The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm


* NASA Langley Research Center, Hampton, Virginia 1 Science Systems and Applications International, Hampton, Virginia # National Institute of Aerospace, Hampton, Virginia

(Manuscript received 23 September 2008, in final form 24 April 2009)

ABSTRACT

Descriptions are provided of the aerosol classification algorithms and the extinction-to-backscatter ratio (lidar ratio) selection schemes for the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) aerosol products. One year of CALIPSO level 2 version 2 data are analyzed to assess the veracity of the CALIPSO aerosol-type identification algorithm and generate vertically resolved distributions of aerosol types and their respective optical characteristics. To assess the robustness of the algorithm, the interannual variability is analyzed by using a fixed season (June–August) and aerosol type (polluted dust) over two consecutive years (2006 and 2007). The CALIPSO models define six aerosol types: clean continental, clean marine, dust, polluted continental, polluted dust, and smoke, with 532-nm (1064 nm) extinction-to-backscatter ratios \( S_a \) of 35 (30), 20 (45), 40 (55), 70 (30), 65 (30), and 70 (40) sr, respectively. This paper presents the global distributions of the CALIPSO aerosol types, the complementary distributions of integrated attenuated backscatter, and the volume depolarization ratio for each type. The aerosol-type distributions are further partitioned according to surface type (land/ocean) and detection resolution (5, 20, and 80 km) for optical and spatial context, because the optically thick layers are found most often at the smallest spatial resolution. Except for clean marine and polluted continental, all the aerosol types are found preferentially at the 80-km resolution. Nearly 80% of the smoke cases and 60% of the polluted dust cases are found over water, whereas dust and polluted continental cases are found over both land and water at comparable frequencies. Because the CALIPSO observables do not sufficiently constrain the determination of the aerosol, the surface type is used to augment the selection criteria. Distributions of the total attenuated color ratios show that the use of surface type in the typing algorithm does not result in abrupt and artificial changes in aerosol type or extinction.

1. Introduction

Aerosol classification can take many forms. For the purpose of estimating human-induced aerosol radiative forcing estimates, aerosols are broadly classified as anthropogenic (urban/industrial pollution and biomass burning) and natural (desert dust, sea salt, biogenic, and volcanic) aerosols. For example, in estimating aerosol radiative forcing values, the most recent Intergovernmental Panel on Climate Change (IPCC) report (Solomon et al. 2007) not only adopts broad anthropogenic and natural categories but also addresses the effects of specific species such as black carbon on snow. When direct measurements of the speciation of particle samples can be made, aerosols are chemically classified by the predominant species (e.g., sulfates, black carbon, organic carbon, etc.). Aerosols have also been classified by their hygroscopicity as being water soluble or water insoluble. Such classification is particularly useful for studies of the aerosol effects on cloud formation, droplet size, and cloud longevity [i.e., the so-called indirect effects of aerosols (Twomey 1977)]. The deployment of satellite-based active [e.g., Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO), Geoscience Laser Altimeter System (GLAS)] and passive instruments [e.g., Multiangle Imaging SpectroRadiometer...
(MODIS), Moderate Resolution Imaging Spectroradiometer (MODIS), Ozone Monitoring Instrument (OMI), and Polarization and Directionality of the Earth’s Reflectance (POLDER) have afforded global daily measurements of optical properties that are useful for aerosol classification.

One of the challenges of satellite-based passive remote sensing of aerosol properties is the separation of the path radiance from the top-of-atmosphere (TOA) radiance. The path radiance is due to atmospheric reflection, whereas TOA radiance includes reflection by the surface. For aerosol retrieval from passive instruments, once the path radiance is isolated, some methods use simulations of the top-of-atmosphere radiation to relate satellite-based observations to aerosol properties by using theoretical models. These models are usually based on measurements or established climatologies (e.g., d’Almeida et al. 1991; WCRP 1986). In most cases these algorithms make assumptions about the vertical distribution of aerosols and surface reflectance, all of which have significant contributions to the top-of-atmosphere radiation.

OMI (Levett et al. 2006) uses two spectral regions—17 wavelengths in the 331–500-nm range and two wavelengths in the near UV—as two independent methods for making aerosol retrievals (Torres et al. 2002). The OMI aerosol algorithm derives aerosol optical properties by comparing the measured reflectance to results from radiative transfer calculations of five major aerosol models: urban-industrial, biomass burning, desert dust, oceanic, and volcanic. These are further categorized into 24 subtypes by size distributions and refractive indices. Like the CALIPSO models described in section 3, the OMI models are derived mainly from measurements of the Aerosol Robotic Network (AERONET) sun photometer observations (Holben et al. 1998).

The MODIS aerosol algorithms use multiwavelength radiances along with the scattering angle to determine dust and nondust aerosol types and the fine-mode fraction of the total optical thickness (Remer et al. 2005). The MODIS collection 05 products are derived from four fine modes and five coarse modes over the ocean. These include water soluble (two fine modes), water soluble without humidity (two fine modes), wet sea salt (three coarse modes), and dustlike (two coarse modes). Recent improvements in the MODIS retrievals, using the “Deep Blue” algorithm, have enabled discrimination of dust plumes from fine-mode pollution particles, even in complex aerosol environments. Deep Blue uses radiance measurements from the Sea-viewing Wide Field-of-view Sensor (SeaWiFS) and MODIS to characterize aerosols over source regions (Hsu et al. 2006).

The MISR standard aerosol retrieval algorithm produces aerosol optical thickness and aerosol type at 17.6-km resolution by using top-of-atmosphere radiances (Kahn et al. 2001). In the latest MISR standard aerosol algorithm, version 17, there are 74 mixtures of eight components, which are grouped by aerosol absorptivity, number modes, shape, and composition (Kahn et al. 2001, 2007) to generate a lookup table of radiances at several view angles. The modeled radiances, which assume that the vertical distribution of the aerosols have a negligible effect on the results, are compared to the measured radiances, and the best-fit model is chosen as the representative aerosol type. For optical depth comparisons, these algorithms have worked very well (Diner et al. 2001; Kahn et al. 2005) and may realize improvements by using CALIPSO’s vertically resolved measurements.

The POLDER instrument is a wide field-of-view imaging spectroradiometer that uses measurements of the spectral, directional, and polarized characteristics of the solar radiation reflected by the earth-atmosphere system to retrieve aerosol properties. POLDER compares a simulated degree of depolarization and total radiance saved in a lookup table as a function of aerosol optical thickness, zenith viewing angle, solar zenith angle, and relative azimuth angle with the measured depolarization and total radiance (Sano and Mukai 2000). An aerosol model that yields the minimum difference in the measured simulated total radiance and degree of polarization is chosen as the representative aerosol type.

CALIPSO (Winker et al. 2009, 2007) is an active sensor that makes range-resolved measurements of atmospheric constituents, and it does not require the assumptions about vertical distribution and surface reflectance that are fundamental to most passive measurements. To determine aerosol type, these algorithms use the integrated attenuated backscatter measurements and the volume depolarization ratio measurements, as well as surface type and layer altitude, to determine aerosol type. The algorithms do make a fundamental assumption (discussed in section 2) for the extinction retrieval of most aerosol layers.

