Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP)

Algorithm Theoretical Basis Document Cover Page

Final Distribution : 26 June 2025

This Algorithm Theoretical Basis Document (ATBD) is being distributed to document the initial designs of the science retrieval algorithms implemented for the Cloud-Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) mission. These algorithms represent state-of-the-art science data product retrievals from space-based lidar measurements circa 2006. For CALIPSO's final (version 5.0) data release, some of these algorithms remain operational in close to their original instantiation. Others have been significantly modified. The ATBDs should therefore be considered as historical points of departure for future developments. Updates to the original algorithms made over the years and the introduction of newly developed algorithms are fully described in the list of peer-reviewed journal articles below.

Instrument and Mission Overview

Hunt, W. H, D. M. Winker, M. A. Vaughan, K. A. Powell, P. L. Lucker, and C. Weimer, 2009: CALIPSO Lidar Description and Performance Assessment, *J. Atmos. Oceanic Technol.*, **26**, 1214–1228, https://doi.org/10.1175/2009JTECHA1223.1.

Winker, D. M., M. A. Vaughan, A. H. Omar, Y. Hu, K. A. Powell, Z. Liu, W. H. Hunt, and S. A. Young, 2009: Overview of the CALIPSO Mission and CALIOP Data Processing Algorithms, *J. Atmos. Oceanic Technol.*, **26**, 2310–2323, https://doi.org/10.1175/2009JTECHA1281.1.

Calibration

Getzewich, B. J., M. A. Vaughan, W. H. Hunt, M. A. Avery, K. A. Powell, J. L. Tackett, D. M. Winker, J. Kar, K.-P. Lee, and T. Toth, 2018: CALIPSO Lidar Calibration at 532-nm: Version 4 Daytime Algorithm, *Atmos. Meas. Tech.*, **11**, 6309–6326, https://doi.org/10.5194/amt-11-6309-2018.

Kar, J., M. A. Vaughan, K. P. Lee, J. Tackett, M. Avery, A. Garnier, B. Getzewich, W. Hunt, D. Josset, Z. Liu, P. Lucker, B. Magill, A. Omar, J. Pelon, R. Rogers, T. D. Toth, C. Trepte, J-P. Vernier, D. Winker, and S. Young, 2018: CALIPSO Lidar Calibration at 532 nm: Version 4 Nighttime Algorithm, *Atmos. Meas. Tech.*, **11**, 1459–1479, https://doi.org/10.5194/amt-11-1459-2018.

Powell, K. A., C. A. Hostetler, Z. Liu, M. A. Vaughan, R. E. Kuehn, W. H. Hunt, K. Lee, C. R. Trepte, R. R. Rogers, S. A. Young, and D. M. Winker, 2009: CALIPSO Lidar Calibration Algorithms: Part I - Nighttime 532 nm Parallel Channel and 532 nm Perpendicular Channel, *J. Atmos. Oceanic Technol.*, **26**, 2015–2033, https://doi.org/10.1175/2009JTECHA1242.1.

Vaughan, M., A. Garnier, D. Josset, M. Avery, K.-P. Lee, Z. Liu, W. Hunt, J. Pelon, Y. Hu, S. Burton, J. Hair, J. Tackett, B. Getzewich, J. Kar, and S. Rodier, 2019: CALIPSO Lidar Calibration at 1064 nm: Version 4 Algorithm, *Atmos. Meas. Tech.*, **12**, 51–82, https://doi.org/10.5194/amt-12-51-2019.

Feature Detection

Vaughan, M., K. Powell, R. Kuehn, S. Young, D. Winker, C. Hostetler, W. Hunt, Z. Liu, M. McGill, and B. Getzewich, 2009: Fully Automated Detection of Cloud and Aerosol Layers in the CALIPSO Lidar Measurements, *J. Atmos. Oceanic Technol.*, **26**, 2034–2050, https://doi.org/10.1175/2009JTECHA1228.1.

Cloud-Aerosol Discrimination

Liu, Z., M. A. Vaughan, D. M. Winker, C. Kittaka, R. E. Kuehn, B. J. Getzewich, C. R. Trepte, and C. A. Hostetler, 2009: The CALIPSO Lidar Cloud and Aerosol Discrimination: Version 2 Algorithm and Initial Assessment of Performance, *J. Atmos. Oceanic Technol.*, **26**, 1198–1213, https://doi.org/10.1175/2009JTECHA1229.1.

Liu, Z., J. Kar, S. Zeng, J. Tackett, M. Vaughan, M. Avery, J. Pelon, B. Getzewich, K.-P. Lee, B. Magill, A. Omar, P. Lucker, C. Trepte, and D. Winker, 2019: Discriminating Between Clouds and Aerosols in the CALIOP Version 4.1 Data Products, *Atmos. Meas. Tech.*, **12**, 703–734, https://doi.org/10.5194/amt-12-703-2019.

Feature Classification

Avery, M. A., R. A. Ryan, B. J. Getzewich, M. A. Vaughan, D. M. Winker, Y. Hu, A. Garnier, J. Pelon, and C. A. Verhappen, 2020: CALIOP V4 Cloud Thermodynamic Phase Assignment and the Impact of Near-Nadir Viewing Angles, *Atmos. Meas. Tech.*, **13**, 4539–4563, https://doi.org/10.5194/amt-13-4539-2020.

Hu, Y., D. Winker, M. Vaughan, B. Lin, A. Omar, C. Trepte, D. Flittner, P. Yang, W. Sun, Z. Liu, Z. Wang, S. Young, K. Stamnes, J. Huang, R. Kuehn, B. Baum, and R. Holz, 2009: CALIPSO/CALIOP Cloud Phase Discrimination Algorithm, *J. Atmos. Oceanic Technol.*, **26**, 2293–2309, https://doi.org/10.1175/2009JTECHA1280.1.

Kim, M.-H., A. H. Omar, J. L. Tackett, M. A. Vaughan, D. M. Winker, C. R. Trepte, Y. Hu, Z. Liu, L. R. Poole, M. C. Pitts, J. Kar, and B. E. Magill, 2018: The CALIPSO Version 4 Automated Aerosol Classification and Lidar Ratio Selection Algorithm, *Atmos. Meas. Tech.*, **11**, 6107-6135, https://doi.org/10.5194/amt-11-6107-2018.

Omar, A., D. Winker, C. Kittaka, M. Vaughan, Z. Liu, Y. Hu, C. Trepte, R. Rogers, R. Ferrare, R. Kuehn, and C. Hostetler, 2009: The CALIPSO Automated Aerosol Classification and Lidar Ratio Selection Algorithm, *J. Atmos. Oceanic Technol.*, **26**, 1994–2014, https://doi.org/10.1175/2009JTECHA1231.1.

Tackett, J. L., J. Kar, M. A. Vaughan, B. Getzewich, M.-H. Kim, J.-P. Vernier, A. H. Omar, B. Magill, M. C. Pitts, and D. Winker, 2023: The CALIPSO version 4.5 stratospheric aerosol subtyping algorithm, *Atmos. Meas. Tech.*, **16**, 745–768, https://doi.org/10.5194/amt-16-745-2023.

Tackett, J. L., R. A. Ryan, A. E. Garnier, J. Kar, B. Getzewich, X. Cai, M. A. Vaughan, C. R. Trepte, R. Verhappen, D. M. Winker and K.-P. A. Lee, 2025: "Mitigating Impacts of Low Energy Laser Pulses on CALIOP Data Products", *EGUsphere* [AMTD], https://doi.org/10.5194/egusphere-2025-2376.

Toth, T. D., G. Schuster, M. Clayton, Z. Li, D. Painemal, S. Rodier, J. Kar, T. Thorsen, R. Ferrare, M. Vaughan, J. Tackett, H. Bian, M. Chin, A. Garnier, E. Welton, R. Ryan, C. Trepte and D. Winker, 2025: Mapping CALIPSO Marine and Dusty Marine Aerosol Lidar Ratios using MODIS AOD Constrained Retrievals and GOCART Model Simulations; egusphere-2025-2832 - manuscript registered with Atmospheric Measurements Techniques, 2025-06-13.

Extinction and Optical Depth Retrievals

Garnier, A., J. Pelon, M. A. Vaughan, D. M. Winker, C. R. Trepte, and P. Dubuisson, 2015: Lidar multiple scattering factors inferred from CALIPSO lidar and IIR retrievals of semi-transparent cirrus cloud optical depths over oceans, *Atmos. Meas. Tech.*, **8**, 2759–2774, https://doi.org/10.5194/amt-8-2759-2015.

Hu, Y., M. Vaughan, Z. Liu, K. Powell, and S. Rodier, 2007: Retrieving Optical Depths and Lidar Ratios for Transparent Layers Above Opaque Water Clouds From CALIPSO Lidar Measurements, *IEEE Geosci. Remote Sens. Lett.*, **4**, 523–526, https://doi.org/10.1109/LGRS.2007.901085.

Ryan, R. A., M. A. Vaughan, S. D. Rodier, J. L. Tackett, J. A. Reagan, R. A. Ferrare, J. W. Hair, and B. J. Getzewich, 2024: Total Column Optical Depths Retrieved from CALIPSO Lidar Ocean Surface Backscatter, *Atmos. Meas. Tech.*, **17**, 6517–6545, https://doi.org/10.5194/amt-17-6517-2024.

Young, S. A. and M. A. Vaughan, 2009: The retrieval of profiles of particulate extinction from Cloud Aerosol Lidar Infrared Pathfinder Satellite Observations (CALIPSO) data: Algorithm description, *J. Atmos. Oceanic Technol.*, **26**, 1105–1119, https://doi.org/10.1175/2008JTECHA1221.1.

Young, S. A., M. A. Vaughan, J. L. Tackett, A. Garnier, J. B. Lambeth, and K. A. Powell, 2018: Extinction and Optical Depth Retrievals for CALIPSO's Version 4 Data Release, *Atmos. Meas. Tech.*, **11**, 5701–5727, https://doi.org/10.5194/amt-11-5701-2018.

Reflectance

Lu, X., Y. Hu, Y. Yang, M. Vaughan, Z. Liu, S. Rodier, W. Hunt, K. Powell, P. Lucker, and C. Trepte, 2018: Laser pulse bidirectional reflectance from CALIPSO mission, *Atmos. Meas. Tech.*, **11**, 3281–3296, https://doi.org/10.5194/amt-11-3281-2018.

CALIOP Algorithm Theoretical Basis Document

Part 3: Scene Classification Algorithms

Primary Authors:

Zhaoyan Liu, National Institute of Aerospace (NIA), Hampton, Virginia, USA
Ali H. Omar, NASA Langley Research Center, Hampton, Virginia, USA
Yongxiang Hu, NASA Langley Research Center, Hampton, Virginia, USA
Mark A. Vaughan, Science Applications International Corp. (SAIC), Hampton, Virginia, USA
David M. Winker, NASA Langley Research Center, Hampton, Virginia, USA

PC-SCI-202 Part 3

Release 1.0

18 October 2005

Table of Contents

1. Purpose and Scope

This document is the first release of the CALIOP (Cloud-Aerosol LIdar with Orthogonal Polarization) Level 2 Algorithm Theoretical Basis Document describing the Scene Classification Algorithm (SCA). This baseline release supplants the descriptions of scene classification found in all prior drafts of the CALIOP Level 2 ATBD.

This baseline release functions as the requirements document to the Algorithm Implementation Team for the CALIOP Level 2 scene classification algorithms. Section 2 provides an overview of the general classification approach to be used. Section 3 lists input parameters required for scene classification and the output parameters produced. Sections 4, 5, 6, 7, 8 and 9 provide discussions of the theoretical basis and design requirements for the classification algorithms that are implemented in the baseline (i.e., Build 5 or Launch Build) production code. The current document describes only the classification of features within the troposphere. Details on the classification of stratospheric features will be included in a future release.

2. Introduction

The SIBYL algorithm (see PC-SCI-202, Part 2) scans lidar profiles throughout the troposphere and stratosphere, identifies regions of enhanced scattering, and records the location and simple characteristics of these atmospheric features. The purpose of the Scene Classification Algorithm (SCA), which is actually a set of algorithms, is to classify these layers by type. In addition to being incorporated into the output data products, some of the type classifications performed by the SCA are also required by the Hybrid Extinction Retrieval Algorithm (HERA).

The functions of the Scene Classification Algorithms are indicated schematically in Figure 2.1. After the SIBYL has found a region in a lidar profile, the SCA first determines if the region is a feature or a "non-feature" based on the flag that has been generated by the SIBYL. Atmospheric features need to be discriminated between cloud and aerosol and then sub-typed by the SCA. Non-feature regions include clear air, surface, subsurface, or totally attenuated regions where no underlying feature or surface is found. If the region is surface, subsurface or totally attenuated, the SCA simply records this information in the Vertical Feature Mask (VFM). If the region is clear air, the SCA records it in VFM and selects the appropriate clear-air lidar ratio for the extinction retrieval in HERA.

If the region is a feature, the SCA checks to see if the feature is elevated (if the molecular scattering signal is available both above and below the feature for layer transmittance retrieval). For an elevated feature, the SCA will derive the lidar ratio using the transmittance-constraint method [*Fernald et al*., 1972; *Young*, 1995]. A description of this method is provided in Section 7. For both elevated and non-elevated layers, the SCA will classify the feature type and assign a lidar ratio to the feature for extinction processing in HERA. Note that if the feature is elevated and a lidar ratio can be computed using the transmittance method, the computed lidar ratio is used; if the feature is non-elevated, a lidar ratio is selected based on the model corresponding to the identified feature type. The selection of aerosol and cloud models is introduced in Sections 8 and 9, respectively.

For the feature classification, the SCA first determines if the feature is tropospheric or stratospheric by checking the base altitude of the feature. The criterion used is an altitude that is equal to the tropopause altitude + n km (n is an offset to be determined). The tropopause altitude is derived from ancillary data obtained from the Global Modeling and Assimilation Office (GMAO). If the feature base is lower (higher) than this altitude, the feature is classified as a tropospheric (stratospheric) feature.

If a feature is in the troposphere, further classifications (using four algorithms) are conducted to sub-type the feature. The SCA first determines if a layer is cloud or aerosol, primarily using the layer mean value of the 532 nm attenuated backscatter coefficient, b'_{532}, and the attenuated color ratio, c', which is defined as the ratio of the mean attenuated backscatter coefficients measured at 1064 nm and 532 nm. If the layer is classified as cloud, the SCA will then determine if it is an ice cloud or water cloud using the measured backscatter intensity and the depolarization ratio profiles, along with ancillary information such as layer height and temperature. The SCA will also use a combination of observed parameters and *a priori* information to select an appropriate extinction-to-backscatter ratio, or lidar ratio (S_a for aerosol layers, S_c for clouds), and multiple scattering function, $h(z)$, required for extinction and optical depth retrieval. To be consistent, the lidar ratio and multiple scattering function must be based on the same underlying aerosol or

cloud particle model. A constant value for the lidar ratio, as well as an array (as a function of range) for the multiple scattering function, are specified for each feature for later use in optical property retrievals.

Stratospheric features are identified but not sub-typed. Stratospheric classification is planned as a part of the post-launch upgrade of the SCA code, where the sub-typing of Polar Stratospheric Clouds (PSCs) will be an important objective. The classification criteria used for features in the stratosphere will be somewhat different than for features found in the troposphere, though the same general classification approach can be used. The strategy of algorithm development of the stratospheric feature sub-typing is further discussed in Section 5.4.

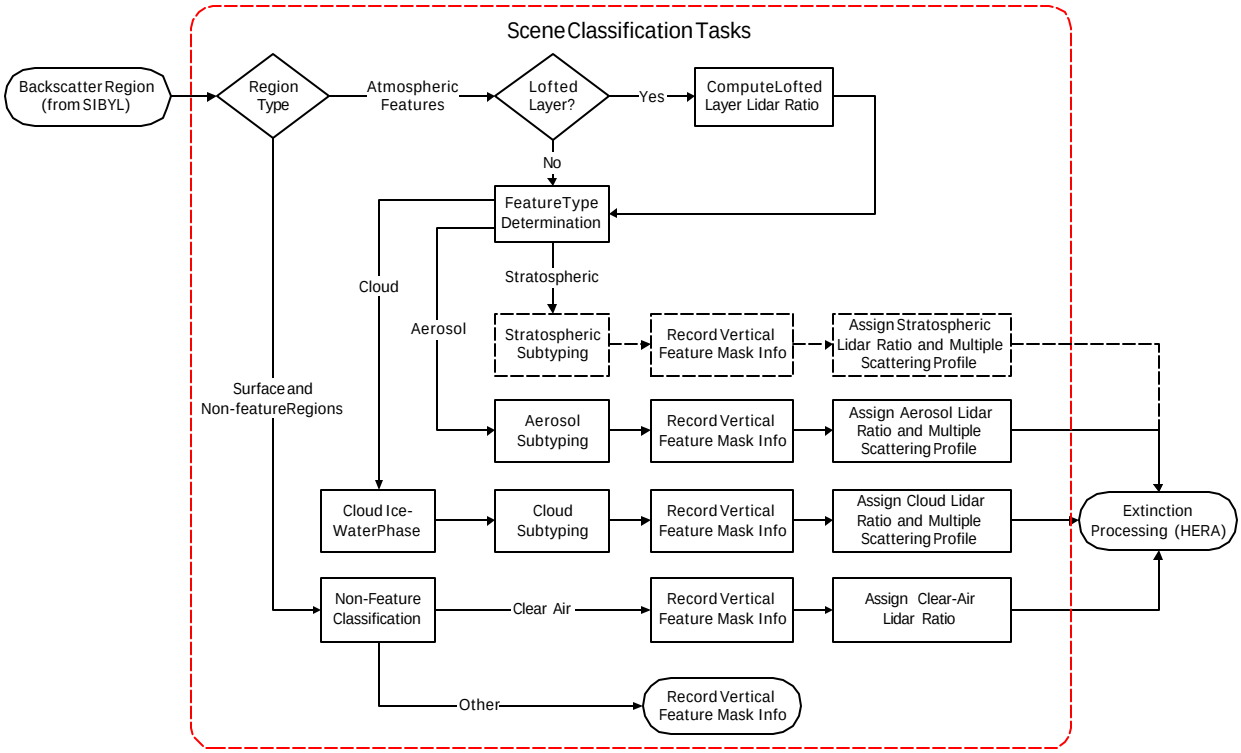


Figure 2.1: Overall flow of the baseline version of the CALIOP Scene Classification Algorithm. (dashed boxes indicate future enhancements)

3. Inputs and Outputs

3.1. Required Input Data

The layer-detection algorithm, SIBYL, computes several successive horizontal averages of lidar profiles and identifies the features which can be found at each stage of averaging. SIBYL also computes a number of simple statistics for each of the identified features. The SCA will use almost all of the outputs from SIBYL. The following bullets list the primary parameters required for the classification of each feature. A complete list of input and output parameters, as well as the sources of the input parameters, is provided in the Scene Classification Algorithm Input and Output Requirements Document (PC-SCI-202 Part 5).

