol parameters is from each of several prespecified classes. The method makes explicit use of uncertainties in input parameters, treating a point and its N-dimensional uncertainty as an extended data point or pseudo-cluster E. It then uses a modified Mahalanobis distance, D_{EC}, to assign an observation to the class (cluster) C that has minimum D_{EC} from the point (equivalently, the class to which the point has maximum probability of belonging). The method also uses Wilks' overall lambda to indicate how well the input data lend themselves to separation into classes and Wilks' partial lambda to indicate the relative discriminatory power of each parameter. We use AERONET-retrieved parameters to define 7 prespecified clusters (pure dust, polluted dust, urban-industrial/developed economy, urban-industrial/developing economy, dark biomass smoke, light biomass smoke, pure marine), and we demonstrate application of the method to a 5-year record of retrievals from the POLDER-3 polarimeter on the PARASOL spacecraft over the island of Crete, Greece. Results show changes of aerosol type at this location in the eastern Mediterranean Sea, which is influenced by a wide variety of aerosol sources.

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,b). 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 has been 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)).

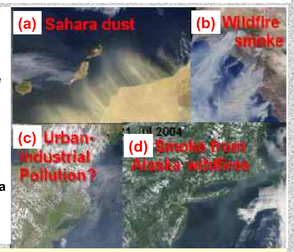


Fig. 1. (a, b) Plumes of Sahara dust and wildfire smoke in MODIS imagery. (c) Image of a large-scale haze over the eastern US and western Atlantic. (d) Image of a large-scale haze over the same area as (c), which was traceable back to Alaska wildfires.

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.

Table 1. Aerosol properties retrieved from various remote sensors (see acronyms in Table 1). λ =wavelength, r_v=volume median radius; σ_v=standard deviation; C_v=volume concentration. r_{eff}=effective radius; v_{eff}=effective variance; FMF=fine mode fraction; N=number concentration. S_a=aerosol extinction-to-backscatter ratio (i.e. lidar ratio). 4STAR is designed to produce AERONET-like retrievals from airborne sun and sky measurements.

	Passive Sensors						Active Sensors	
Aerosol Property	AERONET Version 2 or 4STAR spectro-meter	POLDER-3 on PARASOL: Dubovik et al. (2011)	POLDER-3 on PARASOL: Hasekamp et al. (2011) and updates	Glory APS Mishchenko et al. (2007)	RSP Cairns et al. (1999, 2009), Knobelspiesse et al. (2011).	OMI Jethva and Torres (2011)	HSRL-2 Hosteller et al. (2012)	CALIOP (e.g., Winker et al., 2009; Vaughan et al., 2009; Liu et al 2010)
AOD	7 λ, 340-1020 nm[1]	6 λ, 440-1020 nm	491, 670, 863,1020 nm[1]	≥3 λ, typic[1]y 410-865 nm[1]	9 λ, 410-2250 nm[1]	388 nm[5]	355, 532 nm[6]	532 nm[6]
SSA	4 λ, 440-1020 nm	6 λ, 440-1020 nm	491, 670, 863,1020 nm	≥3 λ, typic[1]y 410-865 nm[1]	From RRI, IRI, size dist.	388 nm[5]		
RRI	4 λ, 440-1020 nm	6 λ, 440-1020 nm	λ–indep, 491-1020 nm[1]	≥3 λ, typic[1]y 410-865 nm[1]	λ–indep, 410-1590 nm[1]			
IRI	4 λ, 440-1020 nm	6 λ, 440-1020 nm	λ–indep, 491-1020 nm[1]		λ–indep, 410-1590 nm[1]			
Size parameters	r_v, σ_v, r_e, $\sigma_v^{1/2}$ FMF(λ_s)	r_v, $\sigma_v^{1/3}$	r_e, $v_e^{1/37}$	r_e, v_e[4]	r_e, v_e[4]			
Column amount	C_v ($\mu m^3/\mu m^2$)[1]	C_v ($\mu m^3/\mu m^2$)[1]	N (cm^{-2})[1]		N($m\mu m^{-2}$)[1]			
Shape indicator	% spheres	% spheres	% spheres in coarse mode	✓			Depolarization (355, 532, 1064 nm)	Depolarization (532 nm)
Layer height					✓[4]		✓	✓
Backscatter coef. B							355, 532, 1064 nm	532, 1064 nm[7]
Extinction coef. E							355, 532 nm	532, 1064 nm[7]
S_a							355, 532 nm	

AOD = Aerosol optical depth.
SSA = Single scattering albedo.
RRI = Real refractive index.
IRI = Imaginary refractive index.