The aerosol subtyping products generated by the algorithms are available to the public at the Langley Atmospheric Sciences Data Center (ASDC) as part of the level 2 suite of products. Vertically resolved aerosol subtyping from spaceborne lidar measurements on an operational basis is unprecedented. Therefore, the first few versions of the algorithm will evolve based on the knowledge gained from CALIPSO measurements and other studies.

2. Extinction-to-backscatter (lidar) ratio

Although backscatter lidar measurements such as CALIPSO are free from surface effects, they do suffer
from a need to assume an aerosol extinction-to-backscatter ratio \( S_a \) (also referred to as the aerosol–lidar ratio) to enable, in most cases, the calculation of extinction from lidar backscatter signals. The few cases (for aerosols) when we do not need to assume an \( S_a \) is when we can estimate the transmission (as in Young 1995) when the feature is bounded by sufficient clear-air regions. This method uses the following equation describing the relationship between optical depth and integrated attenuation backscatter:

\[
\gamma' = \frac{2\pi}{\eta S_a} [1 - \exp(-2\eta\tau)].
\] (1)

Here, \( \gamma' \) is the integrated (from layer base to top) attenuated backscatter, \( \tau \) is optical depth, and \( \eta \) is a multiple scattering parameter. The effective two-way transmittance \( T^2 \) is

\[
T^2 = \exp(-2\eta\tau).
\] (2)

If we define an effective \( S_a \), where \( S = \eta S_a \), then we can rewrite Eq. (1) as follows:

\[
S = \frac{1 - T^2}{2\gamma'}.
\] (3)

The effective two-way transmittance is typically obtained by fitting the return both above and below a feature to a reference profile.

We use \( S_a \) in the extinction algorithm in all other cases because the lidar equation is underdetermined (i.e., there are two unknowns for one equation). A unique solution is only possible if a relationship can be established (or prescribed) between the two unknowns, extinction and backscatter, such that these are combined into one variable (cf. Fernald 1984) so that \( S_a = \sigma_{fa} \).

For illustrative purposes, consider a lidar system with a field of view so small that the multiple scattering effects can be neglected. The single-scatter lidar equation for such a system can be written as

\[
N(z) = \eta \frac{P_{\text{Delta}}}{h_\nu} \frac{A}{4\pi(z - z_s)\tau}\beta_T(z)T^2(z - z_s)\Delta z.
\] (4)

where \( N(z) \) is the received signal photon count from a sample volume of thickness \( \Delta z \) at altitude \( z \). \( \eta \) is the system efficiency, \( \Delta t \) is the measurement integration period, \( P \) is the average laser output power, \( h_\nu \) is the photon energy, \( h \) is Planck’s constant, \( \nu \) is the optical frequency, \( A \) is the receiving telescope area, \( z_s \) is the lidar site altitude, and \( \beta_T(z) \) is the total volume backscatter coefficient. Also, \( T^2(z - z_s) \) is the two-way atmospheric transmission over the range \( z - z_s \) given by

\[
T^2(z - z_s) = \exp\left[-2\int_{z_s}^{z} \sigma_T(l)\,dl\right],
\] (5)

where \( \sigma_T \) is the total atmospheric extinction coefficient.

Both molecular and aerosol components contribute to the atmospheric extinction and backscatter coefficients:

\[
\sigma_T(z) = \sigma_m(z) + \sigma_a(z) = S_m\beta_m(z) + S_a\beta_a(z) \quad \text{and} \quad \beta_T(z) = \beta_m(z) + \beta_a(z),
\] (6)

where the subscripts \( m \) and \( a \) refer to molecular and aerosol scattering, respectively, and \( S \) is the lidar ratio. Molecular scattering varies inversely to the fourth power of the wavelength, whereas the wavelength dependence of aerosol scattering depends on the size distribution, shape, and refractive index of the aerosol particles. The total backscatter coefficient \( \beta_T \) can be calculated from the measured lidar profile using, for example, Eq. (3) in Fernald (1984).

The value of \( S_m \) can be calculated from the Rayleigh scattering phase function of isotropic unpolarized light \( P^R(\pi) = \frac{3}{8}[1 + \cos^2(\pi)] \), and it is equal to \( 8\pi/3 \) sr. The reference altitude \( z_o \) is chosen where the aerosol scattering is negligible [i.e., \( \beta_a(z_o) \approx 0 \)]. For the CALIPSO analyses, the molecular backscatter coefficient profile is computed from theory by using model atmosphere values for the temperature \( T(z) \) and pressure \( P(z) \) obtained from the Global Modeling and Assimilation Office (GMAO; Bloom et al. 2005). The value of \( S_a \), on the other hand, must be determined on a case-by-case basis and depends on the aerosol composition, size distribution, and shape. These properties depend primarily on the source of the aerosol and such factors as mixing, transport, and—in the case of hygroscopic aerosols—hydration. The accuracy of the \( S_a \) value used in the lidar inversions depends on the correct identification of the type of aerosol, a process we refer to as aerosol “typing” or “subtyping.” Though the algorithm assigns an aerosol type to each aerosol layer, the primary driver of this algorithm is the determination of the best estimate of the aerosol lidar ratio.

In following section, we discuss the CALIPSO aerosol models and how we determine the aerosol type, choose the appropriate aerosol model for the type, and determine the appropriate \( S_a \). We also discuss the basis of the \( S_a \) estimates and describe the CALIPSO \( S_a \) selection algorithm. Notwithstanding the choice of \( S_a \), the aerosol subtype product is useful for developing global distributions of aerosol types for science studies, such as source attribution of aerosol forcing and pollution episodes. To determine performance of aerosol subtyping and \( S_a \) selection, we use qualitative measures,
such as the distributions of optical depths, expected aerosol types both geographically and seasonally, and convergence of the extinction calculation when model \( S_a \) values are used. In section 5, we make comparisons of \( S_a \) with the Langley airborne high spectral resolution lidar (HSRL). Although the 532-nm \( S_a \) values are well documented in the literature, there is a dearth of 1064-nm \( S_a \) measurements. Our treatment of the 1064-nm \( S_a \) values reflect this state of knowledge; that is, nearly all our 1064-nm lidar ratio estimates are based on models, whereas the 532-nm values are from both measurements and models.

3. CALIPSO aerosol models

The CALIPSO aerosol models are based on the cluster analysis of a multyear (1993–2002) AERONET dataset to determine characteristic aerosol types grouped using nearly instantaneously observed physical and optical properties (Omar et al. 2005). The cluster analysis identified six aerosol types representative of the aerosol mixtures most frequently observed at the AERONET sites. The CALIPSO models use some of the information derived from AERONET cluster analysis to determine \( S_a \) values. The typing and lidar ratio selection algorithm uses the lidar observables and surface-type information to aid in selecting values of \( S_a \). The goal is to constrain the uncertainty in \( S_a \) to no more than 30%. The observed range of variability of \( S_a \) is between 10 and 110 sr (Anderson et al. 2000) and the modeled range is between 15 and 80 sr (Ackermann 1998). Allowing for an uncertainty of 30%, our selected range of \( S_a \) values for this algorithm spans a range of 14–90 sr.