- Feature base and top altitude
- Temperatures and pressures for base, top, and mid-feature altitudes
- Profiles of the 532-nm volume depolarization ratios within the feature
- Estimates of g' (integrated attenuated backscatter at 532 nm) and the uncertainty in g' for each layer
- Estimates of layer mean attenuated backscatter at 532 nm and 1064 nm and the uncertainties for each layer
- Attenuated backscatter statistics for 532 parallel, 532 perpendicular, 532 total, and 1064
- Estimates of c', layer-integrated attenuated total color ratio (1064/532) and the associated uncertainty, $\Delta c'$, for each layer
- Estimates of T_{532}^2 (two-way transmittance at 532 nm) and the uncertainty in T_{532}^2
- Geographical position (latitude and longitude)
- Knowledge of the season and local (solar) time at the footprint position
- Knowledge of the surface altitude from a digital surface elevation map
- Surface type

3.2. External Outputs

External outputs (for each feature):

- Stratospheric/tropospheric designation

External outputs (for each tropospheric feature):

- Aerosol/cloud flag
- Aerosol type/cloud type
- Classification confidence flags
- For clouds, ice/water phase flag
- Initial lidar ratio, S_a or S_c, and multiple scattering function, $\eta(z)$

4. Generalized Theoretical Basis

Scene classification is performed based on the differences in the physical and optical properties of aerosols and clouds that are reflected in lidar signals. Layers are classified using rules that associate a measured physical or optical attribute (e.g., attenuated backscatter coefficient) or set of attributes with a particular class of atmospheric scatterer. To illustrate the general characteristics of aerosol and cloud scattering, Figure 4.1 was derived using the OPAC software package [*Hess et al*., 1998a, *Hess et al*., 1998b]. The scattering models used are maritime, continental, urban, desert, Arctic and Antarctic type aerosols; stratus, cumulus and cirrus clouds; and fog. Three values of relative humidity (0%, 50%, and 90 %) are considered for the aerosols. Cirrus particles were assumed to be randomly oriented perfect or somewhat distorted hexagonal columns; otherwise, all particles are assumed to be spherical. The predicted properties are not completely accurate, as dust particles are assumed to be spherical rather than irregular, and ice crystals may have different properties than assumed here. The results are qualitatively useful, however. In general, clouds are seen to have larger backscatter coefficients and higher color ratios (~ 1) than aerosols. The exceptions to this general rule are desert aerosols and maritime aerosols under high relative humidity conditions, both of which exhibit relatively large color ratios. These scattering features can be used to distinguish aerosols from clouds. In addition, depolarization ratio is a useful indicator for identifying irregular particles and provides the means to discriminate ice clouds from water clouds and identify dust aerosols.

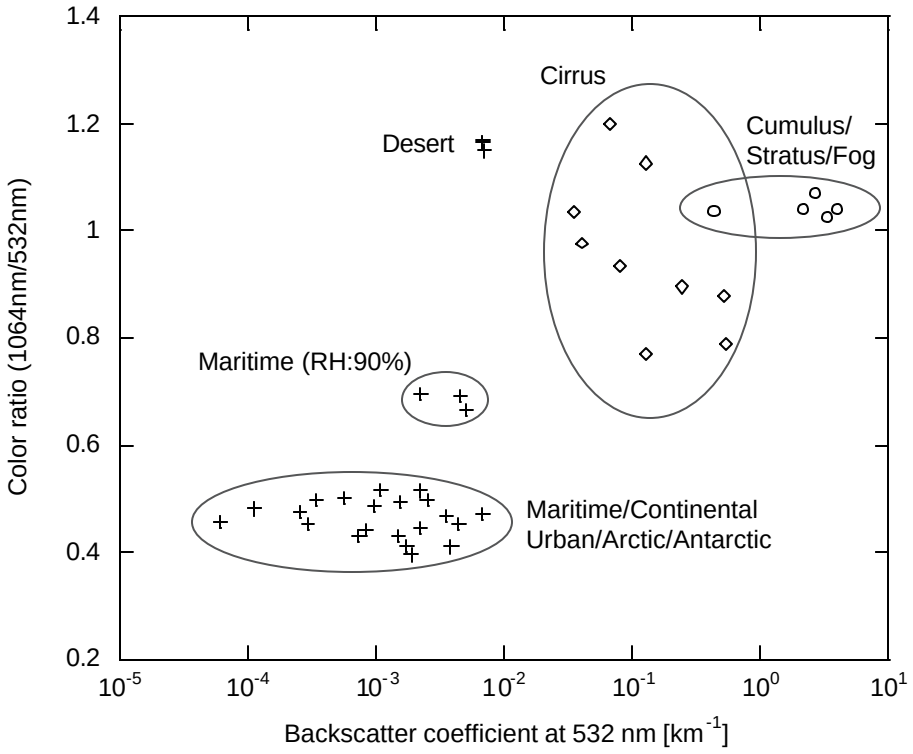


Figure 4.1: Example of modeled aerosol and cloud scattering properties calculated using the OPAC software package.

A generalized approach has been developed to assign a layer to one of two classes, based on one or more attributes. Consider the example of a hypothetical measured attribute X, and two possible classes of scatterers with which X can be associated. Functions $P_1(X)$ and $P_2(X)$ are probability density functions (PDFs) that define the probability that attribute X is associated with class 1 and class 2, respectively.

The functions P_1 and P_2 are plotted in Figure 4.2. In this hypothetical example, if the measured attribute $X < X^*$, the likelihood is very high that the measurement is associated with scatterer class 1. Similarly, if $X > X^{**}$, the likelihood is very high that the measurement is associated with class 2. If the measurement falls in the "gray" area defined by $X^* < X < X^{**}$, it could belong to either class, but, based on probability considerations, will more likely belong to one or the other. In this case, consideration of additional attributes may allow an unambiguous classification. If no additional attributes are available the algorithm uses a selection rule to assign the feature to class 1 or class 2, and estimates the level of confidence in the classification of the feature.

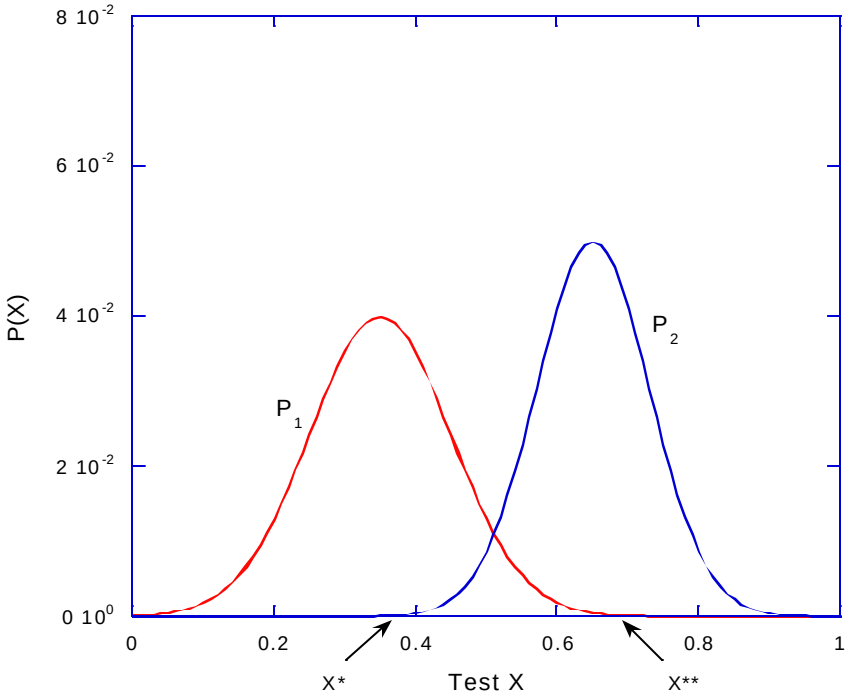


Figure 4.2: Conceptual probability distribution functions for test attribute X and two classes of scatterers.

4.1. Scene Classification Schemes

4.1.1. 1-D Histogram schemes

4.1.1.1. Single-test scene classification

In essence, classification is based on differences in the occurrence statistics of the observables of different feature types. For classification into one of two possible classes, such as aerosol and cloud, a confidence function f can be defined using scene occurrence PDF, P_1 and P_2,

$$f(X)=\frac{n_2(X)-n_1(X)}{n_2(X)+n_1(X)}=\frac{P_2(X)-P_1(X)N_1/N_2}{P_2(X)+P_1(X)N_1/N_2}. \tag{4.1}$$

Subscript 1 or 2 refers to class 1 or 2, respectively. $n_1(X)$ and $n_2(X)$ are the number of occurrences of class 1 and 2 events at a given X. N_1 and N_2 are the total occurrence number of class 1 and 2 events. The sign of f indicates the class: negative values signify class 1, and positive values signify class 2. The absolute value $Q = |f|$ provides the confidence level of a decision. The confidence level Q is related to the success rate R_s, which is the ratio of correctly classified events to the total number of events, and to the false rate,

$$R_f(X)=1-R_s(X), \tag{4.2}$$

by

$$Q=\left|f(X)\right|=2R_s(X)-1=1-2R_f(X). \tag{4.3}$$

It is derived from the relation

$$R_s(X)=\begin{cases}\dfrac{P_2(X)}{P_1(X)+P_2(X)}, & f(X)<0\\ \dfrac{P_2(X)}{P_1(X)+P_2(X)}, & f(X)>0\end{cases} \tag{4.4}$$

$$=\frac{1}{2}\left[\left|f(X)\right|+1\right]$$

The confidence level Q has a linear relation to the success (or false) rate and therefore is a good measure of the classification confidence.

4.1.1.2. Combined single-test scene classification

In general, however, there is an ambiguous region for a given test due to the PDF overlap where the scene cannot be identified clearly. The use of multiple tests is expected to resolve this ambiguity. This is because there is the possibility that, when one test is in its low confidence (overlap) region, another test may be in the high confidence region, thereby providing high confidence in the classification. There still remains the problem of how to combine these single test confidence functions to achieve optimal scene classifications. Several potential combination schemes are summarized in the following:

$$Q_{\max}=\max\left[\left|f(X_1)\right|,\left|f(X_2)\right|,\cdots,\left|f(X_m)\right|\right] \tag{4.5}$$

$$Q_{\text{sum}}=\left|\sum_{i=1}^{m}w_i f(X_i)\right| \tag{4.6}$$

where w_i is a weight for test X_i, and m is the number of available test attributes. The scheme based on Eq.(4.5) is straightforward. It makes a decision according to the test that has maximum confidence. The sign of the confidence of this test determines the class of the scene and its absolute value represents the confidence of the decision.

The scheme based on Eq.(4.6) is an alternative. It makes a decision according to the sum of properly weighted single test confidences. Again, the sign determines the scene class and the absolute value indicates the confidence of the decision.

The confidence levels for the max and sum schemes have the relationships

$$\begin{aligned} Q_{\max}(X_1, X_2, \cdots, X_m) &= \max\left[\left|f(X_1)\right|, \left|f(X_2)\right|, \cdots, \left|f(X_m)\right|\right] \\ &= 2R_s(X_{\max}) - 1 = 1 - 2R_f(X_{\max}) \end{aligned} \tag{4.7}$$

and

$$Q_{\text{sum}}(X_1, X_2, \cdots, X_m) = \left|\sum_{i=1}^{m} w_i f(X_i)\right| = 2\sum_{i=1}^{m} w_i R_s(X_i) - 1 = 1 - 2\sum_{i=1}^{m} w_i R_f(X_i)\,. \tag{4.8}$$

For the max scheme, $Q_{\max}$ is related to the success rate of a single test which has the maximum magnitude of confidence function. On the other hand, for the sum scheme, $Q_{\min}$ is related to a weighted sum of success rates of all tests. Both are however related to single-test success rates rather than the rate at $X_1, X_2, \cdots, X_m$.

4.1.2. Multiple-Dimensional Histogram Based Scheme (Multiple-Test Scheme)

This scheme is based on multi-dimensional PDFs. Let $X_1, X_2, \cdots, X_m$ be the results obtained from multiple tests. Then $p_1(X_1, X_2, \cdots, X_m)$ and $p_2(X_1, X_2, \cdots, X_m)$ are multi-dimensional PDFs of class 1 and 2 scene events, and the multiple-test confidence function is defined by

$$f(X_1, X_2, \cdots, X_m) = \frac{p_2(X_1, X_2, \cdots, X_m) - p_1(X_1, X_2, \cdots, X_m) N_1 / N_2}{p_2(X_1, X_2, \cdots, X_m) + p_1(X_1, X_2, \cdots, X_m) N_1 / N_2} \tag{4.9}$$

The ambiguity encountered in single-test or combined single-test scene classification may be improved greatly by multiple-test scene classification, because in multiple-test (dimension) space different class PDFs may separate well and therefore reduce the low confidence region, which will be discussed further below.

Similar to the single-test scene classification, the multiple-test scheme assigns the scene class according to the sign of its confidence function. Its absolute value indicates the confidence level of the decision, and it is related to the success and false rates by

$$\begin{aligned} Q(X_1, X_2, \cdots, X_m) &= \left|f(X_1, X_2, \cdots, X_m)\right| \\ &= 2R_s(X_1, X_2, \cdots, X_m) - 1 \\ &= 1 - 2R_f(X_1, X_2, \cdots, X_m) \end{aligned} \tag{4.10}$$

Similar to the single-test (one-dimensional) scheme, the confidence level Q for the multiple-dimensional scheme is linearly related to the multiple-test success rate.

4.1.3. Noise Effects

The test PDF can be widened by measurement noise. As a result, the region of PDF overlap is increased and the ambiguity becomes more severe. That is, the noise can further reduce the decision confidence.

If the distribution of test variation induced by measurement noise is $p_{\mathrm{noi}}(X)$, then the PDF of a measured test (noise-affected PDF) can be written as

$$p_n(X) = p(X) \otimes p_{noi}(X) = \int_0^\infty p(X')\, p_{noi}(X - X')\, dX' \tag{4.11}$$

where $p(X)$ is the noise-free test PDF. Figure 4.3 presents (a) artificially modeled test PDFs for two scene classes and noise-widened PDFs, and (b) the corresponding confidence functions. To simplify the simulations, a constant, signal-independent noise level is assumed. Though this assumption is not realistic, it does not materially affect the simulation study's validity in demonstrating the general concept. Gaussian distributions are used for both the noise-free test distribution and the noise-induced test variation distribution. A "noise-to-signal" ratio (NSR) is calculated as the ratio of the standard deviation of the noise and the full width at half maximum of noise-free test distribution. The NSR has a value of two in this example. In Figure 4.3(a) the solid curves are calculated PDFs and dotted curves are simulated PDFs. A normal random number generator is used to yield Gaussian-distributed noise. The noise-affected PDF is seen to be wider than the noise-free PDF, and the PDF overlap region is increased due to the PDF widening, resulting in additional ambiguity in the scene classification.

To accommodate realistic (i.e., noisy) measurements, all of the above-described schemes must be modified to use the noise-broadened PDFs. Rewriting Eqs. (4.3), (4.5), (4.6), and (4.9) by replacing all noise-free PDF terms by the corresponding noise-affected PDF, as in Eq. (4.11), we get

$$f_n(X) = \frac{p_{n,2}(X) - p_{n,1}(X) N_1 / N_2}{p_{n,2}(X) + p_{n,1}(X) N_1 / N_2} \tag{4.12}$$

$$Q_{n,\max} = \max\left[\left|f_n(X_1)\right|, \left|f_n(X_2)\right|, \cdots, \left|f_n(X_m)\right|\right] \tag{4.13}$$

$$Q_{n,\mathrm{sum}} = \left|\sum_{i=1}^{m} w_i\, f_n(X_i)\right| \tag{4.14}$$

$$f_n(X_1, X_2, \cdots, X_m) = \frac{p_{n,2}(X_1, X_2, \cdots, X_m) - p_{n,1}(X_1, X_2, \cdots, X_m) N_1 / N_2}{p_{n,2}(X_1, X_2, \cdots, X_m) + p_{n,1}(X_1, X_2, \cdots, X_m) N_1 / N_2} \tag{4.15}$$

In the presence of measurement noise, equations (4.12) through (4.15) are used in the scene classification schemes.

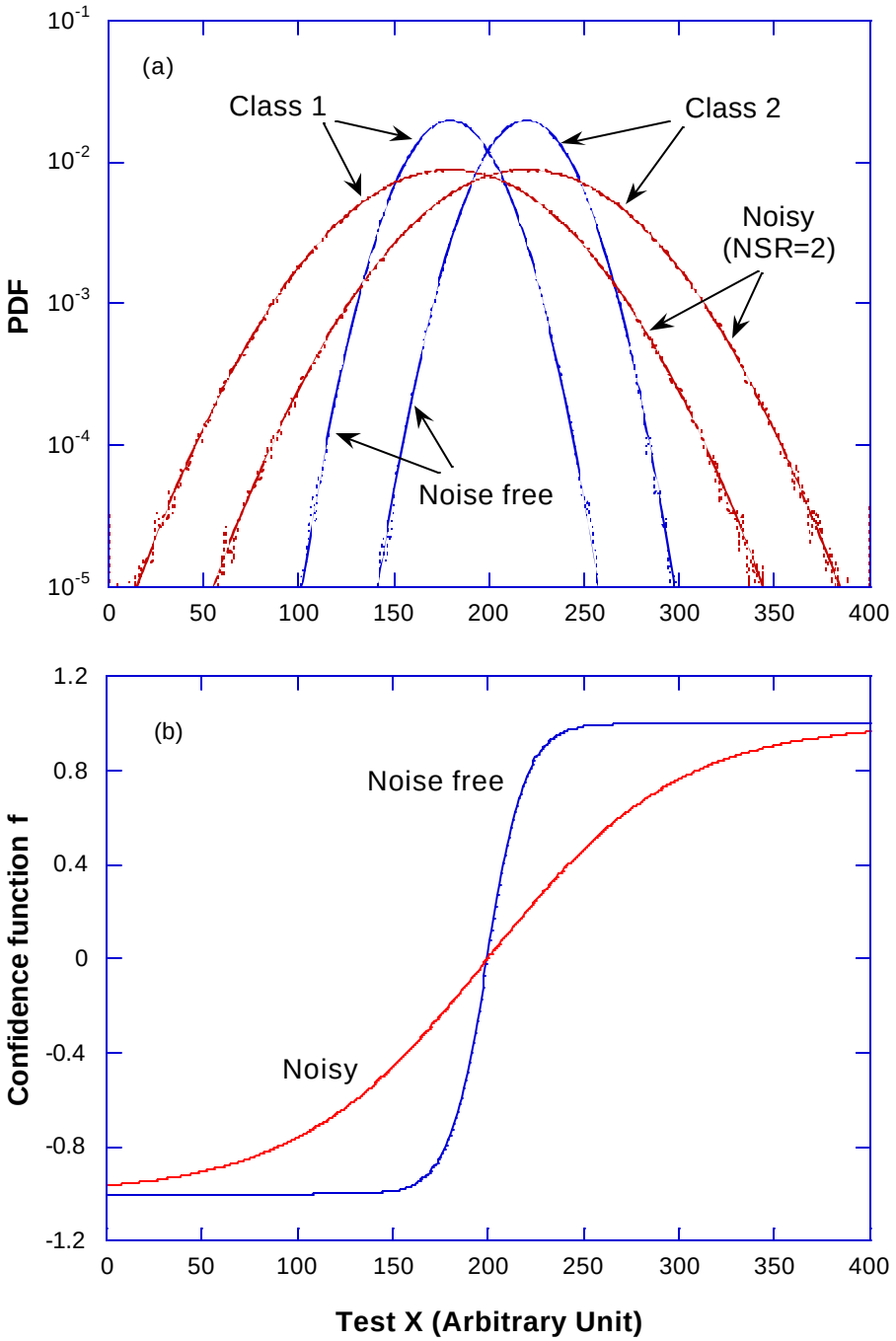


Figure 4.3: Probability distribution functions (PDFs) (a) and confidence functions (b).