From the basic parameters in Table 1, other parameters can be derived, some of which can be very useful for aerosol classification. Examples include:
EAE: Extinction angstrom Exponent.
AAOD: Aerosol absorption optical depth.
AAE: Absorption angstrom exponent.

[1] For each of 2 size modes (fine and coarse).
[2]From 22-bin retrieved volume-size distribution.
[3]From s16-bin retrieved volume-size distribution.
[4]For optimizations that utilize data from the HSRL, the aerosol layer height is fixed by HSRL observations and not changed during optimization.
[5]Plus values at 354 and 500 nm from model.
[6]Vertical integral of extinction profile.
[7]From attenuated backscatter using "assumed" S_a from aerosol layer identification.
[8]For coarse mode

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.
Cattrall, C., et al., Variability of …key aerosol types..., *J. Geophys. Res.*, 110, D10S11,doi:10.1029/2004JD005124, 2005.
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.
Hasekamp, O. P., et al. (2011), Aerosol properties …from PARASOL multiangle photopolarimetric measurements, *J. Geophys. Res.*, 116, D14204, doi: 10.1029/2010JD015469.
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.
Moussiades , L., and A. Vakali, Benchmark graphs for …clustering algorithms, 3rd Int. Conf. Res. Chall. Info. Sci., Fez, Morocco, DOI 10.1109/RCIS.2009.5089283, Apr 2009.
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 et al., A pure marine aerosol model, for use in remote sensing applications, J. Geophys. Res., 117, D05213, doi:10.1029/2011JD016689.
Shinozuka, 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.

2. Aerosol classification method

This poster uses AERONET data in two ways: (1) To illustrate how Mahalanobis classification works with a variable number of parameters, and (2) To define aerosol classes (specified clusters) for use in classifying aerosols observed by the POLDER-3 polarimeter on the PARASOL spacecraft.

2.1 Examples of aerosol parameters in relation to aerosol types

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., Shinozuka 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 (AAOD) are all nearly straight lines in the log-log plot. In other words, they have nearly constant absorption Angstrom exponent (AAE). In Russell et al. (2010) we also noted the overlap in AAE values for some urban-industrial and biomass-burning sites, indicating that aerosol classification using AAE alone could lead to ambiguity.

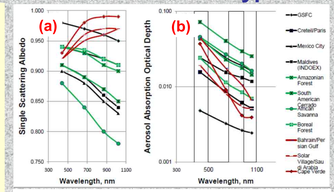


Fig. 2. (a) Spectra of AERONET-derived Single Scattering Albedo (SSA) from Dubovik et al. (2002). Black: Site/season designated by Dubovik et al. (2002) as Urban/Industrial or Mixed; Green: Analogous for Biomass Burning; Red-Brown: Desert Dust. (b) Corresponding Aerosol Absorption Optical Depth (AAOD) spectra.

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.

In the work reported here we have investigated such multidimensional clustering analyses, using both AERONET and PARASOL data. The method we have found most effective is analogous to the method described by Burton et al. (2012). We call this method Specified Clustering and Mahalanobis Classification, and we illustrate it in Fig. 3 and subsequent figures.