For each aerosol model obtained from the cluster analysis of AERONET measurements, a lidar ratio was calculated and compared with field measurements of \( S_a \) that aerosol type. These measurements, primarily at 532 nm, come from a variety of techniques including 180° nephelometer (Anderson et al. 2000), Raman lidar (Ackermann 1998; Ferrare et al. 2001; Franke et al. 2001; Müller et al. 2000), slant path techniques, and transmittance methods (Young 1995). The AERONET-derived model parameters were adjusted, when necessary, to bring the derived lidar ratio into better agreement with the measurements.

Of the six AERONET clusters, three [biomass burning, polluted continental, and polluted dust (dust + smoke)] are adopted as CALIPSO aerosol models. Two models (marine and background/clean continental) were built either directly from measurements of size distributions and complex refractive indices or by adjusting model parameters to generate observed \( S_a \) values. The dust model is based on the work of Kalashnikova and Sokolik (2002). The size distributions based on volume are shown in Fig. 1, and the microphysical and derived properties associated with each of the six CALIPSO aerosol models are listed in Table 1. In the following subsections, we discuss the aerosol types used and the basis for the \( S_a \) values.

a. Dust

The CALIPSO dust model is based on theoretical particle scattering calculations by using the discrete-dipole approximation (DDA) technique, with inputs of realistic compositions and irregular shapes (Kalashnikova and Sokolik 2002). The CALIPSO model \( S_a \) value of 40 sr at 532 nm for desert dust is comparable to the \( S_a \) measurements near the green channel by Voss et al. (2001) using a micropulse lidar (41 ± 8 sr) for African dust, measurements of Liu et al. (2002) of Asian dusts (42–55 sr) found using an HSRL and a combined Raman elastic-backscatter lidar values. During the Aerosol Characterization Experiment-Asia (ACE-Asia) intensive observation period, Murayama et al. (2003) measured lidar ratios of 50.4 ± 9.4 sr for an elevated (>4.5 km) highly depolarizing (>30%) dust layer near Tokyo. Cattrall et al. (2005) report an \( S_a \) value of 42 ± 4 sr for desert dust locations. Müller et al. (2007) used Raman lidars to measure 532-nm dust \( S_a \) values of 59 ± 11, 35 ± 5, and 38 ± 5 sr for the Sahara, Gobi, and Arabian Deserts, respectively. At 1064 nm, the \( S_a \) value used by these algorithms is 55 sr. This value is based on studies of long-range dust transport by Liu et al. (2008) using CALIPSO measurements and two wavelength inversion techniques. These studies found 1064-nm \( S_a \) values of 52 ± 5, 53 ± 5, and 54 ± 13 sr near the northwest coast of Africa, 1300 km from the coast, and 2400 km from the coast, respectively. Another study, using Raman lidars and airborne HSRL to profile of Saharan dust by Tesche et al. (2009), estimated extinction-to-backscatter ratios of 53–55 sr (±7–13 sr) at 355, 532, and 1064 nm. Earlier studies during the Southern Africa Regional Science Initiative (SAFARI) campaign, in which McGill et al. (2003) used AERONET optical depth measurements to constrain the lidar retrievals and determine the appropriate \( S_a \) values for different optical depth regimes, found 1064-nm \( S_a \) values of 31.9 ± 4.7, 33.7 ± 2.7, 35.4 ± 6.0 sr for light, moderate, and heavy aerosol layers, respectively, albeit for dust and smoke mixtures.

b. Polluted continental

The CALIPSO model for polluted continental yields 532- and 1064-nm \( S_a \) values of 70 and 30 sr, respectively. The \( S_a \) measurements by Ansmann et al. (2001) at Sagres, an island off the Portuguese coast, for pollution
emanating from continental Europe varied between 50 and 70 sr. During Indian Ocean Experiment (INDOEX) measurements, $S_a$ values of polluted continental aerosol originating from northern and northeastern India, known for high emissions of black carbon, were made by Franke et al. (2001). They found values ranging from 49 to 70 sr. Cattrall et al. (2005) report an $S_a$ value of $71 \pm 10$ sr for urban–industrial locations. Measurements of a stagnant air mass at Bondville (a polluted continental site) yielded $S_a$ values of $64 \pm 4$ sr (Anderson et al. 2003). Barnaba and Gobbi (2004) obtained the best fit to sun photometer data by using a lidar ratio of 60 sr for their backscatter and an extinction model of continental aerosol.

c. **Polluted dust**

This aerosol model is designed to account for episodes of dust mixed with biomass burning smoke, which are frequent in regions close to strong sources of both [e.g., in West Africa (cf. MODIS images) and Asia (cf. ACE-Asia, INDOEX]. It also accounts for instances of dust mixed with urban pollution as is frequently encountered in parts of Asia and Europe. The CALIPSO polluted dust model is a mixture of the AERONET desert dust (coarse mode) and biomass burning (fine mode) clusters. This model yields an $S_a$ value of 65 sr at 532 nm, which is comparable to similar measurements of polluted
dust (Liu et al. 2002; Voss et al. 2001). At 1064 nm, the $S_a$ value for the CALIPSO polluted dust model is 30 sr.

d. Smoke (biomass burning)

The biomass burning cluster of AERONET measurements is used to model the CALIPSO smoke aerosol. Cluster analysis of the AERONET data yields $S_a$ values of 70 sr at 532 nm and 40 sr at 1064 nm. These values compare well with the 532-nm measurements of Ansmann et al. (2001) of 70 sr for biomass burning–influenced aerosol advected from the Indian subcontinent during INDOEX and with Voss et al. (2001) of 60 ± 6 sr off the west coast of Africa. Cattrall et al. (2005) report values of 71 sr at 532 nm from their AERONET study.

e. Clean continental

This aerosol is also referred to as clean background. The AERONET records of the background cluster from Omar et al. (2005) have low mean optical depths (<0.05 at 673 nm). The microphysical properties derived from these are likely to have large uncertainties (Dubovik et al. 2002). The CALIPSO background aerosol model was derived by fitting size distributions and refractive indices to measurements of $S_a$ of long-range continental transport (Anderson et al. 2000). This generates an $S_a$ value of 35 sr. An $S_a$ value of 32 ± 6 sr for the clean Northern Hemisphere aerosol was measured during the Aerosols99 (Voss et al. 2001). A similar aerosol termed “background like,” originating on the European continent but devoid of any strong biomass or fossil fuel burning signature, yielded an $S_a$ value of 35 sr (Ansmann et al. 2001). Our CALIPSO model for this type of aerosol yields $S_a$ values of 35 sr at 532 nm and 30 sr at 1064 nm.