4.2. Comparisons of Different Schemes

4.2.1. Schematic Comparison

In general, the multiple-dimension histogram-based method is better than any single-test or combined single-test or low-dimensional histogram-based methods. This is further discussed in

this section using a 2-D distribution (2 tests) model shown in Figure 4.4. The discussion is limited to the comparisons between the 2-D method and the single-test (1-D) methods.

There exists an ambiguous, low confidence region in the *f*-function where the class 1 and class 2 PDF curves overlap. This region is area S in the example in Figure 4.4 for the 2-D method, and includes areas B_1, B_2 and **S** for the combined single-test method, and consists of the areas A_2, B_1, B_2 and S for the single-test method using X_1 (or the areas A_1, B_1, B_2 and S using X_2). Obviously, the PDF overlap is the smallest for the 2-D method, and therefore a classification with the lowest false rate should be achieved by using a 2-D method.

Area S still remains as an ambiguous region. The 2-D confidence function method will not produce an unambiguous classification for features falling in this region. However, by increasing the number of tests – and simultaneously increasing the dimension of the confidence function – it may be possible to further reduce the ambiguous region. From a statistical point of view, the separation of different class clusters in higher dimensional space is generally more complete than in lower dimensional space. Increasing the number of dimensions should decrease the region of complete overlap between the two PDFs, thereby reducing the fraction of ambiguous classifications that are retrieved. This can be seen clearly in the example in Figure 4.4 where the 2-D cluster distributions have smaller overlap than the 1-D distributions of any test.

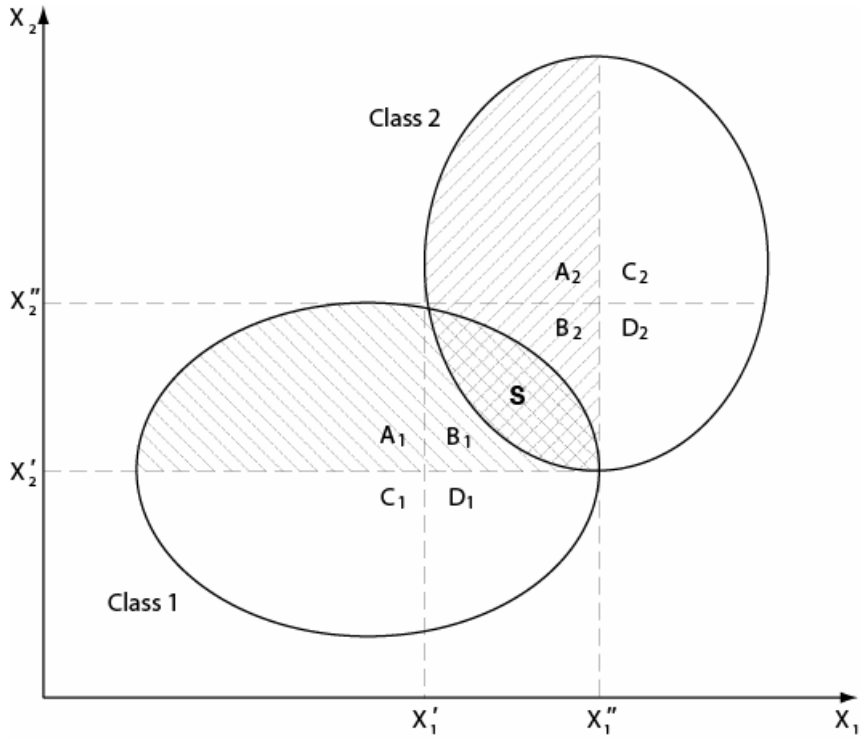


Figure 4.4: A 2-D distribution model.

4.2.2. Simulations

Simulations have been conducted to compare all of the classification schemes described above. This section presents results of simulations of two-class scene classification using artificial test

models. Contour plots in Figure 4.5 show a three-dimensional Gaussian-distributed test model. A normal random number generator is used to yield Gaussian-distributed numbers to simulate the noise (or noise-caused test variation) effects. Again, the noise-to-signal ratio, which is the ratio of standard deviations of noise distribution and noise-free test distribution, is used as a measure of noise. $1x10^8$ signals with noise in X_1, X_2 and X_3 at a level of NSR = 2 for each test have been generated.

All schemes [Eqs. (4.12) – (4.15)] are then tested by applying them to the generated test signals. We define the false classification rate as the ratio of the number of incorrectly classified events to the total number of events at a given test value. Figure 4.6 presents simulated results of false rate as a function of X_1 using different schemes. We have tested with different weight combinations for the sum scheme in the simulations. The equal weight combination (i.e., w_1:w_2:w_3=1:1:1) showed the best classification results (lowest false rate), and we focus our discussion on the equal weight case for the sum scheme.

For the single test scheme, Figure 4.6 shows that, at the point where the PDFs of the two classes cross and are equal (X_1=200), R_f = 0.5; i.e., half of the decisions are wrong, which corresponds to a confidence level of 0. This result cannot be overcome by using any single test alone. However, classification near the crossing point can be improved significantly by using multiple tests. As more tests are used, better classification results can be derived. In general, 3-D results are better than 2-D, and 2-D results better than single-test (1-D).

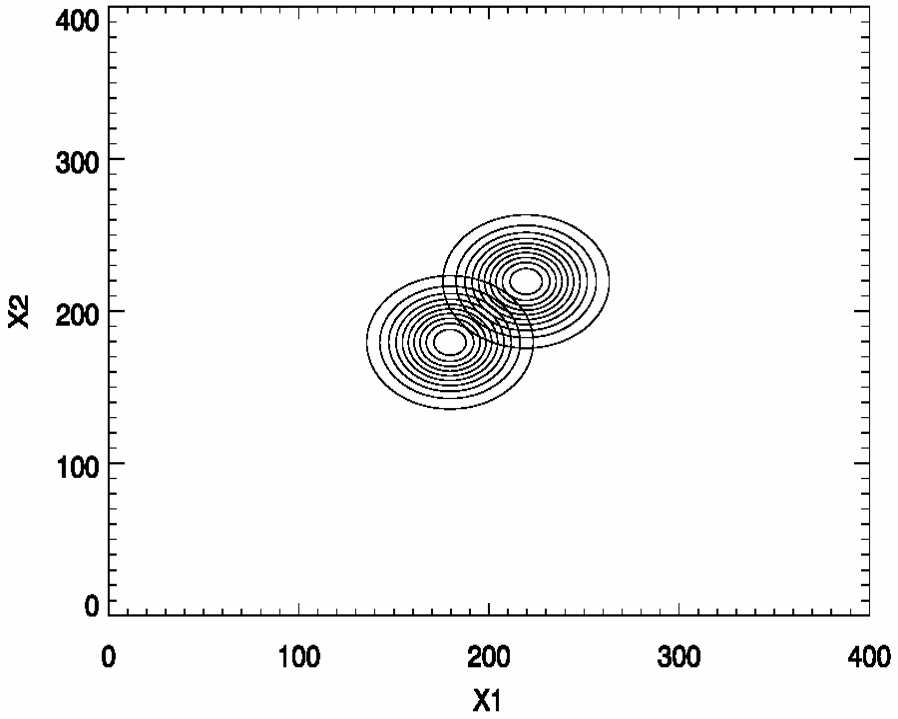


Figure 4.5: X_1-X_2 space contour plots of a 3-D Gaussian-distributed test model for simulations.

We note that, for the model used, the sum scheme with equal weights produces the same classification results as the 3-D scheme. We also note that for some cases the 3-D scheme and the sum scheme produce the same classification results. However, the 3-D scheme has a better

capability for accurately assigning the confidence level than that for the sum scheme (or any combined single-test scheme). The confidence level Q should reflect correctly the success (or false) rate of classifications. As discussed above, Q for the combined single-test schemes (Eq.(4.6) and (4.7)) can only be related to a single-test success rate, or to a weighted sum of all single-test rates, rather than to the success (or false) rate at $[X_1, X_2, \cdots, X_m]$ in the multiple-test space, as the multiple-dimensional scheme does (Eq.(4.9)). To show this, we present simulated false rates as a function of X_1 using the 3-D and sum schemes in Figure 4.7 when X_2=180 and X_3=200, along with the theoretical curves computed using Eq. (4.7) and (4.9). It is clearly shown that the sum scheme overestimates the false rate (except X_1=220), whereas the 3-D scheme produces correct estimates.

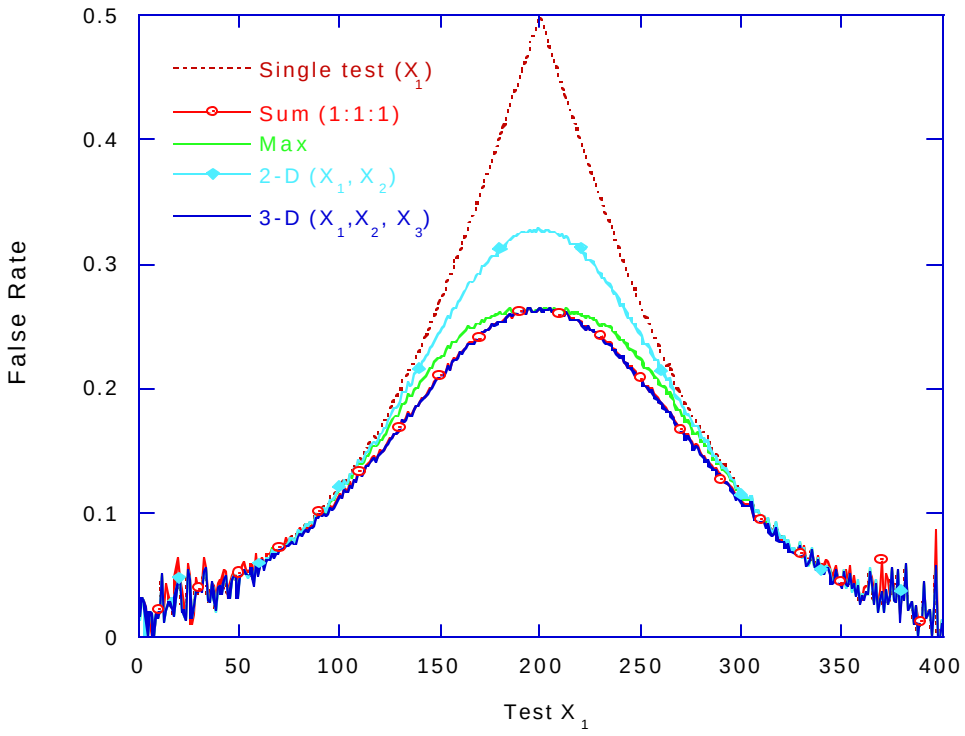


Figure 4.6: False rates of classifications using the single-test (1-D), sum, max, and 2-D and 3-D schemes.

In general, the multiple-dimensional scheme produces better classifications and assignments of confidence levels than combined single-test schemes that use the same number of tests. Furthermore, the multiple-dimensional scheme generates better results when executed in a higher dimensional space (i.e., when using a larger number of tests). For the CALIOP cloud and aerosol discrimination, we employ a 3-D scheme as described below.

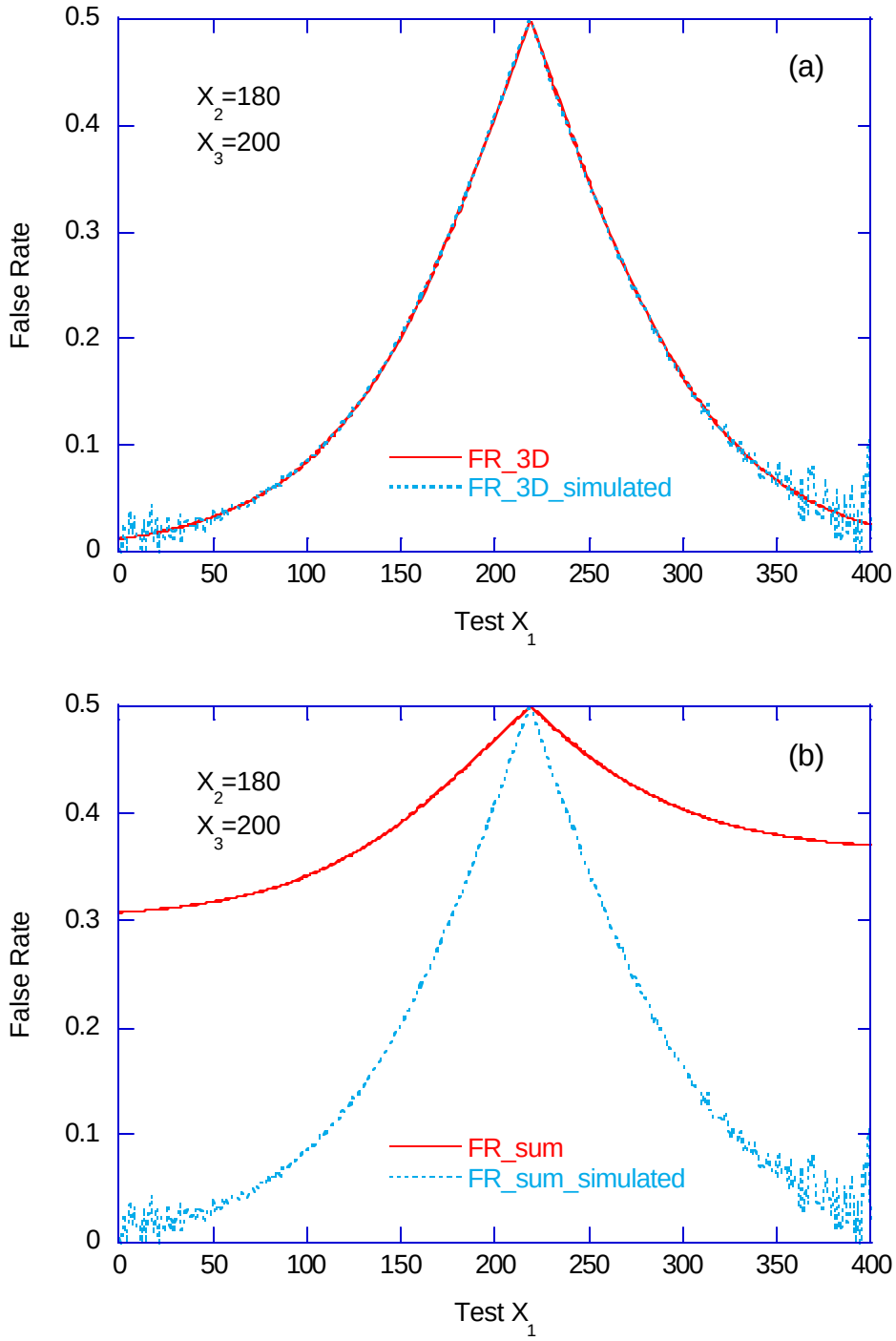


Figure 4.7: Simulation results of false rates as a function of test X1 for the 3-D scheme (a) and sum scheme with equal weight (b) along with theoretical curves from Eq. (4.8) and (4.10).

5. Discriminating Aerosols from Clouds

A 3-D operational algorithm has been developed for the CALIOP data processing [*Liu et al.*, 2004], and tested with data from the Cloud Physics Lidar (CPL) [*McGill et al.*, 2002] and the Lidar In-space Technology Experiment (LITE) [*Winker et al.*, 1996]. Test attributes included are the layer-averaged attenuated backscatter $\overline{b'}$ at 532 nm, the layer-integrated attenuated volume color ratio c', and the layer center altitude z. The layer-integrated volume color ratio is the ratio formed by dividing the layer-integrated attenuated backscatter at 1064 nm by the layer-integrated attenuated backscatter at 532 nm. The limits of the integration are from layer top to layer base. The layer center altitude is the average of the detected layer top and base altitudes. Including height information with the layer optical properties should increase the effectiveness of any cloud-aerosol separation scheme. This is because the optically thinner clouds that contribute most to the overlap of PDF with aerosols generally appear in higher altitudes, whereas aerosols concentrate mostly in the planetary boundary layer (PBL). Details are described below.

5.1. Approximation of noise distribution

To compute the noise-broadened PDFs, distributions of the variations in the test attributes b' and c' due to noise must first be derived or estimated. The CALIOP 532 nm and 1064 nm receiver channels use, respectively, photomultiplier tubes (PMTs) and an avalanche photodiode (APD), all operated in analog detection mode. The Gaussian distribution is a good approximation for the multiplication process for both PMTs and APDs. The algorithm therefore assumes Gaussian distributions for the noise-induced variations in b' at both 532 nm and 1064 nm. If we designate b'_0 and c'_0 as the measured value of b' and c', then the probability that the true value of b' is actually measured as b'_0 is given by

$$p_{noise}(b'_0, b') = \frac{1}{\sqrt{2p}\, s_{b'}} \exp\left[-\frac{(b'_0 - b')^2}{2 s^2_{b'}} \right] \tag{5.1}$$

where

$$s_{b'} = \frac{1}{N_{avg}} \sqrt{\frac{r^2 b'}{C} NSF^2 + \left(\frac{\overline{\Delta P}_{BG}}{N_{shot} G_A} \right)^2} \tag{5.2}$$

and C is the lidar calibration constant, r is the mean range from lidar to the feature, $\overline{\Delta P}_{BG}$ is the background noise in science digitizer readings averaged over a number of laser pulses equal to N_{shot} (corresponding to the horizontal resolution), G_{A} is the variable gain, and NSF is the noise scale factor for the gain-normalized digitizer reading. For a description of NSF, see the CALIOP Level 1 ATBD (PC-SCI-201).