2.2. Specified clustering and Mahalanobis classification

Specified clustering (e.g., Moussiades 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. 3a, a 2-dimensional scatterplot of SSA_{491} vs $EAE_{491,863}$, 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 Cattrall et al. (2005), which specify months during which certain aerosol types tend to dominate at certain sites. This is an example of specified clustering. Fig. 3b, analogous to Fig. 3a, substitutes RRI_{670} for SSA_{491}, to illustrate how the relative differences and overlap between classes can change when different parameters are used. The names of aerosol clusters (equivalently, classes or types) used in Fig. 3 are described more fully in Table 2, along with the AERONET sites and periods used as input when building each class.

Mahalanobis classification (Mahalanobis, 1936; Wikipedia, 2010; Burton et al., 2012) assigns any given N-dimensional point $(x_1, x_2, \ldots, x_N)^T$ to the cluster that has minimum Mahalanobis distance, D_M, from that point. 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_M^2 belongs to a chi-square distribution; hence, the probability $P(D_M)$ of a random point from the distribution 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. 3a illustrates the benefit of using 2 rather than 1 dimension (parameter) for aerosol classification (e.g., $EAE_{491,863}$ helps separate classes that overlap in SSA_{491}). Fig. 3 also demonstrates how using different dimensions (parameters) can change the separation between classes (clusters). For example, replacing SSA_{491} by RRI_{670} in going from Fig. 3a to Fig. 3b increases the separation of BB-White from UrbInd (the corresponding 75% ellipses overlap in Fig. 3a but not in Fig. 3b), but it reduces the separation between BB-White and BB-Dark (the corresponding 75% ellipses are tangent in Fig. 3a but greatly overlapping in Fig. 3b). Use of a 3-dimensional Mahalanobis distance can benefit from the information in all parameters. And the Mahalanobis formulation has the flexibility to accommodate still more dimensions.

Fig. 3. (a) A 2-dimensional scatterplot of data from AERONET Version 2 retrievals at sites/months designated by Dubovik et al. (2002) or Cattrall et al. (2005) to be dominated by certain aerosol types. Abbreviated names of classes/specified clusters (e.g., "DevUrb") are defined in Table 2. (b) As in (a), but using the dimensions (parameters) RRI_{670} vs $EAE_{491,863}$ to illustrate how the relative differences and overlap between classes can change with different dimensions (parameters).

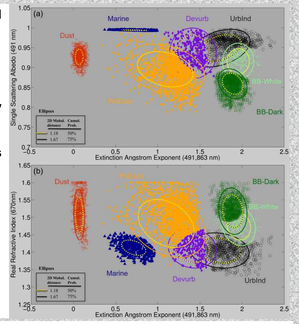


Table 2. Aerosol classes (specified clusters) currently used in our aerosol classification method

Class name	Abrevation	Included sites, numbers of data points, periods
Pure Dust	Dust	Solar Village, SA† (n=436), Capo Verde (n=56), Dakar, SE (n=143), Tamanrasset_INM, AL (n=27), Banizoumbou, NI (n=321), IER_Cinzana, MA (n=186); Total n=1169, Mar-Jul 1992-2008*
Polluted Dust	PolDust, PolD	Beijing, CN (n=1563), Cairo, EG (n=34), Bahrain, BH (n=121); Total n=1718*
Biomass Burning, Dark Smoke	BB-Dark, BB-D	Mongu, ZM (n=1513), Senanga, ZM (n=164), Zambezi, ZM (n=213) Total n=1890, All Aug-Nov 1995-2008*
Biomass Burning, White Smoke	BB-White, BB-W	Alta Floresta, BR (n=501), Los Fieros, BO (n=137). Total n=638, All Aug-Oct 1995-2008*
Urban-Industrial, developed economy	UrbInd, UrbI	GSFC, MD, US (n=836); GISS, NY, US (n=51); Creteil, FR (n=29); Aire Adour, FR (n=9); Lille, FR (n=136). Total n=1061, All Jun-Sep 1992-2009)*
Urban-Industrial, developing economy	DevUrb, DevU	Anmyon, SK (n=77), Bandung, ID (n=76), Chen-Kung, CN (n=354);Total n=507, All 2000-2012*
Pure Marine	Marine, Mari	Lanai, HI, US (n=1212, 11 Nov 1996– 3 Feb 2004). AERONET size distributions + model refractive indices per Sayer et al. (2012)

†Country abbreviations: AL=Algeria. BR=Brazil. BH=Bahrain. BO=Bolivia. CN=China. EG=Egypt. FR=France. ID= Indonesia. MA=Mali. NI=Niger. SA=Saudi Arabia. SE=Senegal. SK=South Korea. US=United States. ZM=Zambia.
To "purify" each class, points are filtered in both parameter values and Mahalanobis distance.