f. Clean marine

Because the AERONET marine aerosol cluster from Omar et al. (2005) is comprised of a small number of records (<4% of the total), the CALIPSO marine aerosol model is derived from the parameters measured during the Shoreline Environmental Aerosol Study (SEAS) experiment (Masonis et al. 2003). Using a marine aerosol model derived from measured size distributions during SEAS yields $S_a$ values of 20 sr at 532 nm and 45 sr at 1064 nm. This 532-nm $S_a$ value for marine aerosols is consistent with marine aerosol $S_a$ estimates by others (Ansmann et al. 2001; Flamant et al. 1998; Reagan et al. 2001). Measurements during INDOEX in the marine boundary layer of the tropical Indian Ocean report a value of 23.5 sr at 532 nm (Müller et al. 2007). In their climatological study of oceanic AERONET sites, Cattrall et al. (2005) report an $S_a$ value of 28 ± 5 sr. Note that most AERONET “oceanic” sites are actually land based and are not free from terrestrial and anthropogenic influences. In the SEAS experiment, the $S_a$ value is measured directly (optical method) or modeled (using measured size distributions). The optical method, using a 180° backscatter nephelometer, an integrating nephelometer, and an absorption photometer, yields a value of $S_a = 25.4 ± 3.5$ sr. The modeled values are 20.3 sr (at 532 nm) using direct size data. The CALIPSO model uses the SEAS direct size data to calculate both the 532-nm value (20 sr) and the 1064-nm value (45 sr).

4. Algorithm design

The goal of the aerosol typing algorithm is to identify the aerosol type closely enough to estimate the

<table>
<thead>
<tr>
<th>Optical/physical property</th>
<th>Dust</th>
<th>Smoke</th>
<th>Clean continental</th>
<th>Polluted continental</th>
<th>Clean marine</th>
<th>Polluted dust</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_r$ fine at 532 nm</td>
<td>1.414</td>
<td>1.517</td>
<td>1.380</td>
<td>1.404</td>
<td>1.400</td>
<td>1.452</td>
</tr>
<tr>
<td>$m_i$ fine at 532 nm</td>
<td>0.0036</td>
<td>0.0234</td>
<td>0.0001</td>
<td>0.0063</td>
<td>0.0050</td>
<td>0.0109</td>
</tr>
<tr>
<td>$m_r$ fine at 1064 nm</td>
<td>1.495</td>
<td>1.517</td>
<td>1.380</td>
<td>1.439</td>
<td>1.400</td>
<td>1.512</td>
</tr>
<tr>
<td>$m_i$ fine at 1064 nm</td>
<td>0.0043</td>
<td>0.0298</td>
<td>0.0001</td>
<td>0.0073</td>
<td>0.0050</td>
<td>0.0137</td>
</tr>
<tr>
<td>$m_r$ coarse 532 nm</td>
<td>1.414</td>
<td>1.517</td>
<td>1.455</td>
<td>1.404</td>
<td>1.400</td>
<td>1.452</td>
</tr>
<tr>
<td>$m_i$ coarse 532 nm</td>
<td>0.0036</td>
<td>0.0234</td>
<td>0.0034</td>
<td>0.0063</td>
<td>0.0050</td>
<td>0.0109</td>
</tr>
<tr>
<td>$m_r$ coarse 1064 nm</td>
<td>1.495</td>
<td>1.514</td>
<td>1.455</td>
<td>1.439</td>
<td>1.390</td>
<td>1.512</td>
</tr>
<tr>
<td>$m_i$ coarse 1064 nm</td>
<td>0.0043</td>
<td>0.0298</td>
<td>0.0034</td>
<td>0.0073</td>
<td>0.0050</td>
<td>0.0137</td>
</tr>
<tr>
<td>Fine cut-off radius ($\mu$m)</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
<td>0.60</td>
<td>1.00</td>
</tr>
<tr>
<td>Fine fraction by volume</td>
<td>0.223</td>
<td>0.329</td>
<td>0.050</td>
<td>0.531</td>
<td>0.025</td>
<td>0.241</td>
</tr>
<tr>
<td>Fine mean radius ($\mu$m)</td>
<td>0.1165</td>
<td>0.1436</td>
<td>0.205 56</td>
<td>0.1577</td>
<td>0.150</td>
<td>0.1265</td>
</tr>
<tr>
<td>Geometric std dev (GSD)</td>
<td>1.4813</td>
<td>1.5624</td>
<td>1.61</td>
<td>1.5257</td>
<td>1.600</td>
<td>1.5112</td>
</tr>
<tr>
<td>Fine fraction by volume</td>
<td>0.777</td>
<td>0.671</td>
<td>0.950</td>
<td>0.469</td>
<td>0.975</td>
<td>0.759</td>
</tr>
<tr>
<td>Fine mean radius ($\mu$m)</td>
<td>2.8329</td>
<td>3.726</td>
<td>2.6334</td>
<td>3.547</td>
<td>1.216</td>
<td>3.1617</td>
</tr>
<tr>
<td>Coarse fraction by volume</td>
<td>1.9078</td>
<td>2.1426</td>
<td>1.8987</td>
<td>2.065</td>
<td>1.600</td>
<td>1.9942</td>
</tr>
<tr>
<td>Coarse mean radius ($\mu$m)</td>
<td>1.80°</td>
<td>2.1426</td>
<td>1.8987</td>
<td>2.065</td>
<td>1.600</td>
<td>1.9942</td>
</tr>
<tr>
<td>GSD coarse</td>
<td>40</td>
<td>70</td>
<td>35</td>
<td>70</td>
<td>20</td>
<td>65</td>
</tr>
<tr>
<td>GSD coarse</td>
<td>55</td>
<td>40</td>
<td>30</td>
<td>30</td>
<td>45</td>
<td>45</td>
</tr>
</tbody>
</table>

\[ S_a \text{ at } 532 \text{ nm (sr)} \]

\[ S_a \text{ at } 1064 \text{ nm (sr)} \]
appropriate value of \( S_a \) to within 30% of the true value. These values are then passed on for use in the extinction retrieval algorithm (Young and Vaughan 2009). The selection scheme uses the observed backscatter strength, that is, the integrated attenuated backscatter at 532 nm (\( \gamma' \)), defined as

\[
\gamma' = \int_{z_{\text{top}}}^{z_{\text{base}}} \beta(z) T(z) \, dz,
\]

where \( \beta \) is the total (molecular \( m \) + particulate \( p \) backscatter) and \( T \) is the atmospheric transmittance due to both molecules and particles. The sum of the perpendicular channel signal within a feature divided by the sum of the parallel channel (i.e., the volume depolarization ratio \( \delta_v' \)) is defined as

\[
\delta_v' = \frac{\sum \beta_{\perp}'(z)}{\sum \beta_{\parallel}'(z)} = \frac{\sum [\beta_{\perp,m}'(z) + \beta_{\perp,p}'(z)]}{\sum [\beta_{\parallel,m}'(z) + \beta_{\parallel,p}'(z)]},
\]

where \( z \) is altitude and the subscripts “top” and “base” refer to the top and base of the feature as determined by the feature finder (Vaughan et al. 2009). The subscripts \( \parallel \) and \( \perp \) refer to the polarized and depolarized attenuated backscatter \( \beta' \) signals, respectively. In the algorithm, \( \delta_v' \) has been corrected to account for the molecular contribution as follows:

\[
\delta_v = \frac{\delta_v'[R_{\text{mas}} - 1](1 + \delta_m) + 1] - \delta_m}{(R_{\text{mas}} - 1)(1 + \delta_m) + \delta_m - \delta_v'},
\]

where \( \delta_v \) is the corrected depolarization ratio (or estimated particulate depolarization ratio), \( R_{\text{mas}} \) is the mean attenuated scattering (mas) ratio (the ratio of the total attenuated backscatter to the molecular backscatter) and \( \delta_m \) is the molecular depolarization ratio. Because the CALIPSO lidar has a very narrow bandwidth filter (~37 pm), the Rayleigh scattering signal is dominated by the central Cabannes line (i.e., \( \delta_m \approx 0.0036 \) at 532 nm). It follows from Eq. (10) that for large \( R_{\text{mas}} \), \( \delta_v \approx \delta_v' \). For the dust threshold of \( \delta_v' = 0.2, \delta_v \approx 0.97 \) at \( R_{\text{mas}} \) of 1.5 and \( \delta_v \approx 0.4 \) at \( R_{\text{mas}} \) of 2.5. Because of this dependence on the layer \( R_{\text{mas}} \), the distributions of the depolarization ratios will spread beyond the threshold values in Fig. 2.

Aerosol type is identified by using \( \gamma' \) and \( \delta_v \) to the extent possible from among one of the six types. However, \( \gamma' \) and \( \delta_v \) are not sufficient to fully constrain the model selection. Surface type is used to exploit differences in aerosol classes over the oceans, deserts, and snow/tundra regions. Information about aerosol layer elevation is also utilized to determine aerosol type, because conditions that favor lifting mechanisms for dust and smoke are more likely than for other aerosol types. As shown in section 6, there is a significant overlap between the distributions of the color ratio \( \chi' \) (defined as the ratio of the attenuated 1064- and 532-nm backscatter). The overlap between \( \chi' \) distributions for several aerosol types does not allow it to be used directly as an aerosol subtyping tool.

The input parameters—altitude, location, surface type, volume depolarization ratio, and integrated attenuated backscatter measurements—are used to identify the type following one of 12 pathways in Fig. 2. The surface types are from the International Geosphere-Biosphere Programme (IGBP). The threshold values of \( \delta_v \) and \( \gamma' \) in Fig. 2 are estimates based on Lidar In-space Technology Experiment (LITE) measurements; in the case of depolarization, they are based on a limited set of observations and models (Barnaba and Gobbi 2004; Gobbi et al. 2000; Murayama et al. 1999; Reagan et al. 2001; Sakai et al. 2003). Recent measurements of Taklimakan Desert dust of 15%–25% depolarization with typical values of 23% during dust storms (Kai et al. 2008) support the 20% threshold for dust particles. The goal is to base typing decisions on these observables as much as possible and to minimize dependence on geographic information. These values are expected to evolve as more information becomes available. Therefore, the threshold values of \( \delta_v \) and \( \gamma' \) will be implemented as runtime parameters that can be adjusted by using a configuration script. When lofted layers are encountered under favorable conditions, \( S_a \) is computed directly from the integrated backscatter and transmission following Eqs. (1)–(3). Combined with a carefully planned validation campaign, this will allow improvements of the method, accuracy, and confidence in the selection of \( S_a \). The algorithm also generates aerosol-type confidence flags as functions of uncertainties in the 532- and 1064-nm integrated backscatter values.

Aerosol layers over polar regions can either be clean continental or arctic haze, depending on the magnitude of \( \gamma' \) (see pathways 1 and 2 in Fig. 2). Arctic haze is similar to continental pollution and, in some cases, it is continental pollution transported to the Arctic region. Measurements of arctic haze at Spitsbergen, Norway, in the spring show that this aerosol type may have a significant soot component and the lidar ratio reaches values as high as 70 sr in the visible wavelengths (Ritter et al. 2004). Arctic haze has a low depolarization ratio ~2% (Ishii et al. 1999), and it is modeled as polluted continental in this algorithm. As currently written, the algorithm does not allow dust and smoke types in polar regions. A near-term modification of the algorithm will
allow these aerosols in polar regions in light of recent observations in the Arctic [cf. Arctic Research of the Composition of the Troposphere from Aircraft and Satellites (ARCTAS) experiment].

We use $d_v$ to identify aerosol types that have a substantial mass fraction of nonspherical particles (e.g., a mixture of smoke and dust in pathways 3 and 5). We use $\gamma'$ to discern instances of transient high aerosol loading over surfaces where this is not usually expected (e.g., a smoke or dust layer over land or the ocean in pathways 10 and 11, respectively). In pathway 4, aerosol layers that have $\delta_v > 0.2$ are identified as dust everywhere but at the poles. Thin aerosol layers over land (other than dust) are identified as clean continental (pathway 6) unless these layers are elevated, in which case they are identified as smoke (pathway 11). If, however, they are not elevated but still scatter more strongly than the clean continental threshold ($\gamma' = 0.0005$), then we identify them as polluted continental (pathway 7). Over the ocean, all elevated nondust aerosol layers are identified as smoke. In the marine boundary layer (MBL), aerosol layers can either be clean marine characterized by optically thick layers consisting of spherical particles (pathways 9 and 10) or optically thin layers with some nonspherical particles, as would be found near coastal regions and classified as polluted continental (pathway 8). Once the type is identified, $S_n$ is chosen from a lookup table that currently consists of the six pairs of 532- and 1064-nm values shown in Table 1.

5. Algorithm performance

a. Case studies of aerosol scenery

Figures 3, 4, and 5 show the backscatter browse image (Figs. 3a, 4a, 5a), cloud/aerosol mask (Figs. 3b, 4b, 5b), and classification of the aerosol found by the aerosol subtyping algorithm (Figs. 3c, 4c, 5c) for three scenes observed on 12 August 2006. Figure 3 shows the distribution of aerosol types as the satellite passes over western Russia, continental Europe, across the Mediterranean Sea, and to the eastern Sahara Desert. The subtyping algorithm captures the evolution of the aerosol from smoke and polluted dust to pure dust in the Sahara. The aerosol classification algorithm only operates on those inputs from the cloud aerosol discrimination (CAD) algorithm (Liu et al. 2009) that have been classified as aerosols. The CAD output is shown in Fig. 3b. The dust layer stretches for about 4000 km and is mixed...
with smoke from Siberian fires at the leading edge. The horizontal extent and the aerosol types shown (smoke, polluted dust, and dust) verify that the algorithm finds the expected types of aerosol.

Figure 4 is another example of the subtyping result showing an aerosol layer that has been classified as predominantly smoke extending from land to the deep ocean. Note that the aerosol type on the ocean surface is classified as marine aerosol. The algorithm identifies this 3000-km layer of smoke found in southwestern Africa. Carbon monoxide (CO) measurements from the Measurements of Pollution in the Troposphere (MOPITT) by Bremer et al. (2004) show that CO in this region is at its peak during September–November (SON), although the number of southern African fires may peak in June–August (JJA), with some interannual variation.