The noise distribution for c' can then be derived from the Gaussian noise distributions of b' at 532 nm and 1064 nm wavelengths:

$$p_{noise}(c', c_0') = \int_0^\infty p_{noise}(b_{532}', b_{532,0}')\, p_{noise}(c'b_{532}', b_{1064,0})\, b_{532}'^2\, db_{532}'$$
$$= \frac{1}{2p\, s_{b_{532,0}'}\, s_{b_{1064,0}'}} \int_0^\infty \exp\left[-\frac{(b_{532}' - b_{532,0}')^2}{2s_{b_{532,0}'}^2} - \frac{(c'b_{532}' - b_{1064,0}')^2}{2s_{b_{1064,0}'}^2} \right] b_{532}'^2\, db_{532}'. \quad (5.3)$$

5.2. Description of Operational Tropospheric Algorithm

Figure 5.1 presents a flowchart of the CAD algorithm. The algorithm first selects noise-free PDF files from an existing PDF database that has been compiled based on previous measurements. The stored PDF files have a 100x100x20 array structure: Dimension 1 (*ib*=1:100) is test $Xb = \ln b_{532}'$, starting from Xb_{start} = -12, with an increment ΔXb = 0.014 [$m^{-1}sr^{-1}$]; Dimension 2 (*ix*=1:100) is $Xx = c'$, starting from 0, with ΔXx = 0.02; and Dimension 3 (*iz*=1:20) is Xz, starting at 0, with ΔXz =1 [km]. Here *Xb*, *Xx*, and *Xz* represent the test values, and *ib*, *ix*, and *iz* represent the index of the corresponding test (*Xb*, *Xx*, *Xz*) into the PDF array. For each feature, layer products b_{532}', b_{1064}', c', and z are read as input along with estimated uncertainties $\Delta b_{532}'$, $\Delta b_{1064}'$ and $\Delta c'$. In the following description of the algorithm, subscript 0 is used to indicate these measured values, i.e., $b_{532,0}'$, $b_{1064,0}'$, c_0', and z_0, and the corresponding index is $ib_0 = Int(\ln(b_{532,0}')/\Delta Xb)$, $ix_0 = Int(c_0/\Delta Xx)$, $iz_0 = Int(z_0/\Delta Xz)$, where *"Int"* represents conversion to an integer value.

Assuming the *z* values are noise free, the computation is essentially a 2-D (b' and c') problem. The noise-affected confidence function, $f_n(b_{532,0}', c_0', z_0)$, can then be computed as follows. First, the 2-D noise-free PDFs for $iz = iz_0$, $P_{aerosol}(ib = 1{:}100, ix = 1{:}100, iz_0)$ and $P_{cloud}(ib = 1{:}100, ix = 1{:}100, iz_0)$, are selected. PDFs affected by noise in *b*' at $ib_{l,0} = Int(\ln(b_{l,0}')/\Delta Xb)$ and iz_0, $P_{aerosol}(ib_0, ix = 1{:}100, iz_0)$ and $P_{cloud}(ib_0, ix = 1{:}100, iz_0)$, are then computed using

$$P_{j,n}(ix) = \frac{\Delta Xb}{\sqrt{2p}\, s_{ib}} \sum_{ib=1}^{100} P_j(ib, ix, iz_0)\, b(ib) \exp\left(-\frac{[b(ib) - b(ib_0)]^2}{2s_{ib}^2} \right) \quad (5.4)$$
$$b(ib) = \exp\left(Xb_{start} + ib \cdot \Delta Xb \right)$$

where subscript *j* indicates "aerosol" or "cloud". s_{ib} can be computed using Eq.(5.2). The noise distribution for c' can then be computed numerically [Eq.(5.3)] using,

$$p_{\chi',n}(ix) = \begin{cases} \dfrac{1}{2\pi\Delta\beta_{532,0}'\Delta\beta_{1064,0}'} \displaystyle\sum_{i=1}^{2\cdot ix_0} \exp\left[-\frac{(i \cdot dBeta - \beta_{532,0}')^2}{2(\Delta\beta_{532,0}')^2} - \frac{(i \cdot dBeta \cdot ix \cdot \Delta Xx - \beta_{1064,0}')^2}{2(\Delta\beta_{1064,0}')^2} \right] (i \cdot dBeta)^2\, dBeta, & \chi_0' < 2 \\ \dfrac{1}{\sqrt{2\pi}\Delta\chi'} \exp\left(-\dfrac{[(ix - ix_0)\Delta Xx]^2}{2(\Delta\chi')^2} \right) dBeta. & \chi_0' \geq 2 \end{cases} \quad (5.5)$$
$$(\text{ix} = 1, 2, \cdots, 200)$$

Where $dBeta = 2\max[b'_{532}, b'_{1064}, \Delta b'_{532}, \Delta b'_{1064}]$. The color ratio dimension of the PDF files covers a range of [0, 2). The real (1064/532) color ratio of natural scatterers that lidar measures is generally in this range. However, there exists the possibility that some measured features have a color ratio outside this range. Negative values could be frequently observed for faint scatterers (mostly aerosols) because the backscattered signal from these features at 1064 nm is relatively small and noisy, which may result in negative value of backscatter. For features having negative color ratio, c' can be set to zero. In practice, however, the code sets it to be 0.02, as required by the programming (to avoid returning empty call). A value of color ratio larger than 2 may be derived due to system characteristics such as the different transient response (or saturation) of the detectors used in the two channels. In this case c' is set to 1.98 (ix_0=100). Noise distribution for the color ratio (a 200-element array) is computed covering a range from 0 to 4. If the color ratio is larger than 2, a Gaussian distribution centered at ix_0=100 is computed.

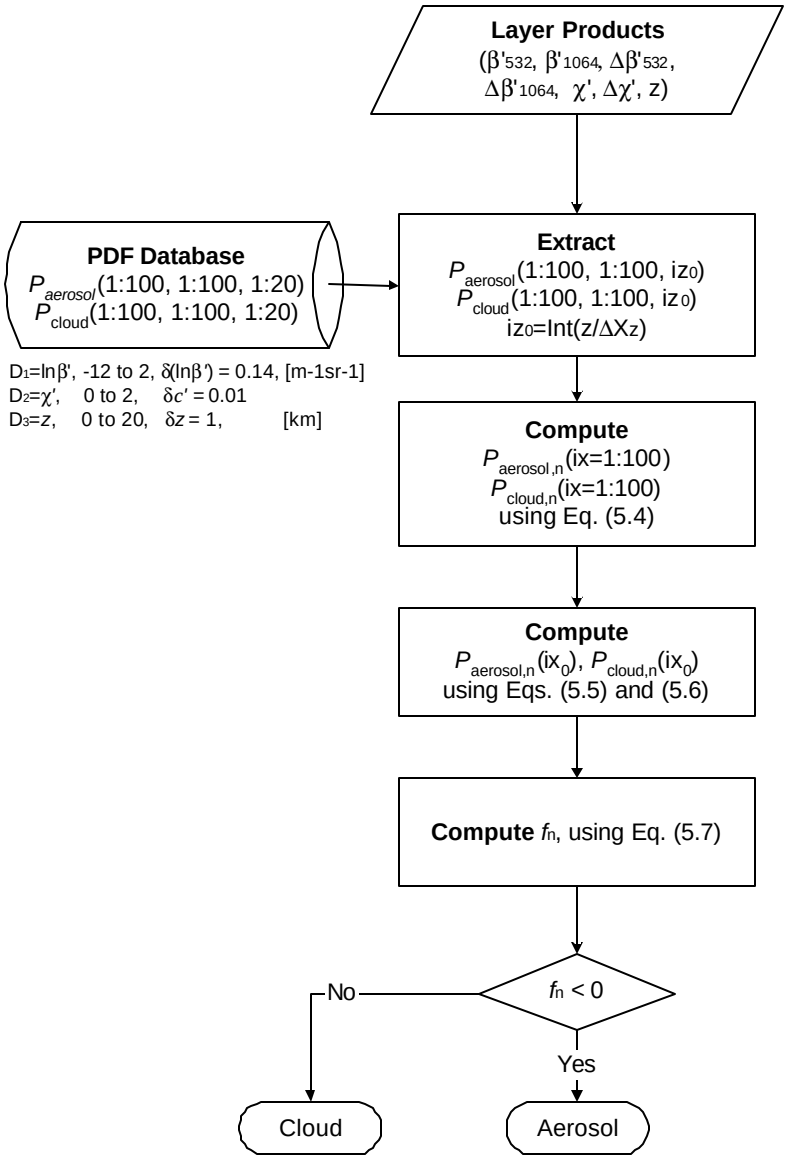


Figure 5.1: Flowchart of the 3-D CALIOP CAD algorithm.

The noise-affected PDFs can be derived using

$$P_{j,n} = \begin{cases} \sum_{ix=1}^{2\cdot ix_0} P_{j,n}(ix)\cdot p_{x',n}(2\cdot ix_0 - i), & ix_0 \le 50 \\ \sum_{ix=1}^{100} P_{j,n}(ix)\cdot p_{x',n}(2\cdot ix_0 - i). & ix_0 > 50 \end{cases} \quad \text{(j = "aerosol" or "cloud")} \qquad (5.6)$$

Finally, the noise-affected confidence function is computed using

$$f_n = \frac{P_{2,n} - P_{1,n}}{P_{2,n} + P_{1,n}}. \qquad (5.7)$$

If the computed value of confidence function is negative, the feature is classified as aerosol; if the computed value is positive, the feature is classified as cloud; and if the computed value is zero, the feature is classified as “undetermined”.

Classification results will be further processed by the aerosol typing algorithm if the feature is classified as an aerosol or by the phase classification algorithm if the feature is classified as a cloud. The details for these additional classification algorithms are described in Sections 6, 8, and 9.

5.3. Algorithm Tests

The Cloud Physics Lidar (CPL) is a three-wavelength polarization-sensitive airborne lidar system [*McGill et al.*, 2002] that detects the perpendicular and parallel components of the backscatter profile at 1064 nm, and total backscatter profiles at 532 nm and 355 nm. Included among the comprehensive suite of CPL post-flight data products are feature boundaries, layer type (i.e., cloud or aerosol), and range-resolved extinction and backscatter coefficients. The feature types were classified by the GLAS prototype layer discrimination algorithm [*Palm et al*. 2002]. The CPL datasets are thus an ideal data source for CALIOP algorithm tests, as direct, straightforward comparisons can be made between the CALIOP test results and the CPL data products. A dataset acquired during the 2003 THORPEX-PTOST campaign conducted in Honolulu, Hawaii from February 18 to March 14, 2003 was used to test the performance of the cloud-aerosol discrimination algorithm. During THORPEX-PTOST, the CPL acquired backscatter data on nine scheduled ER-2 flights. The CALIOP algorithm test set consists of all of the data acquired during these flights. In addition, all data acquired on the ER-2 transit flight from California to Hawaii is also used. In total, the test set consists of approximately 49 hours of CPL measurements.

Figure 5.2(a) presents an example of 532 nm attenuated backscatter signals acquired by CPL. The data shown was acquired between 19:45:32 and 20:15:26 on February 19, 2003. Four high cloud layers above 5 km are observed. PBL aerosols as well as some low broken clouds are also seen. An optically thin aerosol layer exists above the PBL and below 5 km. Layer boundaries have been determined by a threshold-based feature finder incorporated into the CPL data analysis software. Feature locations, together with the classifications assigned by the GLAS algorithm are shown in Figure 5.2(b). Colors indicate feature type: red is aerosol; blue is cloud. Most layers have been successfully detected, excepting only those optically very thin layers or layers obscured by overlying, optically dense cloud layers. Application of the CPL layer finder

to the entire 10-flight dataset yields a total of 228264 features, all of which are used as input to test the CALIOP cloud-aerosol discrimination algorithm.

For each layer, the mean attenuated 532 nm backscatter, b'_{532}, is derived by averaging the calibrated and range-corrected 532 nm signals over the detected layer upper and lower boundaries. The 1064 nm to 532 nm volume color ratio, c', is obtained from the ratio of the averaged attenuated 1064 nm and 532 nm backscatters, b'_{1064} and b'_{532}. Uncertainty is calculated for each parameter based on error propagation theory [*Bevington and Robinson*, 1992]. Random errors due to noise are dominant when features are optically thin. Only random error is taken into account. Measured layer altitudes are assumed to be accurate.

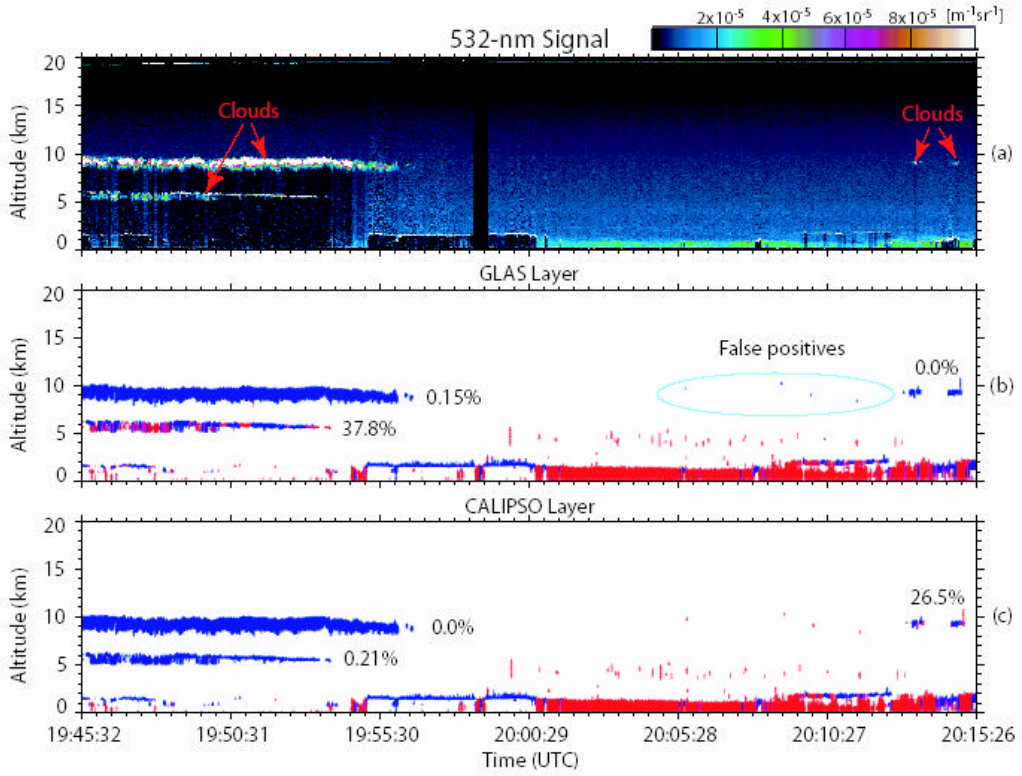


Figure 5.2: An example of cloud-aerosol classification.

A PDF database was developed using lidar observations and model data. Consistent with the measurements of *Beyerle et al.* [2001], the mean attenuated volume color ratio for clouds is characterized using a normal distribution. Based on the author's analyses of the LITE data, a normal distribution was also selected for use with aerosol color ratios. The initial means and standard deviations for the color ratio PDFs were derived from model studies using OPAC. That normal distributions represent appropriate approximations for both cloud and aerosol color ratio distributions can be further verified by inspection of the CPL THORPEX observations. A lognormal distribution has been employed to represent the mean attenuated backscatter coefficient for aerosols. LITE data analyses as well as other lidar measurements,including those acquired at numerous EARLINET stations, show that aerosol backscatter can be well characterized by a lognormal distribution [Post, et al., 1982; *Matthias and Bösenberg*, 2002; *Bösenberg et al.*, 2001]. The aerosol backscatter distribution parameters used in this work were derived from the LITE aerosol measurements. For cloud backscatter coefficients we employ a

multi-modal distribution retrieved from the LITE cloud data. In order to automatically accommodate the huge amount of new data that will be collected during the lifetime of CALIOP, the distribution parameters for both clouds and aerosol can be adjusted iteratively during the classification process. This iterative improvement procedure was employed during the tests with the THORPEX-PTOST data set.

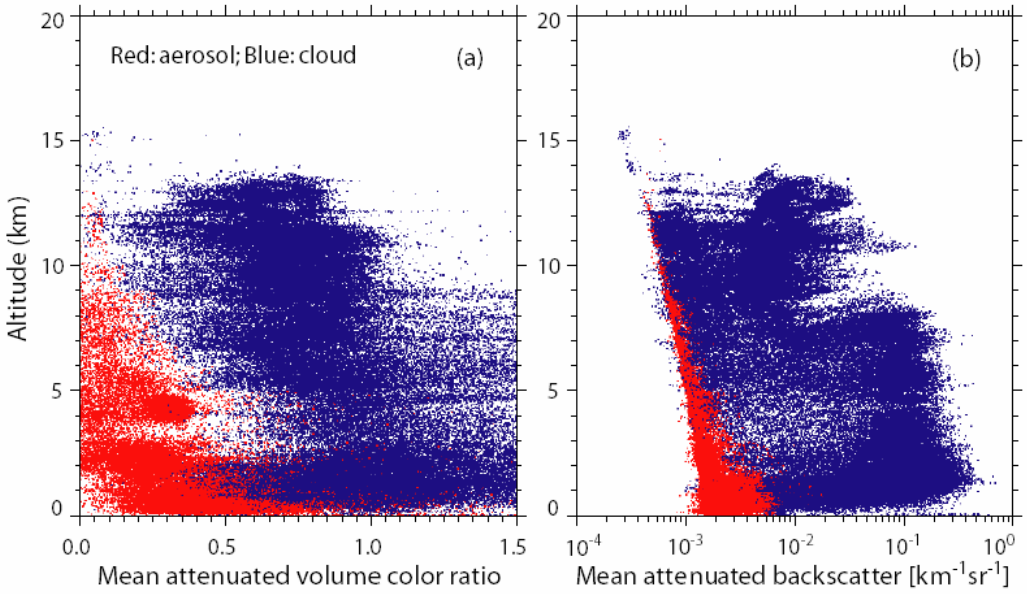


Figure 5.3: Scatter plots of mean attenuated volume color ratio and backscatter of nearly 230,000 features observed by CPL during the THORPEX-PTOST mission.

The results produced from the CALIOP CAD algorithm are quite consistent with the GLAS results: only 5.7% of total 228,264 features have been classified as different type by the GLAS and CALIOP algorithms. Our case studies (e.g., Figure 5.2) showed, however, that the CALIOP CAD algorithm can provide better classification for high optically thin clouds. This is because a better separation of cloud and aerosol clusters can be achieved in this b'-c'-z space, and the degree of separation of cloud and aerosol clusters is an essential limit on the performance of any scene classification scheme. Figure 5.3 presents scatter plots of the mean attenuated backscatter and volume color ratio of the features observed by CPL during the THORPEX-PTOST mission. Good separation between clouds and aerosols in the volume color ratio-altitude space is seen above ~1.5 km. In addition, the mean attenuated volume color ratio is less sensitive to the attenuation of overlying clouds, because the cloud scattering has weak wavelength dependence in the visible and near-infrared region and, as a result, taking ratio of the two attenuated backscatter values can cancel the effect of the attenuation of overlaying clouds.

5.4. Algorithm Enhancement Strategy

The 3-D CAD algorithm described above is used for the CALIOP cloud-aerosol discriminations in the troposphere. In the future, additional classification schemes will be added for the stratosphere, The CAD algorithm for the stratosphere will be somewhat different from the tropospheric algorithm, because there are polar stratospheric clouds (PSCs) in the polar region.