3. Application to POLDER-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. We applied our classification method to aerosol parameters retrieved by an updated version of the Hasekamp et al. (2011) algorithm, which uses more wavelengths and includes particle non-sphericity (though in a different way than the AERONET algorithm). Our POLDER classification uses 4 parameters, SSA_{491}, $EAE_{491,863}$, RRI_{670}, and $dSSA_{863,491}$. Polder-retrieved uncertainties are used in two ways: to filter input points ($\delta SSA_{491} \leq 0.075$, $\delta EAE_{491,863} \leq 0.6$, $\delta RRI_{670} \leq 0.1$, $\delta dSSA_{863,491} \leq \sqrt{(2)} * 0.075$) and to define a modified Mahalanobis distance, D_{EC}, that treats an N-dimensional data point and its N-dimensional error bar as a pseudo-cluster.

Fig. 4 shows results of applying our aerosol classification technique to a PARASOL data set from FORTH-Crete, an island in the Eastern Mediterranean that can experience different aerosol types at different times. Fig. 5 shows the classification results as a time series. Fig. 6 examines the case of April 20, 2008 (dashed vertical line in Fig. 5) by showing the many ancillary data sets that are consistent with our POLDER classification result (red color indicates it is closest to Pure Dust).

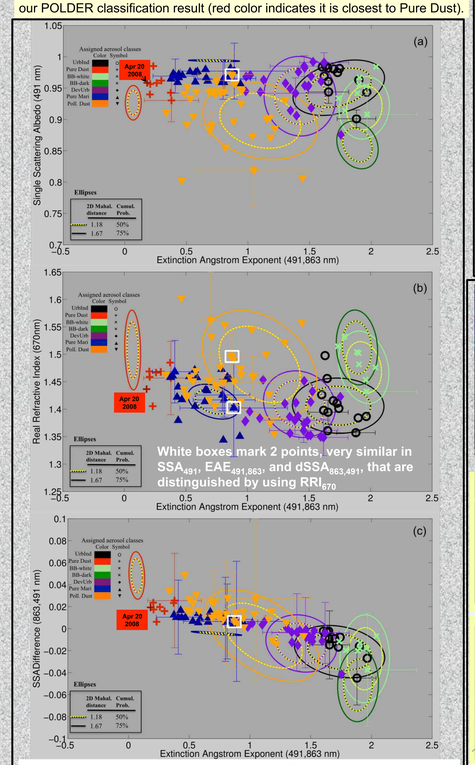
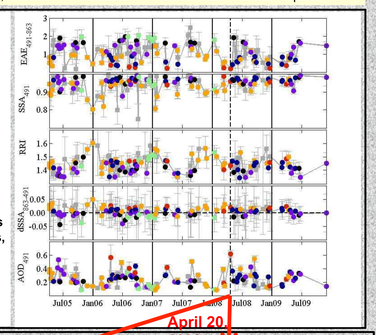


Fig. 4. Results of applying our 4-parameter (SSA_{491}, $EAE_{491,863}$, RRI_{670}, $dSSA_{863,491}$) aerosol classification technique to the updated Hasekamp et al. (2011) retrievals from PARASOL at FORTH-Crete. Colors of ellipses identify the prespecified clusters (aerosol classes) from Table 2; symbol shapes and colors of the FORTH-Crete data points identify the aerosol class to which they are assigned. Error bars show retrieval uncertainty for representative PARASOL-retrieved points.