CALIPSO observed this layer on 12 August 2006, one day after the MODIS instrument on *Aqua* reported a large number of fires (see the 11 August 2006 MODIS Rapid Response System images; available online at http://rapidfire.sci.gsfc.nasa.gov/gallery/) in the same region. The image also illustrates that the algorithm is not dependent on surface type, because the layer stretches from land to the Atlantic Ocean (i.e., the use of surface types in the algorithms do not result in abrupt and artificial changes in aerosol type). The continuity of the type from land to ocean means the extinction and optical depth will exhibit the same uniform transition.

In Fig. 5, all six aerosol types were found in this section of an orbit stretching from Eastern Europe to North Africa. The aerosol type in Europe is primarily polluted continental in the north and smoke and polluted dust in the south near the Mediterranean Sea. The algorithm detects and identifies the clean marine aerosol near the surface of the sea. As the spacecraft flies south, the lidar signal encounters a 5-km-deep dust layer stretching across the Sahara.

### b. Comparison with HSRL measurements

In general, HSRLs take advantage of the differences between the molecular return and the aerosol return to separate the two signals. The spectrum of the molecular lidar return is Doppler broadened by the thermal motion of the molecules, whereas the slow-moving aerosol particles generate very little spectral broadening. An iodine absorption cell is used to absorb the aerosol return, whereas the spectral wings of the Doppler-broadened molecular signal are transmitted with little attenuation. This allows the instrument to measure the extinction directly. Details of the Langley HSRL can be found in Hair et al. (2008). Figure 6 shows a comparison of the probability distribution functions (PDFs) of the 532-nm backscatter ratio (Fig. 6a), extinction coefficient (Fig. 6b), and $S_{\alpha}$ (Fig. 6c) of nearly coincident measurements of CALIPSO and the Langley HSRL. The distributions shown are an amalgamation of measurements.
taken during the CALIPSO CloudSat Validation Experiments (CCVEX), the CALIPSO Twilight Zone Experiments (CATZ), and HSRL–CALIPSO validation experiments, which are collectively referred to as the CCVEX/CATZ/CALIPSO dataset. HSRL has the capability of making direct $S_a$ measurements, and Fig. 6c shows continuous change in $S_a$ values from HSRL and CALIPSO’s high frequency of fixed 532-nm $S_a$ values shown by the stove pipes at 20, 35, 40, 65, and 70 sr. The biases for this dataset are 7.4 sr (24.3%; HSRL lower) for the lidar ratio and 0.0029 km$^{-1}$ (23.9%; HSRL lower) for extinction. Figure 7 shows PDFs of the CALIPSO and HSRL measurements of the 532-nm backscatter ratio (Fig. 7a), extinction coefficient (Fig. 7b), and $S_a$ (Fig. 7c) during the Gulf of Mexico Atmospheric Composition and Climate Study (GoMACCS). As in Fig. 6, the backscatter ratio and extinction coefficient distributions show good agreement between CALIPSO and HSRL. The CALIPSO $S_a$ distributions show a high frequency of fixed $S_a$ values—40 (dust), 65 (polluted dust) and 70 sr (smoke, polluted continental)—whereas the HSRL measurements show a continuous distribution spanning these values. The lidar ratio bias for this dataset is 9.2 sr (37.5%, HSRL lower) and the extinction bias is 0.015 km$^{-1}$ (58.9%, HSRL lower).

c. Optical depth and lidar ratio distributions

The extinction retrieval provides a constraint on lidar ratio, albeit very weakly for aerosols. As explained in Young and Vaughan (2009), the retrieved profile of the backscatter may diverge from the correct values if incorrect parameters are used. The profile solver detects if the retrieval is diverging and changes the lidar ratio values in successive intervals to obtain a convergent, nonnegative solution of the extinction equation. Figure 8 shows cases where the lidar ratio has been reduced to prevent divergence of the extinction solution for the month of August 2006. The few cases where the lidar ratio adjustment is to increase the lidar ratio [extinction quality-control (QC) value of 4 in the version 2 data products] are not shown in Fig. 8. In Fig. 8, most of the cases where a lidar ratio change is warranted occur for very thick aerosol layers with optical depths greater than one. Table 2 shows the statistics of the change in lidar ratio values for the 5-km layers (i.e., layers found at a 5-km resolution) for the same month. Table 2 show that most of the aerosol layers produced a solution with the CALIPSO aerosol model $S_a$ value. Although this is not a sufficient condition for the correctness of these values, it is a necessary condition. In other words, although we cannot know that these model $S_a$ values...
correctly characterize the actual aerosol layers observed, we do know that they produce physically realistic solutions. By the same token, the large number of polluted dust lidar ratios (14.4% of the total polluted dust layers for August 2006) that were adjusted downward indicates that the model value of 65 sr may be high for this air mass. It is also noteworthy that, because the most frequent lidar ratio reductions occur at high optical depths (>1.0), the result shown in Fig. 8 may include an artifact of the cloud–aerosol discrimination algorithm, where some clouds may have been misclassified as aerosols.

Figure 9 shows the probability distribution functions of the optical depth $\tau$ of the six aerosol types. The plots show the $\tau$ distributions for unconstrained layers in blue and all layers in red. The unconstrained layers are the layers whose $S_a$ is unchanged (i.e., the optical depth is retrieved by using the CALIPSO model $S_a$ values). The distributions are nearly lognormal for each type. The clean continental and clean marine types show that most
of the layers are unconstrained. The rest of the aerosol types are characterized by a smooth tail with a small hump near an optical depth of 1.5. Clean continental and clean marine have the smallest 532-nm lidar ratios (35 and 20 sr, respectively) and therefore result in low optical depth, high-transmittance layers. The hump in Fig. 9 occurs because of highly attenuating layers; that is, the transmittance $T^2$ is nearly zero, which results in a

![Figure 7](https://via.placeholder.com/150)

**FIG. 7.** CALIPSO and HSRL measurements during GoMACCS.

![Figure 8](https://via.placeholder.com/150)

**FIG. 8.** Distributions of the change in lidar ratios needed to reach a convergent solution in the extinction algorithm for August 2006.
nonconvergent negative solution of the extinction equation, and this necessitates an adjustment of the lidar ratio. Thus, following Young (1995), we write

\[ T^2 = 1 - 2\gamma S_a = 0 \quad \text{and} \quad \tau = -0.5 \ln(1 - 2\gamma S_a). \]  

(11) (12)

Solving for \( S_a \) in Eq. (11) and reducing \( S_a \) by 5% and substituting Eq. (12), we get

\[ \tau = -0.5 \ln \left[ 1 - 2(0.95S_a) \right] = -0.5 \ln(0.05) \approx 1.5. \]

The reduction of the lidar ratio by 5% in search of a nonnegative convergent solution therefore results in an optical depth of ~1.5. This is the case irrespective of the aerosol type or lidar ratio provided \( S_a \) is adjusted down by 5%. The hump is most prominent for polluted dust, which has the largest number of opaque layers (14.4%; shown in Table 2). The figures also show that, where the lidar ratio needs adjustment, a convergent solution is frequently obtained after the first iteration [i.e., at \( S_a(\text{new}) = 0.95S_a(\text{old}) \)].