There exist three major types of PSCs: types Ia, Ib, and II [*Browell et al.*, 1990]. Type Ib PSCs have low backscatter and low depolarization, and usually consist of liquid $HNO_3/H_2SO_4/H_2O$ particles. The other two types of PSCs, on the other hand, have high depolarization ratios. Type Ia PSCs are composed mostly of nitric acid trihydrate (NAT) crystals. Type II PSCs consist mainly of water ice particles (and are therefore very similar to cirrus clouds). A two-step algorithm for distinguishing between these classes can then be proposed as follows:

- By testing the volume depolarization ratios using the 1-D scheme, the features in the stratosphere can be separated first into two classes. Class one will contain all aerosols and type Ib PSCs, and class two will include type Ia and type II PSCs.
- Subtyping within each of these two classes (e.g., discrimination between aerosols and type Ib PSCs, and between type Ia and type II PSCs) will be accomplished using the same 3-D algorithm that is currently used for tropospheric classification. The PDF files required for stratospheric discrimination will be developed when CALIOP polar data becomes available after the CALIPSO launch.

The volume depolarization ratio is a useful indicator for identifying irregular particles [e.g., *Sassen*, 1991], and it provides the means to discriminate between ice clouds and water clouds and between dust layers and spherical aerosols. As described above, depolarization ratios can be used for stratospheric scene classification. Furthermore, it should also prove useful in the identification of diamond dust in the polar region. Depolarization is also a candidate for inclusion in the parameter space used for future improvement of cloud-aerosol discriminations. At lower altitudes, the depolarization ratio should be especially helpful in distinguishing dense desert dust aerosols from water clouds and diamond dust (consisting of nonspherical ice crystals) from arctic haze. Nonspherical dust particles can produce high depolarization ratios [*Gobbi et al.*, 2000; *Murayama et al.*, 2001], as can ice crystals, whereas, when multiple scattering can be neglected, the spherical droplets in water clouds generally yield no depolarization. However, we note here that for space-borne lidars the depolarization due to the multiple scattering in dense water clouds could be an issue [*Hu et al.*, 2001]: the high altitude of the orbit (e.g., 705 km for the CALIPSO satellite) results in a large footprint on the ground even for the small receiver field of view. Introducing a depolarization ratio test into the CALIOP algorithm would thus require the development of PDFs for cloud and aerosol depolarization ratios that correctly account for the effects of multiple scattering. Unfortunately, as neither LITE nor GLAS was configured as a polarization-sensitive instrument, no space-based measurements of depolarization ratio are available for testing such PDFs. Therefore, it is currently difficult to make reliable assessments regarding the utility of an additional depolarization ratio test. However, revisiting the issue will become possible once sufficient data has been collected by the CALIPSO mission itself.

6. Discriminating Water and Ice Clouds

This section introduces the cloud ice/water phase discrimination algorithm in the CALIOP baseline code, and discusses studies of mixed phase cloud identification. The CALIOP baseline Ice-Water Algorithm (IWA) will primarily use layer integrated depolarization ratio, d, and temperature to classify cloud phase as ice or water. We begin by discussing several issues which must be considered in developing an ice-water phase algorithm

6.1. Theoretical background

6.1.1. Particle Depolarization

The transmitted lidar beam is nearly 100% linearly polarized. Depolarization of the return signal from the molecular atmosphere is about 0.35% due to the narrow optical bandwidth of the receiver which blocks most of the highly polarized Stokes and anti-Stokes lines. It is well known from ground-based and airborne observations that backscattering from ice crystals results in appreciable signal in a polarization plane perpendicular to the plane of the transmitted lidar beam (*Sassen*, 1991). Depolarization from ice crystals depends on crystal shape and aspect ratio but is typically in the range of 30%-50%. Lower values can be seen when horizontally oriented particles are present (*Sassen* and *Benson*, 2001). In contrast, backscattering from spherical water droplets preserves the polarization of the incident light. However, if a water cloud is optically thick, multiple scattering can give rise to appreciable depolarization. Monte Carlo modeling has been used to study the details of multiple scattering in ice and water clouds. Preliminary results shown below indicate ice/water phase can be distinguished even for dense cloud by considering the shape of the depolarization ratio profile.

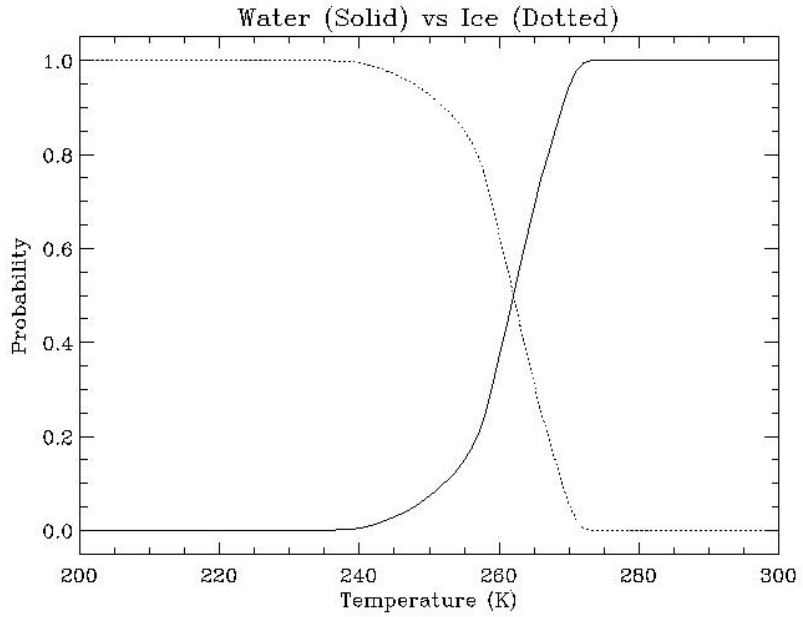


Figure 6.1: Cloud temperature statistics from MODIS cloud products.

6.1.2. Cloud Temperature Test

In cases where classification using the depolarization signal is ambiguous, it may be possible to perform the classification using cloud temperature. The lidar profile provides unambiguous cloud height, and the temperature is then determined using temperature profiles from a gridded analysis. Statistics on the observed frequency of cloud ice and water as a function of cloud temperature can be derived from various observations, such as MODIS cloud measurements (see Figure 6.1). Cloud temperature can be estimated from the observed cloud height and temperature profiles obtained from gridded synoptic analysis products. If the temperature at cloud base is lower than -45°C, it can be assumed the cloud is an ice cloud [*Pruppacher*, 1995]. If the temperature at cloud top is higher than 0°C, then it can be assumed that the cloud is a water cloud. If the top/base temperatures fall between these extremes, either ice, water, or a mixture of the two may exist.

6.1.3. Evaluating Multiple Scattering Effects

Compared with ground-based lidars, the footprint size (on the order of 100 meters or more) for space-based lidars is large. A significant portion of the lidar signal from a cloud can come from multiple scattering. A Monte Carlo model with full-Stokes vectors has been developed to quantitatively evaluate the impact of the multiple scattering on cloud phase discrimination [*Hu et al.*, 2001] and the simulations shown in Figure 6.2 were based on this model. As demonstrated by these results, multiple scattering increases the depolarization ratio, especially for water clouds, and thus creates difficulties in cloud phase discrimination.

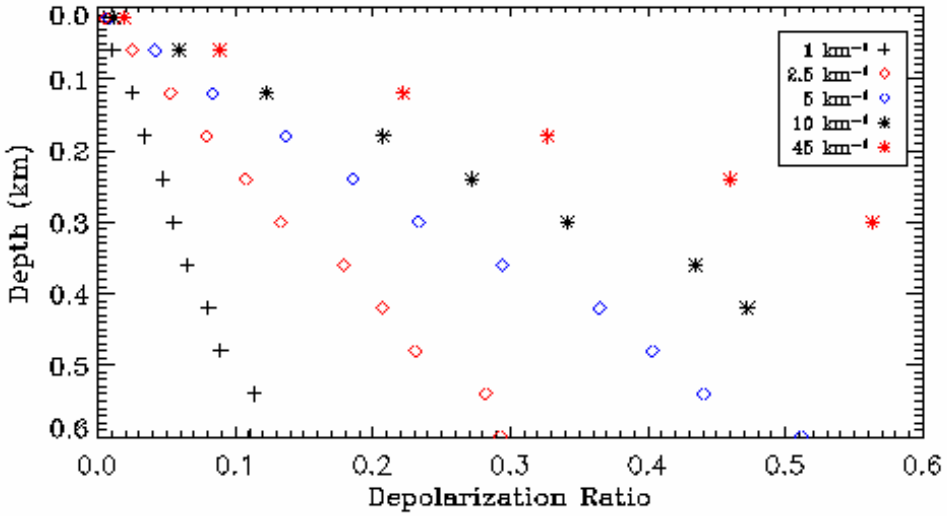


Figure 6.2: Depolarization of water cloud backscattering due to multiple scattering.

6.1.4. Mixed-Phase Clouds

Mixed-phase clouds are common, especially in the Arctic [*Shupe* et al., 2001)]. While many mixed-phase clouds are optically dense and cannot be fully profiled with lidar measurements, some mixed-phased clouds, having a thin-layer of supercooled water above ice [*Rauber* and *Tokay*, 1991], can be clearly identified by CALIOP. While a cloud layer of pure water or pure ice

can be identified from the layer-integrated depolarization ratio, identification of mixed-phase cloud requires vertical profiles of depolarization. For example, we need to look for a discontinuity in depolarization (from zero to more than 30%) for a thin layer of supercooled water on top of cold ice clouds.

6.1.5. Anomalous Backscatter for Oriented Particles

Clouds composed primarily of horizontally oriented ice crystals occur and have a signature of low depolarization [*Sassen* and *Benson*, 2001; Platt, 1978], so in this case the depolarization test fails to identify ice particles. Horizontally oriented crystals also exhibit a strong backscatter signal and a lidar ratio which is "anomalously" low (*Platt* and *Dilley*, 1981) because the backscatter coefficient is greatly enhanced compared to randomly oriented crystals. Supercooled water droplets also exhibit a strong backscatter signal, but have a lidar ratio typical of other water clouds, of about 18. Therefore, values of the cloud lidar ratio, S_c, derived from the integrated attenuated backscatter and two-way transmittance of transparent layers (refer to Section 7) can be used to identify horizontally oriented ice crystals and discriminate them from supercooled water droplets.

6.2. Algorithm Description

6.2.1. Inputs and Outputs of the Algorithm

The primary input data for cloud phase determination includes:

1. cloud top and base temperatures,
2. layer integrated depolarization ratio and uncertainty

The layer integrated volume depolarization ratio is the ratio of the integrals of the parallel and perpendicular components of the attenuated total backscatter,

$$d = \int_{base}^{top} b_{\perp}'(r)dr \Big/ \int_{base}^{top} b_{\parallel}'(r)dr \,. \qquad (6.1)$$

where $\beta'_{\perp}(r)$ and $\beta'_{\parallel}(r)$ are perpendicular and parallel attenuated backscatter profiles.

The key output parameters are:

1. index of cloud phase (0 = undetermined; 1 = ice; 2 = water; and 3 = oriented plates; however, recognition of oriented plates has not been implemented in the baseline version of the algorithm)
2. uncertainty flag (0 to 4: from minimum confidence to maximum confidence)

6.2.2. Probability Functions

There will be 3 groups of probability functions (for ice, water, and oriented plates) for each individual parameter or parameter set:

a) cloud temperatures (top and base)
b) layer integrated depolarization ratio

The baseline version of the algorithm, however, only implements identification of ice and water. The probability functions of ice and water as a function of temperature are modeled as

$$P_w(T_{top}) = \frac{1}{1+\exp(-a\cdot(T_{top}-c)+d)}; \qquad P_i(T_{top}) = 1 - P_w(T_{top}). \tag{6.2}$$

Where T_{top} is the temperature at the cloud top, subscripts *w* and *i* indicate water and ice, *a*, *c*, and *d* are constants that will be updated after the launch. The computation of the probability functions of ice and water for the depolarization ratio is discussed in the following subsection.

6.2.3. Probability Function for Depolarization Ratio in the Presence of Noise: Sigmoidal Functional Form

We start with the following example. Assuming that we know the ice cloud threshold value is C=0.2 for a certain cloud optical depths, what is the probability of the cloud phase being ice if the measured depolarization ratio is 0.15 and the signal-noise-ratio (SNR) is 3? The answer can be found by calculating the shaded area under the curve shown in Figure 6.3:; which is

$$P[d > c] = \int_c^{\infty} p(d')dd' = \int_c^{\infty} e^{-\frac{(d'-d)^2}{s^2}} dd' \tag{6.3}$$

By applying the sigmoidal approximation illustrated in Figure: 6.4:, the integration can be simplified into an analytic form which speeds up the estimation of the probability:

$$P[d > c] \approx \int_c^{\infty} \frac{d}{dd'}\left\{1+\exp\left[\frac{-2.3\cdot(d'-d)}{s}\right]\right\}^{-1} dd' = \left\{1+\exp\left[-2.3\cdot(d-c)\frac{SNR}{d}\right]\right\}^{-1} \tag{6.4}$$

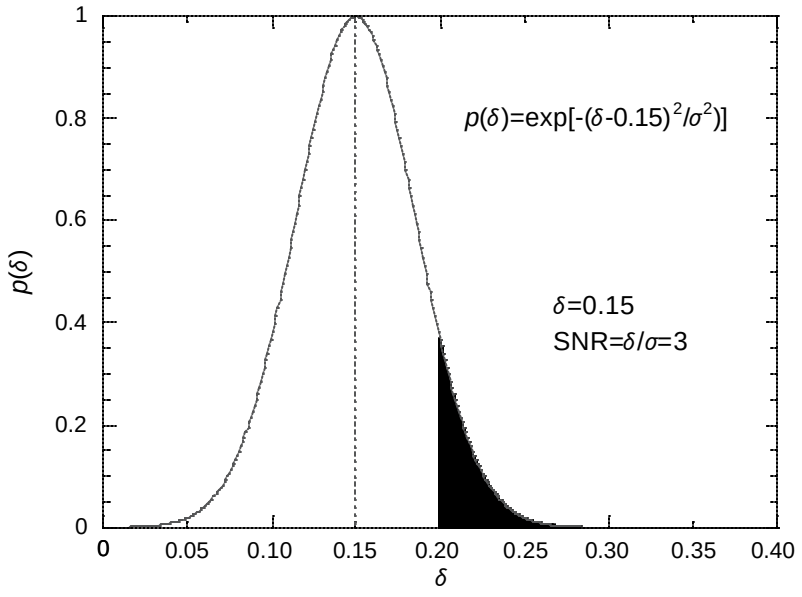


Figure 6.3: Probability that depolarization, *d*, is greater than ice cloud threshold value *c*.

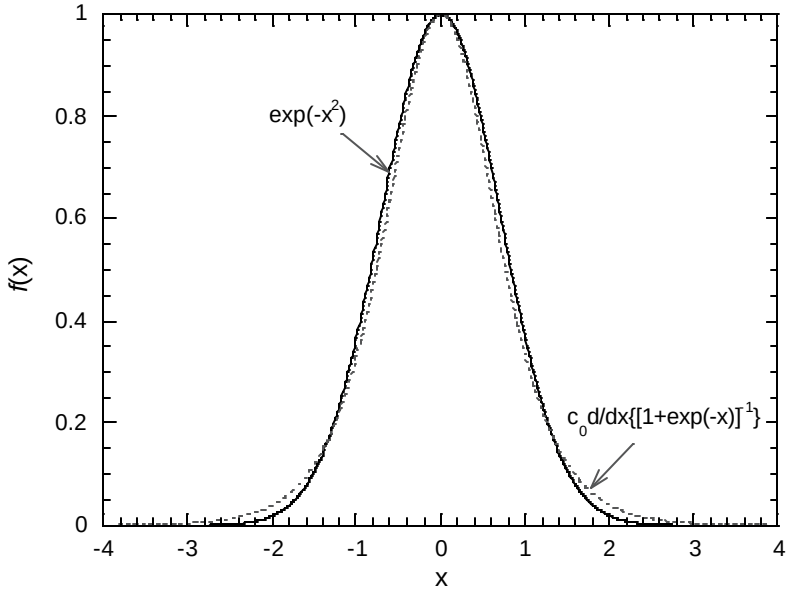


Figure: 6.4: Simplification of the integration of a Gaussian using a sigmoidal approximation.

In the baseline version of the algorithm, the probability functions for the water and ice phase are computed using

$$P_w(d, \Delta d) = \frac{1}{1 + \exp[-2.3 \frac{SNR}{d}(d - c_w)]}$$
$$P_i(d, \Delta d) = 1 - \frac{1}{1 + \exp[-2.3 \frac{SNR}{d}(d - c_i)]} \quad (6.5)$$

Where, c_w and c_i are coefficients that are computed as a function of layer-averaged attenuated backscatter. $SNR = d/\Delta d$ and Dd is the uncertainty of d. Figure 6.5 presents examples of probability functions computed using Eq. (6.5) for noisy and low-noise cases, i.e., SNR=2 and 1000. For the low-noise case, it is shown that there is a low depolarization ratio threshold of around 0.05 and a high threshold of 0.25. When the depolarization ratio is smaller than the low threshold, the feature is a water cloud; when the depolarization ratio is larger than the high threshold, the feature is an ice cloud. Theoretically, if the depolarization ratio lies between the low and high thresholds, the feature should be a mixed phase cloud. However, the depolarization ratios of both water clouds and ice clouds can lie between the low and high threshold due to noise in the signals (refer to the dotted curves). On the other hand, a mixed phase cloud could have a depolarization ratio of either smaller than 0.05 or larger than 0.25, again due to the noise.

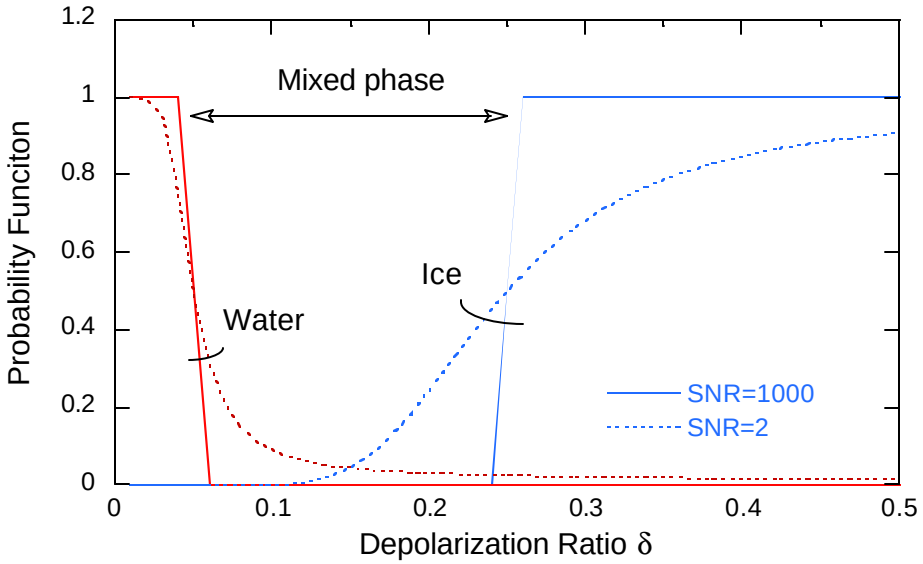


Figure 6.5: Examples of probability functions for the water and ice phase for SNR=2 and 1000.