Fig. 5. The classification results of Fig. 4 (colored points) shown as a 5-year time series in the 4 parameters used by the classification algorithm: $EAE_{491,863}$, SSA_{491}, RRI_{670}, and $dSSA_{863-491}$. The extensive variable, AOD_{491}, is shown in the lowest frame for reference only; it is not used by the classification algorithm. Gray points have error bars that exceed our input limits, and hence were not classified. Dashed vertical line marks April 20, 2008, the dust case explored in Fig. 6.



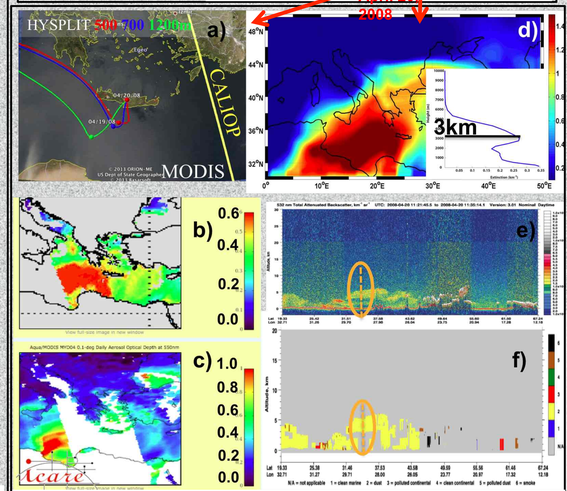


Fig. 6. Ancillary results re the Apr 20, 2008 aerosol over FORTH-Crete, for which our algorithm classifies POLDER results as Dust. a) MODIS image for April 20, 2008, 11:20 UTC showing dust flowing from Libya, with HYSPLIT back trajectories from FORTH-Crete at 3 altitudes, and CALIOP track for April 20, 2008, 11:22-11:35 UTC. b) POLDER-retrieved AOD_{865}. c) MODIS-retrieved AOD_{550}. d) GEOS-Chem predictions of dust AOD_{550}. Insert: GEOS-Chem extinction coefficient profile (550 nm, km^{-1}) above FORTH-Crete. e) CALIOP-retrieved total attenuated backscatter at 532 nm. f) CALIOP-assigned aerosol type.

Summary and Conclusions

- Our method, **Mahalanobis classification using prespecified clusters (classes), defines the pre-specified clusters (Table 2) using parameters retrieved from AERONET stations** where a single aerosol type tends to dominate in certain months. To "purify" each class, points are filtered in both parameter values and Mahalanobis distance.
- **We applied the method to aerosol parameters retrieved from the POLDER-3 polarimeter on the PARASOL spacecraft.** The POLDER retrieval is based on that described by Hasekamp et al. (2011), updated to use more wavelengths and include particle non-sphericity (though in a different way than the AERONET algorithm).
- **We used a modified Mahalanobis distance, D_{EC}, that takes explicit account of the N-dimensional uncertainty on N-dimensional input points retrieved from POLDER.**
- **Applying the classification algorithm to a 5-year series of parameters retrieved from POLDER measurements at Crete yielded classifications into six of our seven prespecified clusters** (Figs. 4 and 5), with only Biomass Burning—Dark Smoke receiving no data points.
- The **April 20, 2008** retrieval at Crete, **classified as closest to Pure Dust, is consistent with a variety of ancillary measurements and analyses**, including MODIS RGB imagery, HYSPLIT trajectories, POLDER and MODIS AOD maps, CALIOP-retrieved vertical cross sections, CALIOP-assigned aerosol type, and GEOS-Chem modeling (Fig. 6).

Outlook

- The classification method could be applied to POLDER-retrieved data sets more global in scope and extensive in time, when they become available in the future.
- And also to other parameters and sensors (e.g., Table 1) because of the flexible applicability of the Mahalanobis and modified Mahalanobis distance measures used here.
- Comparisons to coincident results from other methods (e.g., more extensive modeling; in situ measurements; and classifications using HSRL- or 4STAR-retrieved parameters) can help to reveal the relative strengths of each method.