Less than 1% of the marine aerosols needed any adjustment at all. This is expected because marine layers are frequently optically thin. Nearly 5% of the dust and smoke aerosols and 6% of the polluted continental aerosols needed adjustments mostly to reduce the lidar ratio. A significant proportion of polluted dust (14.4%) needed adjustments for a convergent extinction solution.

Figure 10 shows the mean optical depth of each type for the four seasons. For each type, the seasonal optical depth mean does not vary appreciably. Note that this is different from the type apportioned optical depth (i.e., these values do not add up to yield a mean aerosol optical depth for the season). Clean continental and clean marine have the lowest means, and polluted dust has the highest means, except in SON, when it is smoke. Dust and polluted dust show the same variation patterns and also have the highest seasonal variability.

**Table 2.** The number of aerosol layers found at a 5-km resolution and the percentage of layers whose lidar ratio has been adjusted (%\( \Delta S_a \)) for the month of August 2006.

<table>
<thead>
<tr>
<th>Type</th>
<th>All 5-km layers</th>
<th>( \Delta S_a ) layers (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean continental</td>
<td>4551</td>
<td>0.02</td>
</tr>
<tr>
<td>Clean marine</td>
<td>138 686</td>
<td>0.09</td>
</tr>
<tr>
<td>Dust</td>
<td>34 924</td>
<td>4.48</td>
</tr>
<tr>
<td>Polluted continental</td>
<td>15 492</td>
<td>5.63</td>
</tr>
<tr>
<td>Polluted dust</td>
<td>38 076</td>
<td>14.43</td>
</tr>
<tr>
<td>Smoke</td>
<td>23 360</td>
<td>5.28</td>
</tr>
</tbody>
</table>

**Figure 9.** PDFs of the optical depths by aerosol type. The two histograms show the two cases where the optical depth is retrieved using the initial lidar ratio (unconstrained layers; shown in blue) and all other cases (shown in red), which includes cases where the initial lidar ratio may have been reduced or increased to obtain a solution during the retrieval of the optical depth.
6. Probability distribution functions of aerosol parameters and subtypes

a. Seasonal fractional frequencies of the integrated attenuated color ratio $\chi'$

The integrated attenuated backscatter and the volume depolarization ratio are inputs into the subtyping algorithm, and their distributions reflect the selection scheme shown in Fig. 2. The total attenuated color ratio is the parameter whose distribution has the most significance, because it is not used in the subtyping algorithm and is therefore an independent quantity. In Fig. 11, although the $\chi'$ distributions are not identical, there is very little variation in the mean values between the three types shown. Though one would expect there to be more variation between dust and smoke, because of particle size differences, Figs. 11a,c show that $\chi'$ is centered around 0.5 for both types. This is because $\chi'$ is modulated by the scattering ratio and is therefore not directly indicative of particle size. Although $\chi'$ are not intensive properties and cannot be used to unambiguously identify aerosol types, preliminary studies of the gradient of the $\chi'$ in smoke layers from layer top to layer bottom show that there might be a useful signature for identifying smoke layers in these gradients. Of more significance is the overall similarity of the total attenuated color ratio distributions over land and water. This is an indication of the relative independence of the algorithm from surface type.

b. Seasonal fractional frequencies of the optical properties $\delta_v$ and $\gamma'$

Figures 12 and 13 show the fractional frequencies of the volume depolarization ratio $\delta_v$ and the integrated attenuated backscatter $\gamma'$, respectively. The aerosol $\delta_v$ fractional distributions shown in Figs. 12a–f for the aerosol types by season exhibit multiple modalities for dust, polluted dust, and polluted continental. The polluted continental $\delta_v$ distributions are significantly different in December–February (DJF) than the other seasons. In DJF, the $\delta_v$ mode shifts to more spherical particles, which may be an effect of hydration. The dual modalities in dust and polluted dust may simply be a result of multiple selection pathways (Fig. 2). Overall, the differences in the mean and standard deviations are small within types and significant without, reflecting the type dependence on $\delta_v$ inherent in the algorithms. These distributions for $\delta_v$ do not strictly conform to the threshold values depicted in the flowchart in Fig. 2 because of the correction illustrated in Eq. (10).

The integrated attenuated backscatter ($\gamma'$) probability distribution functions are lognormal for all types and multimodal for dust, polluted dust, and smoke (Fig. 13). There are significant differences in the mean values between types. Both of these characteristics are relics of the design of the subtyping algorithm that uses $\gamma'$ to

![Fig. 10. The seasonal variation of the type attributed optical depth for June 2006–May 2007.](image-url)
discriminate between types. The multiple modes reflect the different branches of the subtyping flowchart, along which a given type can be determined.

c. Fractional frequencies of aerosol types

The first full year (June 2006–May 2007) of version 2 of the CALIPSO measurements were used to generate the seasonal fractional frequencies by type for the JJA, SON, DJF, and March–May (MAM) seasons shown in Fig. 14. The figure also shows the three resolutions (5, 20, and 80 km) at which the features are found. In the figure, the solid red and blue bars (land and water surface types) sum to 1, the green bars (each type as a fraction of all aerosol layers) across the rows sum to 1, and the hatched bars (fraction of the detection resolution) within each box sum to 1. The aerosol layers with the highest backscatter intensity are found at the smallest spatial resolution (5 km). For a detailed description of the CALIPSO layer detection algorithm, see Vaughan et al. (2009). In Fig. 14, “ALL” refers to the fraction of all aerosol layers that are a given type (e.g., nearly 20% of all aerosol layers observed by CALIPSO are clean marine layers during DJF).

Most of the clean continental layers are found over land, with a significant fraction (35%) over the oceans. A large number of clean continental layers over the
oceans are found in the southern ocean and over sea ice, a direct result of pathway 1 in Fig. 2. The fractions of marine, dust, polluted dust, and smoke layers are all comparable at nearly 20%, with the clean continental at about 10% and the polluted continental at less than 5%. This pattern is repeated for all seasons except MAM, when the dust and polluted dust frequencies dominate, most likely because of Asian dust episodes in the spring (Darmenova et al. 2005; VanCuren et al. 2005).

7. Global detection resolutions and aerosol distributions

a. Detection resolutions and underlying surface type

Except for clean marine and polluted continental, more aerosol layers are found at the 80-km resolution than the 5- or 20-km resolution. Nearly 80% of smoke and 60% of polluted dust are found over water, whereas dust and polluted continental are found over both land and water at comparable frequencies. Clean continental is nearly equiprobable over land and water in SON and MAM, but it slightly favors land in JJA.