6.2.4. Composite Probability Functions

For features whose classification is indeterminate based solely on depolarization (mostly for $0.5 < d_{\text{feature}} < 0.25$), we introduce the composite probability function

$$P_K(T,d) = P_K(T) * P_K(d), \qquad (K = 1,2) \tag{6.6}$$

where K=1 represents ice, K=2 represents water, and T represents the feature top temperature. After calculating both composite probabilities, the cloud phase is selected as the one with the largest probability.

6.2.5. Flowchart of Ice-Water Algorithm

The baseline IWA performs cloud phase classifications of water and ice based on tests of depolarization ratio and temperature at the feature top. A cloud is classified as "water", "ice", or "unknown". These unknown features could be mixed phase clouds, features that have very poor SNR, or mistakenly classified non-cloud features. The subtyping of mixed phase clouds (ice over water, water over ice, ice-plate, or mixture of ice particles and water droplets) is an ongoing study and the theoretical basis is described in Subsection 6.5. Classification of clouds as "mixed phase" has not yet been implemented in the IWA code.

A flowchart of the baseline IWA is presented in Figure 6.6:. For each feature that has been classified as cloud by the CAD module, the IWA will further identify its phase (water or ice). The IWA first reads the cloud layer-integrated depolarization ratio and its uncertainty (d, Δd), the temperature at the feature top (T_{top}), and the layer-averaged attenuated backscatter at 532 nm (b'). It then computes the probability functions for the depolarization ratio and temperature using, respectively, Eqs. (6.2) and (6.5). The classification is first conducted based on the depolarization ratio test. When the result of the depolarization ratio test is ambiguous, the

classification is then conducted based on the composite probability function of depolarization ratio and cloud-top temperature.

A confidence value, Q, ($0 \le Q \le 1$) is also computed. Q is evaluated as the maximum of $P(\delta,\Delta\delta)$ and $P(T_{top})$ when the decision is made using the depolarization ratio only, or as the square-root of the product of $P(d,\Delta d)$ and $P(T_{top})$ when the composite probability function is used. For values of Q greater than 0.75, we assign a 'maximum confidence' level (QC flag = 1) to the IWA classification. Similarly, for Q between 0.50 and 0.75, the IWA classification is judged to have 'high confidence' (QC flag = 2). The 'low confidence' regime (QC flag = 3) covers the range from $0.25 \le Q \le 0.5$. Finally, if Q is less than 0.25, then we assign 'no confidence' (QC flag = 0).

Table 6.1 QC flag assignments based on confidence value Q.

Q	QC Flag
> 0.75	1
0.50 < Q < 0.75	2
0.25 < Q <0.50	3
< 0.25	0

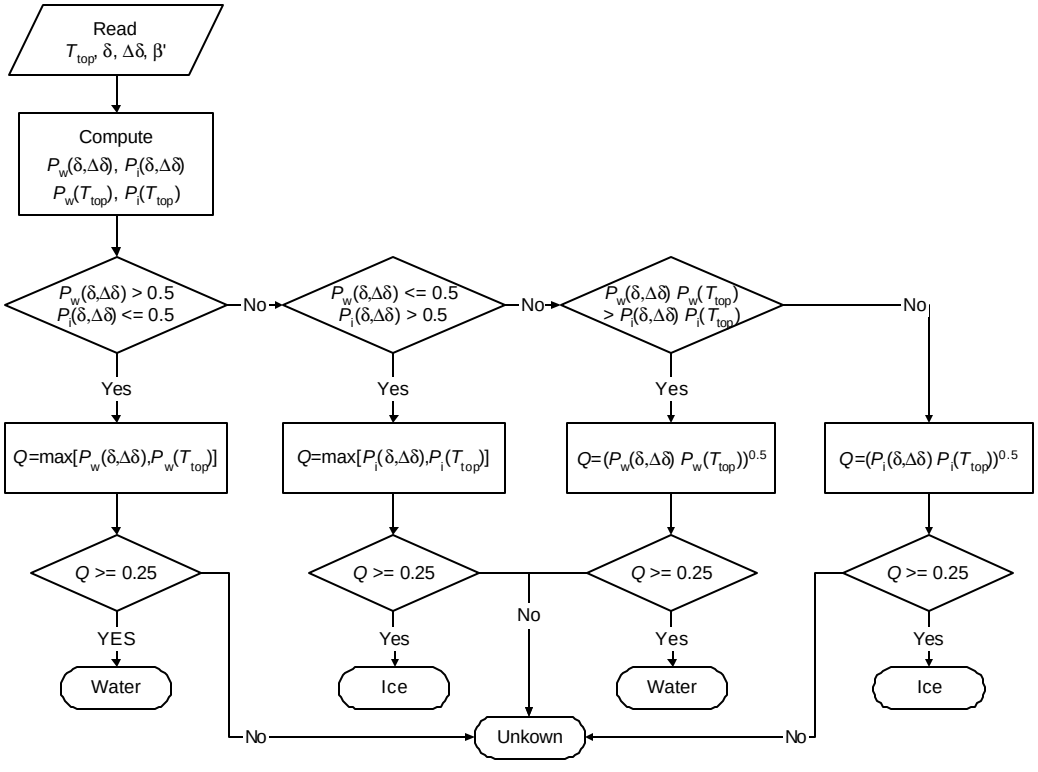


Figure 6.6: Flowchart of the baseline version of the ice-water algorithm.

6.3. Test results

Test results using artificially generated cloud fields are given below. The results show that the algorithm can identify all the water clouds properly. Ice clouds with depolarization ratios greater than 0.3 can also be easily identified.

However, for mixed phase clouds and oriented plates mixed with ice clouds, of which the layer integrated depolarization ratio can drop far below 0.3, it is likely that they are identified as water clouds by the baseline code. As an ongoing improvement on the baseline code, the concept of the discrimination of mixed phase clouds is described in Section 6.5.

Table 6.2 Some test results of IWA

Depol. ratio	*b*'	SNR	*T*	Phase [a)]	Conf *Q*
0.500	0.005	8.00	230.0	2	1.000
0.500	0.055	13.00	235.0	2	1.000
0.500	0.105	18.00	240.0	2	1.000
0.500	0.155	23.00	245.0	2	1.000
0.500	0.205	28.00	250.0	2	1.000
0.500	0.255	33.00	255.0	2	1.000
0.500	0.305	38.00	260.0	2	1.000
0.500	0.355	43.00	265.0	2	1.000
0.500	0.405	48.00	270.0	2	1.000
0.500	0.455	53.00	275.0	2	1.000
0.500	0.505	58.00	280.0	2	1.000
0.012	0.005	4.00	230.0	1	0.968
0.032	0.055	5.00	235.0	1	0.960
0.052	0.105	6.00	240.0	1	0.943
0.072	0.155	7.00	245.0	1	0.916
0.092	0.205	8.00	250.0	1	0.875
0.112	0.255	9.00	255.0	1	0.823
0.132	0.305	10.00	260.0	1	0.924
0.152	0.355	11.00	265.0	1	0.982
0.172	0.405	12.00	270.0	1	0.996
0.192	0.455	13.00	275.0	1	0.999
0.212	0.505	14.00	280.0	1	1.000

[a)] 1: water; 2: ice; 3: undetermined

Depol. ratio	*b*'	SNR	*T*	Phase [a)]	Conf *Q*
0.300	0.005	5.00	230.0	2	0.998
0.300	0.055	7.00	235.0	2	0.993
0.300	0.105	9.00	240.0	2	0.987
0.300	0.155	11.00	245.0	2	0.993
0.300	0.205	13.00	250.0	2	0.996
0.300	0.255	15.00	255.0	2	0.997
0.300	0.305	17.00	260.0	2	0.997

Depol. ratio	*b*'	SNR	*T*	Phase [a)]	Conf *Q*
0.300	0.355	19.00	265.0	2	0.996
0.300	0.405	21.00	270.0	2	0.994
0.300	0.455	23.00	275.0	2	0.988
0.300	0.505	25.00	280.0	2	0.969
0.012	0.005	3.50	230.0	1	0.952
0.032	0.055	4.00	235.0	1	0.927
0.052	0.105	4.50	240.0	1	0.892
0.072	0.155	5.00	245.0	1	0.846
0.092	0.205	5.50	250.0	1	0.793
0.112	0.255	6.00	255.0	1	0.735
0.132	0.305	6.50	260.0	1	0.924
0.152	0.355	7.00	265.0	1	0.982
0.172	0.405	7.50	270.0	1	0.996
0.192	0.455	8.00	275.0	1	0.999
0.212	0.505	8.50	280.0	1	1.000

[a)] 1: water; 2: ice; 3: undetermined

Depol. ratio	*b*'	SNR	*T*	Phase [a)]	Conf *Q*
0.200	0.005	5.00	230.0	2	0.707
0.200	0.055	7.00	235.0	2	0.700
0.200	0.105	9.00	240.0	2	0.675
0.200	0.155	11.00	245.0	2	0.609
0.200	0.205	13.00	250.0	2	0.462
0.200	0.255	15.00	255.0	2	0.256
0.200	0.305	17.00	260.0	3	0.202
0.200	0.355	19.00	265.0	1	0.411
0.200	0.405	21.00	270.0	1	0.996
0.200	0.455	23.00	275.0	1	0.999
0.200	0.505	25.00	280.0	1	0.000
0.052	0.005	4.50	230.0	1	0.897
0.072	0.055	6.00	235.0	1	0.881
0.092	0.105	7.50	240.0	1	0.839
0.112	0.155	9.00	245.0	1	0.769
0.132	0.205	10.50	250.0	1	0.672
0.152	0.255	12.00	255.0	1	0.731
0.172	0.305	13.50	260.0	1	0.643
0.192	0.355	15.00	265.0	1	0.591
0.212	0.405	16.50	270.0	1	0.541
0.232	0.455	18.00	275.0	1	0.513
0.252	0.505	19.50	280.0	1	0.517

[a)] 1: water; 2: ice; 3: undetermined

6.4. Monte Carlo Simulation Study

6.4.1. Single-Phase Clouds - Plate Crystal Ice Cloud

A Monte Carlo simulation scheme has been developed to assess the impact of multiple scattering on the depolarization signal. The code models the photon scattering processes within the cloud layer using a ray tracing scheme and carries the full phase matrix through each scattering event [*Hu et al.*, 2001]. Figure 6.7 shows profiles of particle depolarization ratio $d(z) = b_{\perp}(z) / b_{\parallel}(z)$ and backscatter intensity $[b_{\perp}(z) + b_{\parallel}(z)]$ calculated for ice clouds composed of hexagonal plate crystals with optical depths, t, of 0.9, 2.7, and 50. The single-scattering depolarization ratio at 180 degrees for this crystal type is ~ 30 %. For all three cases, $d(z)$ immediately below cloud top is greater than 0.3.

$d(z)$ remains virtually constant with penetration depth for the two lower optical depth cases, but increases to about 0.6 in the topmost 100 meters for the $t = 50$ case due to multiple scattering. The $t = 50$ case is probably unrealistically dense, corresponding to a backscatter coefficient of about 2.5 [$km^{-1}sr^{-1}$], even for a convective cirrus cloud. This case however does indicate that depolarization due to multiple scattering will have to be considered if ice clouds are encountered which are optically very dense.

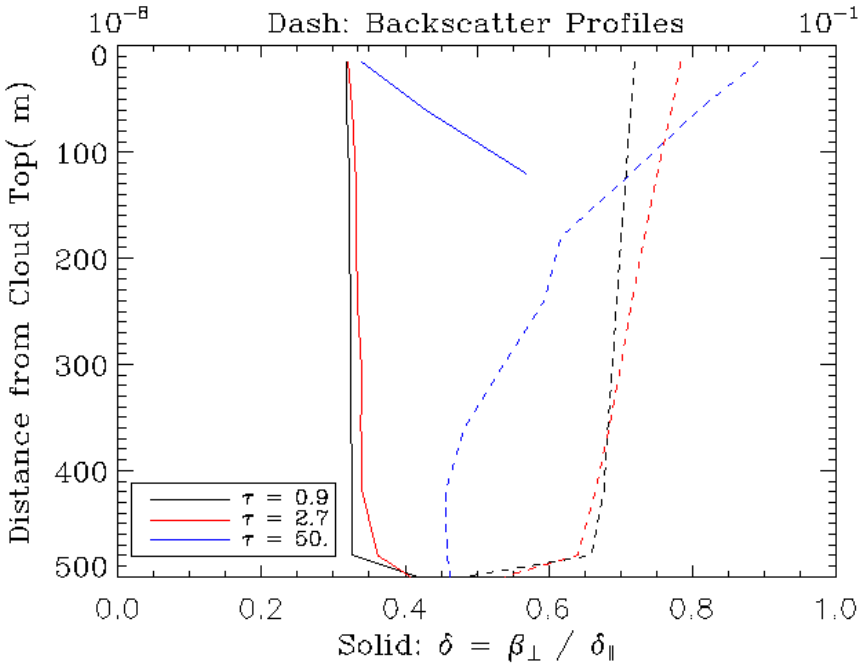


Figure 6.7: Calculated profiles of backscattered intensity $[b_{\perp}(z) + b_{||}(z)$, dashed curves] and depolarization ratio $[b_{\perp}(z) / b_{||}(z)$, solid curves] for plate crystal ice clouds with optical depths (t) of 0.9, 2.7, and 50.

6.4.2. Water Cloud

Figure 6.8 shows curves of backscatter intensity and depolarization ratio for water clouds having the same optical depths as the ice clouds in Figure 6.7. The backscatter intensity profiles for these clouds are nearly identical to those seen for the ice cloud in Figure 6.7, but the profiles of depolarization ratio are markedly different. δ is near zero in the first bin within the cloud for all values of τ, and increases with penetration into the cloud due to multiple scattering. For the t = 0.9 and t = 2.7 cases the increase in depolarization is relatively small. However, for the t = 50 case the increase in depolarization in the first 100 meters of the cloud is even greater than in the cirrus case above. The depolarization at a depth of 100 meters into the cloud is about the same as in the cirrus case, roughly 50%, but the rate of increase in depolarization is greater than for the cirrus cloud.

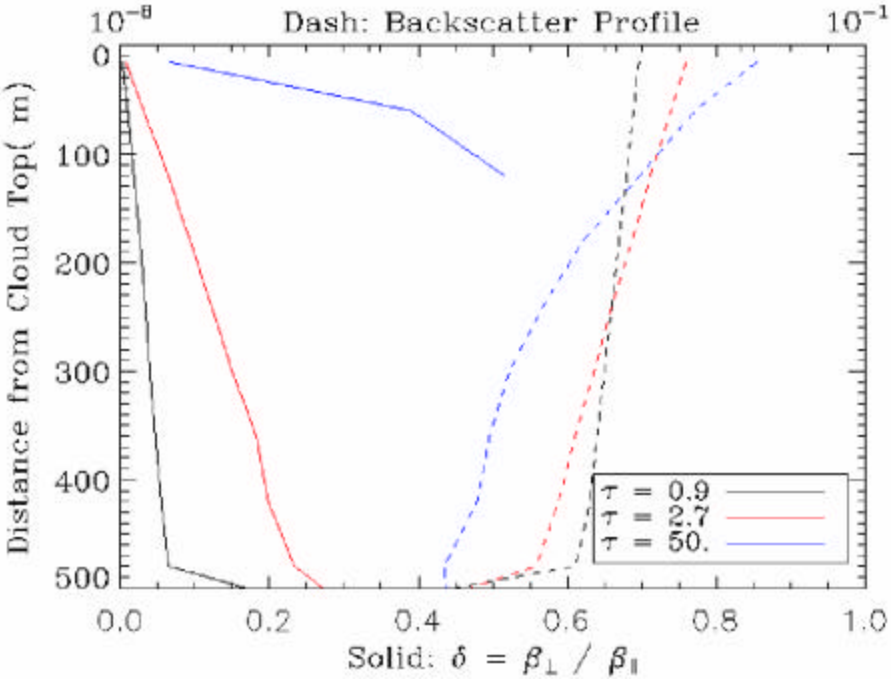


Figure 6.8: Calculated profiles of backscatter intensity [$b_\perp(z) + b_{||}(z)$, dashed curves] and depolarization ratio [$b_\perp(z) / b_{||}(z)$, solid curves] for water clouds with t = 0.9, 2.7, and 50.

6.4.3. Simulations of Mixed Phase Clouds

A great deal of experimental evidence exists showing that near-range depolarizing layers do not impair the ability of lidar to detect non-depolarizing layers further away [e.g., *Winker and Osborn*, 1991]. Therefore, the distribution of water and ice within mixed-phase clouds can be observed, subject to the limitation that the lidar signal will be unable to completely penetrate optically dense clouds.

To demonstrate this, Figure 6.9 shows results from Monte Carlo simulations run for clouds of optical depth 2 having identical extinction coefficient profiles (shown as the red and blue step-function curves in the figures), but with different ice/water phase profiles. Frame (a) represents two layers of water cloud, while frame (b) represents two layers of ice cloud. Frame (c)

represents a 100-m water layer overlying a 900-m thick ice cloud and frame (d) represents a 100-m ice layer overlying a 900-m thick water cloud. While the backscatter intensity (dotted black) profiles are very similar for all four cases, the depolarization ratio (solid black) profiles are quite different. In the future, tests will be developed to identify mixed-phase clouds and to locate regions of ice and water within the clouds.

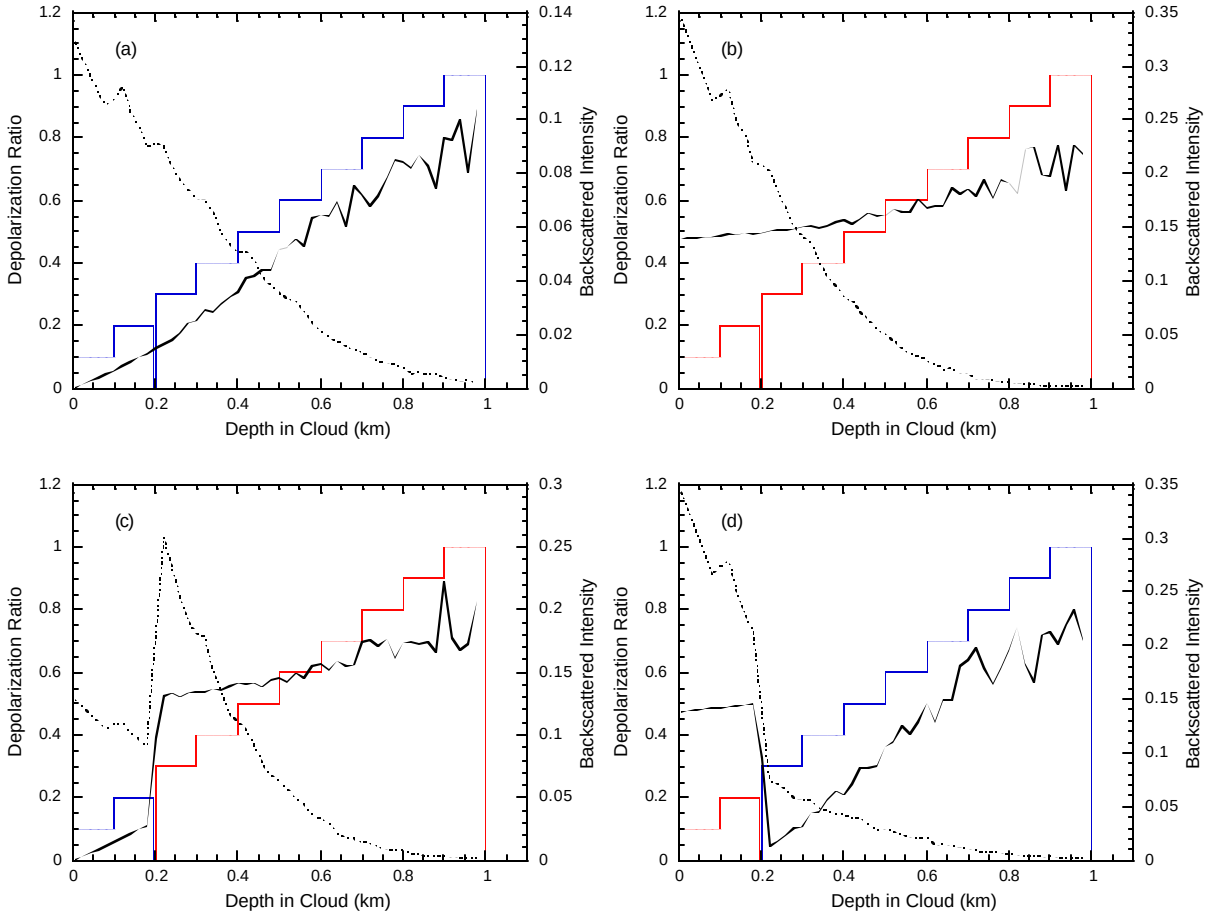


Figure 6.9: Calculated profiles of backscatter intensity [$b_{\perp}(z) + b_{||}(z)$, dashed curves] and depolarization ratio [$b_{\perp}(z) / b_{||}(z)$, solid curves] for single-phase and mixed-phase clouds with $t = 2.7$.

6.5. Ongoing improvements on mixed-phase clouds

Future algorithm improvements will include the identification of mixed-phase clouds and of oriented plate particles. The type of mixed-phase clouds to be classified will include

1. Water over ice
2. Ice over water
3. Plates over ice
4. Ice over plates
5. Water mixed with ice

Characteristics to consider in discriminating water and ice in mixed phase clouds are:

- the single scattering depolarization ratios for water clouds are close to zero
- depolarization ratios for ice clouds with randomly oriented crystals are much larger (in the neighborhood of 0.4)
- multiple scattering increases depolarization ratios of water clouds and makes cloud phase discriminations more difficult
- depolarization ratios for some ice clouds (e.g., oriented plates) can be similar to water clouds

Figure 6.10 presents an example of the water-over-ice identification using layer-integrated depolarization. Dotted curves are simulated layer integrated depolarization ratios as a function of layer integrated backscatter based on the Monte-Carlo simulation. The integration is from the cloud top toward the base. Six situations of layer thickness have been considered. For all cases, the integrated depolarization ratio increases gradually with increasing penetration into the water layer (increasing the layer integrated backscatter) due to the multiple scattering, though the single scattering depolarization would be zero if we assume spherical water droplets. A sharp increase occurs at the interfaces of the water and ice layers, because of high depolarization ratio of the ice particles.

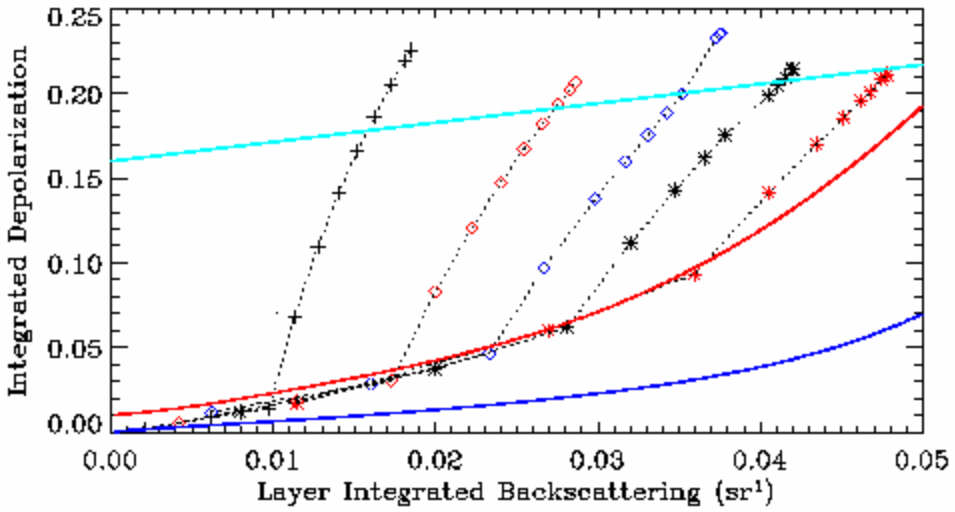


Figure 6.10: Relationship between integrated depolarization and integrated backscatter for water over ice. The solid blue line indicates the oriented plate's threshold; the solid red line shows the sphere/non-sphere threshold; and the solid green line is the ice threshold.

The layers for this example can be identified by setting three threshold curves (solid curves in Figure 6.10). For any layer, if the integrated depolarization rises above the green curve, the cloud consists of ice over water. If the integrated depolarization remains between the red and blue curves for the entire depth of the cloud, the cloud is pure water. If the integrated depolarization falls below the blue curve, a region of oriented plates (ice) is indicated. However, if the integrated depolarization lies in between the green and red curves, the cloud phase is undetermined (mixed water and ice or mixed water, ice and plate, etc.) Details will be provided in the next release of this document.

7. Computing Lidar Ratio for Elevated Layers

The extinction-to-backscatter ratio or the lidar ratio, defined as $S = s / b$, where s and b are respectively extinction and backscatter coefficients, is an intensive particle property. It does not depend on the number density of the particles but rather on such physical and chemical properties as size distribution, particle shape, and composition. In the aerosol case, these properties depend primarily on the source of the aerosol and such factors as mixing, transport, and hydration. The lidar ratio is an important parameter used in the retrieval of feature extinction, and subsequently the optical depth, from lidar backscatter measurements. There are two unknowns, extinction and backscatter, in the lidar equation. A unique solution is only possible if a relationship can be established (or prescribed) between the two unknowns such that these are combined into one variable. A general and widely used assumption is that the lidar ratio is constant throughout a feature layer. Thus, analytical solutions can be derived from the lidar equation, based on which solutions lidar signal can be inverted to derive particulate scattering parameters. This assumption has been adopted in the CALIOP data processing.

For some layers where molecular scattering signals are available both above and below the feature, the SCA will compute the lidar ratio from lidar measurements using the transmittance-constraint lidar retrieval method. The transmittance method is introduced in this section. For the non-elevated layers, on the other hand, the SCA will choose a model and lidar ratio based on the identified feature type. The selection of model and lidar ratio is introduced in Sections 8 and 9, respectively, for the non-elevated aerosol and cloud layers.

7.1. Transmittance-Constraint Lidar Retrieval Method

The transmittance method requires clear air (molecular scattering) above and below the layer, from which the layer transmittance and consequently optical depth can be determined [*Fernald et al*., 1972; *Young*, 1995]. The basis of the method is the relationship between optical depth and integrated attenuated backscatter, described by the following equation [*Platt*, 1973]:

$$g' = \frac{1 - \exp(-2ht)}{2hS} \tag{7.1}$$

Here g' is the integrated attenuated backscatter from layer top to base,

$$g' = \int_{top}^{base} b(r) T^2(r)\, dr \tag{7.2}$$

where t is optical depth and h is a layer-effective multiple scattering parameter. Note that Eq. (7.1) is valid for the single component layers where molecular scattering is negligibly small compared with particulate scattering (e.g., in clouds). However, in aerosol layers, the molecular scattering can be significant and must be taken into account in using Eq. (7.1) to compute lidar ratio. To make use of Eq. (7.1), a technique has been developed to correct for the molecular scattering. Figure 7.1 illustrates the computation of particulate integrated attenuated backscatter using the molecular scattering correction technique. The technique approximates the molecular scattering contribution to the total integrated attenuated backscatter with a trapezoid as illustrated in Figure 7.1 (the area in gray). The particulate integrated attenuated backscatter is estimated by

subtracting the trapezoid from the total integrated attenuated backscatter, i.e., the area in blue in Figure 7.1.

If we define an effective lidar ratio, $S^* = hS$, and substitute the effective two-way transmittance, $T^2 = \exp(-2ht)$, we can rewrite Eq. (7.1) as follows:

$$S^* = \frac{1 - T^2}{2g'} \tag{7.3}$$

The effective two-way transmittance is typically obtained by fitting the return both above and below a feature to a reference profile [*Young*, 1995]. The SCA then estimates the uncertainties in the observationally-derived value of S and sets a flag indicating whether the extinction retrieval (see PC-SCI-202, part 3) should perform a retrieval using the observational value or the model-derived value.

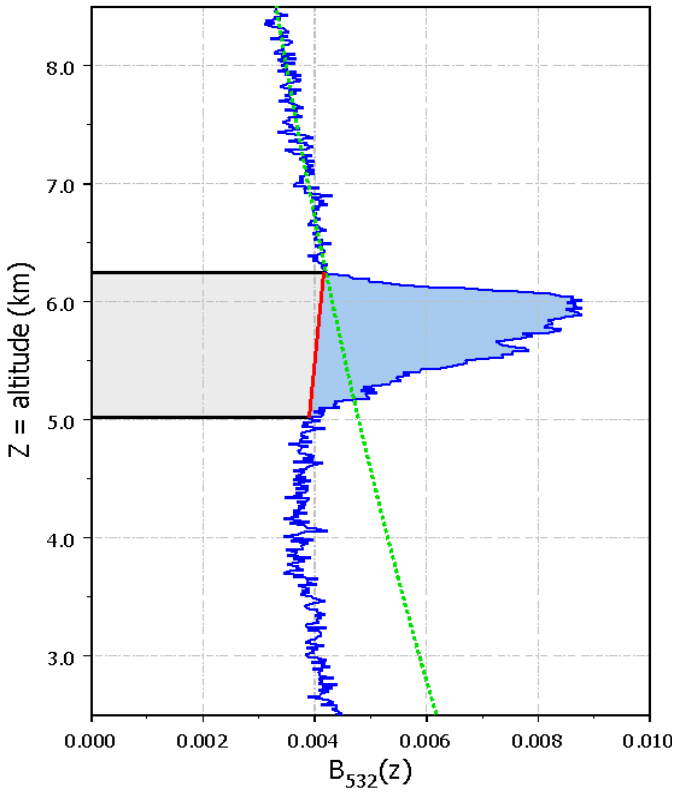


Figure 7.1: Computation of particulate integrated attenuated backscatter for elevated layers. Molecular scattering contribution to the total integrated attenuated backscatter is approximated as the trapezoid area in gray; the area in blue is the estimate of particulate integrated attenuated backscatter that is used to compute S^* using Eq. (7.1) or (7.3).

7.2. Computing Effective Lidar Ratio Uncertainty

The expression for the lidar ratio uncertainty is

$$\frac{Var\left(S^*\right)}{\left(S^*\right)^2}=\left(\frac{1}{1-T^2}\right)^2\left(\frac{Var\left(C_{hi}\right)}{C_{hi}^{\ 2}}\right)+\left(\frac{T^2}{1-T^2}\right)^2\left(\frac{Var\left(C_{lo}\right)}{C_{lo}^{\ 2}}\right)+\left(\frac{Var\left(g'\right)}{g'^2}\right). \qquad (7.4)$$

C_{hi} and C_{lo} are normalization constants calculated both immediately above (C_{hi}) and immediately below (C_{lo}), and are computed by normalizing the measured signal to the molecular model over some altitude range. Normalization constants are related to the profile system constant via the following equation: $C_{norm}\left(r\right)=C_{cal}\cdot T^2\left(r\right)$. The third term on the right hand side is the contribution due to the error in the computed particulate integrated attenuated backscatter including measurement noise and the error of trapezoid approximation of molecular scattering. The relative error of the computed effective lidar ratio S^* is the square-root of the left hand side of Eq. (7.4).

8. Selection of Aerosol Model

8.1. Background

For cases where the aerosol lidar ratio, S_a, is not computed from layer transmittance, the value used in the lidar inversions is derived from identification of the type of aerosol, a process we shall refer to hereafter as 'aerosol typing'. In this section we discuss how the SCA determines the aerosol type, chooses the appropriate aerosol model for the type, and determines the appropriate S_a and scattering phase function.

8.2. Algorithm Overview

After discriminating cloud and aerosol layers, the SCA attempts to identify the type of aerosol in the layer. A look-up table is used to associate a lidar ratio and other properties with that layer. This has several purposes: attribution of aerosol radiative forcing to natural or anthropogenic emissions requires the determination of the source of the aerosol; aerosol radiative properties vary significantly by type; and, most directly, determination of aerosol type allows an estimate of S_a. By "type" we mean an aerosol mixture which is characteristic of a region or an air mass. The mixture observed at a given location depends on the local and remote aerosol sources and wind trajectories, internal and external mixing, the state of hydration, and chemical processes which may have occurred during transport.

In contrast to the approach adopted by MISR, which is to define pure aerosol types which are assumed to be externally mixed [*Kahn et al.*, 2001], in the CALIOP approach each aerosol type is assumed to be a mixture of aerosol particles of different compositions, where the mixing can be internal, external or both. The underlying paradigm of the type classification is that the variety of emission sources and atmospheric processes will act to produce air masses with a typical, identifiable aerosol "type". This is an idealization, but one that allows us to classify aerosols based on observations and location in a way to gain insight into the geographic distribution of aerosol types and constrain the possible values of S_a for use in aerosol extinction retrievals.

The S_a value is determined on a case by case basis and depends on the aerosol composition, size distribution, and shape. Several studies have been conducted to determine S_a values characteristic of different aerosol types. Although the assumption that S_a is constant throughout the vertical column has been widely used in the lidar community, recent studies [*Ansmann et al.*, 2001; *Ferrare et al.*, 2001] have shown that S_a can vary significantly with altitude. In the baseline version of the algorithm S_a is selected independently for each aerosol layer in a vertical column, but S_a is assumed to be constant within each layer.

8.3. CALIOP Aerosol Models

The multi-year AERONET data archive has been analyzed using a clustering algorithm to determine characteristic aerosol types that are defined in terms of physical and optical properties such as complex refractive indices, geometric mean radii and standard deviations, single scattering albedos, etc. [*Omar et al.*, 2005]. AERONET is a global network of multi-wavelength sun photometers designed to make sky radiance measurements, allowing inversions of aerosol microphysical properties, which has collected data for up to a decade in some locations [*Dubovik*

and King, 2000; *Holben et al.*, 1998]. AERONET retrieves aerosol fine and coarse mode size distribution parameters and refractive indices at several wavelengths, based on the assumption of a bimodal log normal distribution and spherical or spheroidal particles. From the size and refractive index data retrieved by AERONET, Mie theory is used to compute the lidar ratios at the two CALIOP wavelengths and the aerosol backscatter color ratio, as well as the complete phase function, which is needed to estimate multiple scattering effects. The cluster analysis identified six aerosol 'mixture types' representative of the aerosol mixtures most frequently present at the AERONET sites.

The CALIOP aerosol product includes both aerosol extinction and aerosol type. Aerosol is classified as one of six types: desert dust, smoke from biomass burning, clean continental, polluted continental, marine, and polluted dust. Arctic haze is modeled as polluted continental. Desert dust is mostly mineral soil; biomass burning is an aged smoke aerosol consisting primarily of soot and organic carbon (OC), background aerosol (also referred to as clean continental) is a lightly loaded aerosol consisting of sulfates (SO_4^{2-}), nitrates (NO_3^-), OC, and ammonium (NH_4^+); polluted continental is background aerosol with a substantial fraction of urban pollution; marine aerosol consists primarily of seasalt (NaCl); and polluted dust is a mixture of desert dust and smoke. While this set does not cover all possible aerosol mixing scenarios, it accounts for a majority of mesoscale aerosol characteristics. In essence the algorithm trades off complex transient multi-component mixtures for relatively stable layers with large horizontal extent (10-1000 km).

8.4. CALIOP Type-Specific S_a

CALIOP will produce two Level 2 aerosol extinction products derived using different choices of S_a. An approximate extinction product will be derived using a fixed extinction-to-backscatter ratio (currently, S_a = 35 sr). This value corresponds to S_a of the clean continental and is intermediate between high and low values of S_a encountered in the atmosphere. Analyses of the AERONET dataset described in *Omar et al.* [2005] show that clean background aerosol is frequently encountered in the atmosphere. In addition, experience with LITE measurements shows that this value is likely to result in stable extinction retrievals. The second extinction product is derived using the best estimate of S_a, derived from CALIOP measurements and the most up-to-date field observations. This algorithm uses information on surface type to aid in selecting values of S_a, which may result in abrupt and artificial changes in aerosol extinction and optical depth, and regionally correlated errors in the retrieved aerosol properties. Comparison of the best-estimate retrieval with the fixed-S_a retrieval provides a measure of the dependence of the derived extinction and optical depth on the algorithm used to select S_a. We discuss the basis of the S_a estimates and describe the selection algorithm below. The goal is to constrain the uncertainty in S_a to no more than 30%. Given the observed range of variability of S_a between 10 and 110 sr [*Anderson et al.*, 2000b] and the modeled range of 15 – 80 sr [*Ackermann*, 1998], we only need to select among a small set of values to meet the 30% requirement over the whole range of observed S_a values.

The derivation of S_a from AERONET measurements has not yet been validated. In addition, because AERONET produces column retrievals, care must be taken because multiple layers of different aerosol types will lead to erroneous estimates of the intensive aerosol properties. Therefore, 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 for that aerosol type. These measurements, primarily at 532 nm, come from a variety of techniques: 180-nephelometer [*Anderson et al.*, 2000], Raman lidar [*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 measurements.

8.5. Aerosol Types and S_a values

Of the six AERONET clusters, three (desert dust, biomass burning, polluted continental) were adopted directly as CALIOP aerosol models. Three additional models (marine, polluted dust, and clean continental) were built either directly from measurements of size distributions and complex refractive indices, or by adjusting AERONET-derived model parameters to generate observed S_a values. Volume size distributions of the 6 CALIOP aerosol models are shown in Figure 8.1, and the associated microphysical and derived properties are listed in Table 8.1. Once the aerosol type is identified, the associated S_a value is the best-guess value to be used in the extinction retrieval. The set of aerosol models described here is an initial result. Further research will be conducted to improve these models. There is currently no dataset suitable for testing the performance of the algorithm. Therefore, much of this work will have to wait till we have on-orbit data and have conducted validation campaigns.

8.5.1. Background Aerosol

Since the AERONET observations corresponding to the background cluster 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 CALIOP background (clean continental) 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], resulting in a S_a value of 35 sr. Note that the S_a value for this aerosol type is used to generate the approximate extinction product described above. A S_a value of 32 ±6 sr for clean northern hemisphere aerosol was measured during Aerosols99 [*Voss et al.*, 2001]. A similar aerosol termed 'background-like' aerosol and originating on the European continent but devoid of any strong biomass or fossil fuel burning signature yielded a S_a value of 35 sr [*Ansmann et al.*, 2001].