The land/water frequencies are not necessarily a reflection of the actual land and ocean occurrences of aerosol layers in nature but rather a reflection of where CALIPSO actually makes the measurements (i.e., if CALIPSO’s footprint has a water bias the frequencies will reflect this bias). To get a realistic land/water aerosol layer frequency, one can still use the CALIPSO measurements appropriately weighted by the land/water residence of the footprint.

b. Global distributions of dust and polluted dust aerosols

In this last section of our data analyses, we examine the global distributions of dust aerosols by the four seasons: JJA, SON, DJF, and MAM. Figure 15 shows the seasonal distributions of dust layers as a fraction of all layers in 5° × 5° grid boxes. These are the number of dust layers divided by the number of all aerosol layers found in a grid box during the season. Note the maxima in dust fractions for Saharan dust occurs in JJA and Asian dust in MAM, both according to several studies and measurements (Darmenova et al. 2005; VanCuren et al. 2005; Zender et al. 2003). The dust frequency in the western United States gradually increases from SON to DJF and reaches a maximum in MAM. However, there is some evidence of Asian dust transport to the western United States in MAM. There is more compelling evidence of Saharan dust transport to the eastern United States, especially in JJA. The pictures also present some evidence of Saharan dust transport to South America during MAM and JJA. Figure 16 is a study of the interannual variability of polluted dust during the same season. The figure is a plot of the number of layers of polluted dust found in 5° × 5° grid boxes in JJA 2006 and 2007. Notwithstanding slight differences in the number of sampling days for these two periods, the polluted dust
FIG. 14. Fractional frequencies of the aerosol types computed globally by season, surface type, and resolution of finding (5, 20, or 80 km). The seasons are (a) JJA, (b) SON, (c) DJF, and (d) MAM spanning June 2006–May 2007. The aerosol types are clean continental, clean marine, dust, polluted continental, polluted dust, and smoke. The ALL category refers to the fraction of all aerosol layers that are a given type.
distributions are remarkably similar, an expected outcome for these aerosols. Most polluted dust layers are found near deserts and biomass burning regions (West Africa and South America) or where long-range transport of dust and smoke can be realized.

c. Sensitivity of the classification scheme to algorithm threshold values

The CALIPSO measurements have a higher signal-to-noise ratio for the nighttime measurements than the daytime measurements. Although this has no direct effect on the aerosol classification scheme, it does have an effect on the number of features found by the feature finder (Vaughan et al. 2009) and the number of layers classified by the cloud aerosol discrimination algorithm (Liu et al. 2009) as aerosol layers. The day and night biases are therefore external to this algorithm. The algorithm is quite sensitive to the threshold values shown in the classification flowchart (Fig. 2). Figure 17 shows the performance of the algorithm when the threshold values are perturbed by 25% in either direction. The figure shows that when the threshold values are increased (Fig. 17a), most of the smoke aerosols are unaffected, but a few of the dust aerosols are now identified as polluted dust (i.e., a mixture of dust and smoke). Some of the polluted dust aerosols are now identified as smoke. When lower thresholds are used, the converse happens (i.e., some polluted dust layers are identified as dust and a few smoke layers are identified as polluted dust). Smoke and polluted dust both have strong absorption at 532 nm, with lidar ratios of 70 and 65 sr for smoke and polluted dust, respectively. The cross classification of these two has little effect on the retrieved optical depth. However, the misclassification of dust ($S_a = 40$ sr) as polluted dust ($S_a = 65$ sr), results in a significantly larger error ($\sim 47\%$) in the optical depth estimates. In the example shown, most of the dust layers retain their nominal classification.

![Fig. 15. Seasonal distributions of dust fractions of the aerosol layers in 5° × 5° grid boxes.](image1)

![Fig. 16. Number of polluted dust layers found in 5° × 5° grid boxes for June, July, and August of (a) 2006 and (b) 2007.](image2)
8. Summary and conclusions

Using data from the lidar measurements of CALIPSO during its first year of flight, we have examined the performance of the CALIPSO aerosol classification algorithms and the extinction-to-backscatter ratio (lidar ratio) selection schemes for the level 2 aerosol layer products. The CALIPSO model 532-nm lidar ratio values for clean continental, clean marine, dust, polluted continental, polluted dust, and smoke of 35, 20, 40, 70, 65, and 70 sr, respectively, are within the range of expected values for these types of aerosols. A comparison with the Langley high spectral resolution lidar (HSRL) resulted in a mean bias of the lidar ratio of -7.4 sr (-24.3%; HSRL lower) for the CCVEX/CATZ/ CALIPSO data and -9.2 sr (-37.5%; HSRL lower) for the GoMACCS data.

The aerosol-type distributions, when partitioned according to surface type (land/ocean), show no dependence on surface type. Polluted dust layers are found most frequently near deserts and biomass burning regions (West Africa and South America). Most of the clean continental layers are found over land with a significant fraction (35%) over the oceans. Marine, dust, polluted dust and smoke layers each account for about 20% of all aerosol layers. The clean continental fraction is about 10% and polluted continental is less than 5%. The most frequent detection resolution for smoke, dust, clean continental, and polluted dust is 80 km. The frequency of clean continental over water is enhanced by the large number of clean continental aerosol types detected over sea ice. The dust studies found that the maxima in dust fractions for Saharan dust occurs in JJA and Asian dust in MAM. An analysis of interannual variability (using JJA 2006 and 2007) of polluted dust distributions showed little interannual variation.

The similarity of the optical properties of polluted continental and smoke aerosols poses a challenge for the algorithm and is one of the limitations of the subtyping scheme. Both of these aerosol types are mainly spherical, their size distributions are dominated by the fine mode, and they are both moderately absorbing. The backscatter and depolarization signatures are not good discriminators of these two types of aerosols. This algorithm exploits the probability that smoke aerosols are likely to be lofted leaving unresolved cases where the smoke is in the boundary layer. Because the main objective of the algorithm is to select an appropriate lidar ratio, the mistyping these two types has no effect of the retrieved extinction value because both types are assigned the same \( S_a \) (70 sr). A sensitivity analysis on one scene in August 2006 shows that most of the aerosol layers retain

---

**Fig. 17.** An aerosol feature mask of a scene observed by CALIPSO on 12 Aug 2006 that shows the performance of the classification scheme when (a) high (1.25 of the nominal) and (c) low thresholds (0.75 of the nominal) are applied. (b) The nominal thresholds are the values shown in the classification flowchart in Fig. 2.
their nominal classification when the thresholds are perturbed. However, the few cases where dust is misclassified as polluted dust lead to significant errors (~47%) in the retrieved optical depth.

Recent measurements of dust lidar ratios by Müller et al. (2007) have shown that dust aerosol characteristics and lidar ratios are dependent on the source (e.g., Sahara, Gobi, Taklamakan, and Arabian). The current algorithm does not discriminate between sources. It is not clear that it should for the purpose of lidar ratio assignment, but with more information it is conceivable that a variable source dependent dust lidar ratio assignment scheme may be adopted.

The algorithm described is fully operational and generates aerosol subtypes as part of the CALIPSO level 2 products. Aerosol subtyping from spaceborne lidar measurements has been accomplished for the first time and each season of measurements provides new insights for improvements. The subtyping scheme is therefore a living algorithm that will certainly evolve with experience from CALIPSO measurements and recommendations of users.

Acknowledgments. We acknowledge the support of CALIPSO, a joint NASA CNES mission, and in particular the algorithm development and data management teams at Langley. We are also indebted to the anonymous reviewers for the many useful suggestions and constructive criticism, all of which have significantly improved this paper.

REFERENCES