8.5.2. Marine

Since the AERONET marine aerosol cluster contains only a small number of records (< 4% of the total), the CALIOP marine aerosol model is instead derived from parameters measured during the SEAS experiment [*Masonis et al.*, 2002]. This study estimates a S_a for marine aerosols of 20 sr and is consistent with marine aerosol S_a estimates by others [*Ansmann et al.*, 2001; *Flamant et al.*, 1998; *Reagan et al.*, 2001]. In their climatological study of oceanic AERONET sites, *Cattrall and Reagan* (2005), report a S_a value 28±5 sr. In the SEAS experiment, the S_a value is measured directly (optical method) or modeled (using measured size distributions). The optical method, using 180^o-backscatter nephelometer, integrating nephelometer, and an absorption photometer, yields a value of S_a = 25.4±3.5 sr. The modeled values are 20.3 sr and 16.5 sr using direct size data and log normal distributions, respectively.

8.5.3. Polluted Dust

The polluted dust model is designed to account for episodes of mixed dust and biomass burning smoke which are frequent in regions close to strong sources of both types, for example in West Africa (*cf.* MODIS images) and Asia (*cf.* ACE-Asia, INDOEX). The CALIOP 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 which is comparable to similar measurements of polluted dust [*Liu et al.*, 2002; *Voss et al.*, 2001].

Table 8.1 Physical and Optical Characteristics of the CALIOP Aerosol Models

Optical/Physical Property	Desert Dust	Biomass Burning	Clean Continental	Polluted Continental	Marine (Seasalt)	Polluted Dust
m_r fine @ 532 nm	1.414	1.517	1.380	1.404	1.400	1.452
m_i fine @ 532 nm	0.0036	0.0234	0.0001	0.0063	0.0050	0.0109
m_r fine @ 1064 nm	1.495	1.541	1.380	1.439	1.400	1.512
m_i fine @ 1064 nm	0.0043	0.0298	0.0001	0.0073	0.0050	0.0137
m_r coarse 532 nm	1.414	1.517	1.455	1.404	1.400	1.452
m_i coarse 532 nm	0.0036	0.0234	0.0034	0.0063	0.0005	0.0109
m_r coarse 1064 nm	1.495	1.541	1.455	1.439	1.390	1.512
m_i coarse 1064 nm	0.0043	0.0298	0.0034	0.0073	0.0005	0.0137
? @532 nm	0.91	0.70	0.90	0.88	0.99	0.79
Fine cut-off radius (μm)	1.00	1.00	1.00	1.00	0.60	1.00
Fine fraction by volume	0.223	0.329	0.050	0.531	0.025	0.241
Fine mean radius (μm)	0.1165	0.1436	0.20556	0.1577	0.150	0.1265
GSD Fine	1.4813	1.5624	1.61	1.5257	1.600	1.5112
Coarse fraction by volume	0.777	0.671	0.950	0.469	0.975	0.759
Coarse mean radius (μm)	2.8329	3.726	2.6334	3.547	1.216	3.1617
GSD coarse	1.9078	2.1426	1.8987	2.065	1.600	1.9942
Color Ratio	0.79	0.67	1.39	0.72	0.53	1.1
S_a @ 532 nm (sr)	38.1	71.3	37.7	72.7	19.1	65.3
S_a @ 1064 nm (sr)	29.3	38.9	28.2	30.9	43.2	30.9

Note: m_r(m_i) are complex real (imaginary) parts of the refractive index, ? is the single scattering albedo, color ratio is the ratio of the particle backscatter at 1064 nm to 532 nm and GSD is the geometric standard deviation of a log normal distribution.

8.5.4. Biomass Burning

The biomass burning cluster of AERONET measurements is used for the CALIOP biomass burning aerosol model. The AERONET data yields an S_a value of 70 sr, most likely driven by the high imaginary refractive indices (Table 8.1). This value is not at variance with the measurements of *Voss et al.* [2001] of 60 ± 6 sr off the west coast of Africa, and *Ansmann et al.* [2001] of 70 sr for biomass burning influenced aerosol advected from the Indian subcontinent during INDOEX. Recent studies of a limited climatology of 26 AERONET sites [*Cattrall et al.*,

2005] at which biomass burning, coal combustion, urban/industrial, oceanic, and dust are the predominant types found S_a values of 60±8 sr for biomass burning.

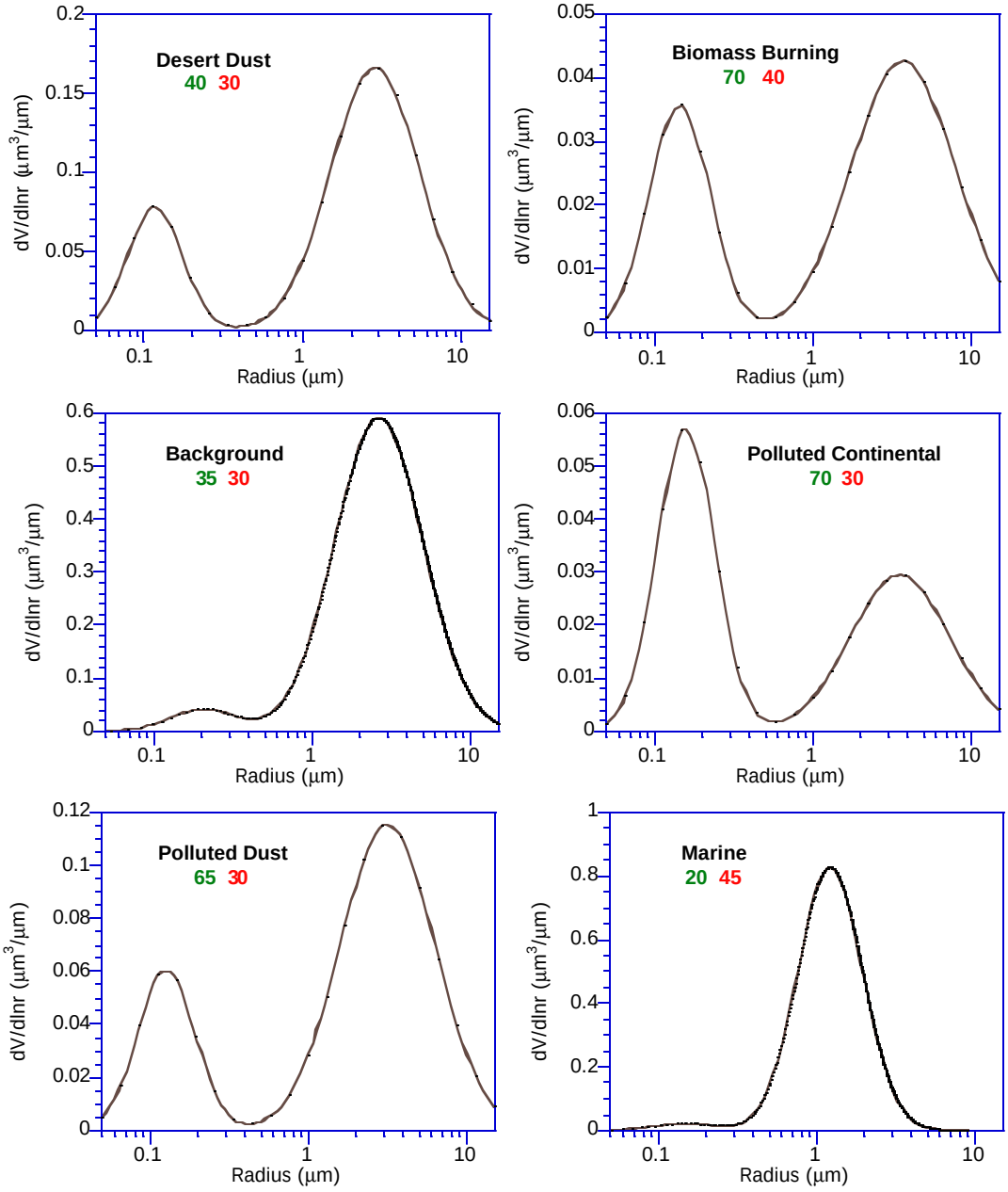


Figure 8.1: CALIOP Aerosol Models. The green and red numbers are the 532- and 1064-nm extinction-to-backscatter ratios, respectively.

8.5.5. Desert Dust

In the case of desert dust, the AERONET retrievals used in the cluster analysis were based on the assumption of spherical particles. Dust particles are non-spherical, though, which has an

unquantifiable effect on the accuracy of the results. Therefore, the CALIOP dust model is based on theoretical calculations using the dipole-dipole approximation (DDA) technique, using realistic compositions and irregular shapes [*Kalashnikova and Sokolik*, 2002]. The CALIOP model S_a value of 40 sr for desert dust is comparable to S_a measurements by *Voss et al.* (2001) using a Micropulse Lidar (41 ± 8 sr) for African dust, *Sasano and Browell* [1989] (52 ± 10 sr) and measurements of *Liu et al.* [2002] of Asian dusts (42 - 55 sr) found using a high-spectral-resolution lidar and combined Raman and elastic-backscatter lidar values. *Cattrall and Reagan* (2005) report an S_a value of 42 ± 4 sr for desert dust locations.

8.5.6. Polluted Continental

The CALIOP model for polluted continental aerosol yields an S_a value of 70 sr. Measurements by *Ansmann et al*. [2001] at the Sagres island off the Portuguese coast showed the S_a value for pollution emanating from continental Europe to vary between 50 and 70 sr. During INDOEX, S_a measurements of polluted continental aerosol originating from the northern and northeastern part of 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 and Reagan* [2004] report an S_a value of 71 ± 10 sr for urban/industrial locations. Measurements of a stagnant airmass at Bondville (a polluted continental site) yielded S_a values of 64 ± 4 sr [*Anderson et al.*, 2000].

8.6. Type Identification and S_a Selection Scheme

One of the objectives of the aerosol typing algorithm is to estimate the appropriate value of S_a within 30% of the true value. The strategy is to identify aerosol type and then use a look-up table to select values of S_a and $h(z)$ appropriate for the layer. These values are then passed on for use in the extinction retrieval algorithm, if necessary. The selection scheme uses the observed backscatter strength and depolarization to identify aerosol type, to the extent possible, from among one of the six types. The depolarization is directly related to the hydration state of the aerosol (ie: solid or liquid). The backscatter and depolarization signals are not sufficient to fully constrain the model selection, however. Therefore, additional data is used to narrow the choices of aerosol types based on the lidar observables. The S_a selection algorithm uses the inputs shown in Table 8.3, along with the IGBP surface types listed in Table 8.2.

Table 8.2 IGBP surface types

1. Evergreen Needleleaf Forest 2. Evergreen Broadleaf Forest 3. Deciduous Needleleaf Forest 4. Deciduous Broadleaf Forest 5. Mixed Deciduous Forest 6. Closed Shrubland 7. Open Shrubland 8. Woody Savanna 9. Savanna	10. Grassland 11. Permenant Wetland 12. Cropland 13. Urban 14. Crop/Natural Veg. Mosaic 15. Permanent Snow/Ice 16. Barren/Desert 17. Water Bodies 18. Tundra

Table 8.3 Inputs and outputs for the S_a selector algorithm

INPUTS	SOURCE	NOTES	OUTPUTS
CALIOP Observables			
1. Elevated layer flag	SIBYL	Determine lofted layer for transmittance calculation	Estimated S_a using $S_a=T^2/(1-2g')$
2. Feature base and top heights	SIBYL		
3. Two-way feature transmittance (T^2)	SIBYL		
4. Uncertainty in the two-way feature transmittance ($?T^2$) 5. Normalization constant uncertainty	SIBYL	Estimate uncertainty in the S_a calculated from lofted aerosol layers using the transmittance method	
6. Integrated attenuated 532- and 1064-nm backscatter (g') 7. Uncertainty in the integrated attenuated 532-and 1064-nm backscatter ($?g'$)	SIBYL SIBYL	 Estimate a confidence index of aerosol type	Aerosol type
8. Integrated attenuated color ratio (χ) 9. Uncertainty in the integrated attenuated color ration ($?c$)	SIBYL	Uncertainty flag in aerosol type Estimate a confidence index of aerosol type	Aerosol type
10. Integrated volume depolarization ratio(d) 11. Uncertainty in the integrated volume depolarization ratio ($?d$)	SIBYL	Uncertainty flag in aerosol type Estimate a confidence index of aerosol type	Aerosol type
Ancillary Data Sets			
1. Latitude, Longitude	GIS	Local sources and	Aerosol type
2. Surface type	IGBP	influence on the aerosol type, composition, size, and shape	
4. Local solar time	Level I	Season	Aerosol type (Polar)

The input parameters - altitude, location, surface type, depolarization ratio, and mean attenuated backscatter coefficient measurements - are used to identify the aerosol type following one of eleven pathways in Figure 8.2. After launch the distributions of the integrated attenuated color ratio, $c = g'_{1064} / g'_{532}$, for each aerosol type will developed and used for type identification in later versions of this algorithm. The depolarization ratio is used to identify aerosol types that have a substantial mass fraction of non-spherical particles, e.g., a mixture of smoke and dust in pathways 3 and 5. $?'_{532}$ is used 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, pathways 10 and 11, respectively. Once the type is identified, S_a is chosen from a lookup table consisting of six pairs of S_a values at 532 nm and 1064 nm.

In Figure 8.2, pathway 1 is a lightly loaded aerosol layer found over snow/ice/tundra regions such as Antarctica and the clean Artic. Arctic haze, by virtue of the high integrated backscatter value, will be classified as polluted continental following pathway 2. Desert dust with a volume depolarization ratio greater than 0.2 is expected to be the predominant selection pathway for layers with substantial fractions of non-spherical particles (Pathway 4). Pathways 3 and 5 allow for mixing with biomass burning smoke, which will depress the volume depolarization ratio to a value below 0.2. Pathway 6 is a clean, non-desert land surface at which the aerosol loading is close to the background values. Pathway 7 is a highly polluted land surface such as would be found in urban areas. Pathway 8 accounts for continental pollution advected off the coast and entrained into the marine boundary layer. Pathway 9 is a clean marine boundary layer aerosol, usually found in the remote ocean, and consisting primarily of sea salt. Pathways 10 and 11 are lofted aerosol layers (based on the Lofted Layer Flag from the SIBYL) over land and ocean, respectively, of biomass burning smoke. Note that elevated dust layers are found in Pathway 4.

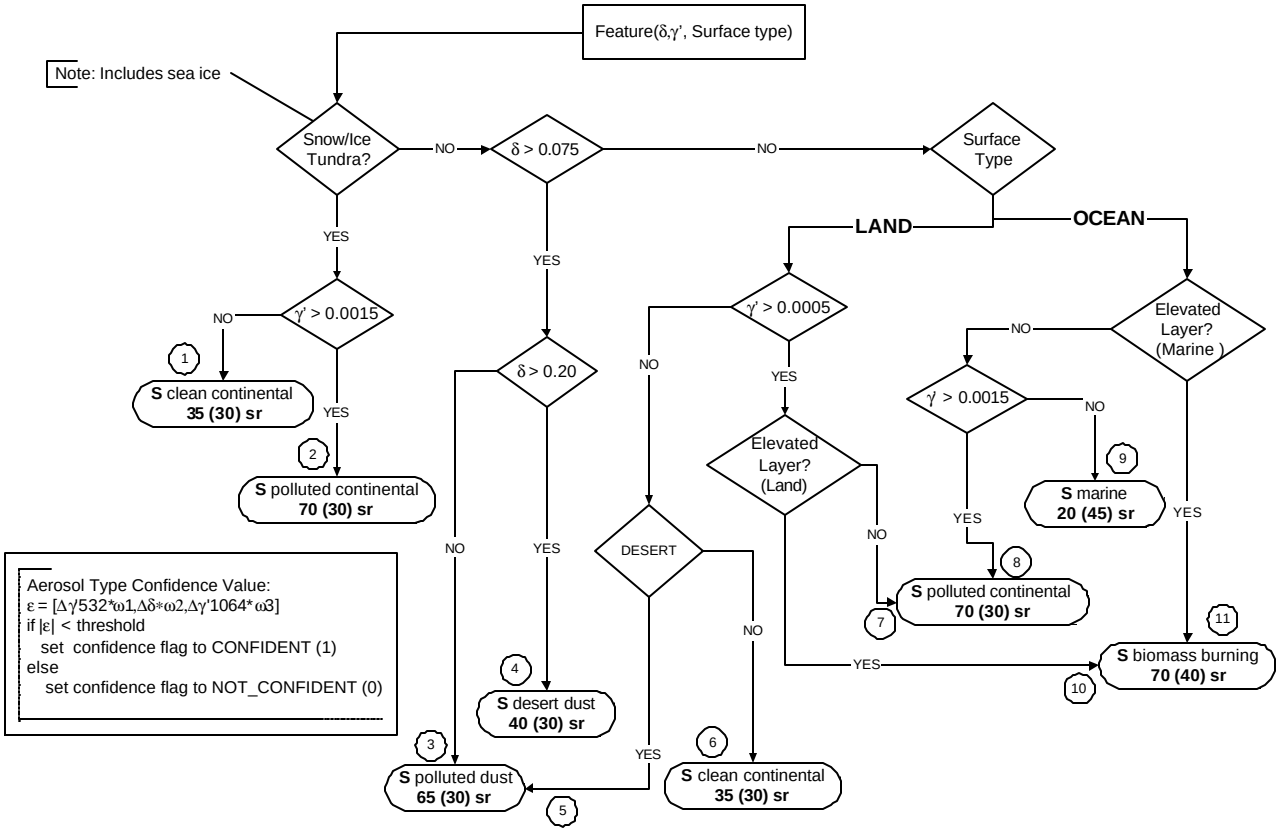


Figure 8.2: Flowchart of the CALIOP S_a selection scheme for tropospheric aerosols. The values shown are the extinction-to-backscatter ratios at 532 nm and 1064 nm in parentheses. Note that γ' is the integrated attenuated backscatter coefficient.

Studies are underway to determine optimum threshold values of d, $?'$, and (eventually) $?$ to be used in the typing scheme. The values shown in Figure 8.2 are initial estimates based on LITE measurements and, in the case of depolarization, on a limited set of observations and models [*Gobbi et al.*, 2000; *Murayama et al.*, 1999; *Reagan et al.*, 2001; *Sakai et al.*, 2003]. The goal is to base typing decisions on these observables as much as possible and avoid the major dependence on geographic information. Therefore, the threshold values of d, $?'$, and $?$ have been implemented as runtime parameters that can be adjusted using a configuration script. When lofted layers are encountered under favorable conditions, S_a is computed directly from the integrated backscatter and transmission.

Currently, data from the Cloud Physics Lidar (CPL, a two-wavelength, polarization lidar which flies on the NASA ER-2, [*McGill et al.*, 2003]) is being used to develop distributions of the lidar observables for use in the selection algorithm. We expect the algorithm will evolve significantly after the CALIPSO launch, when CALIOP will provide a much larger set of d, $?'$, and $?$ measurements than is currently available. Combined with carefully planned validation campaigns, this will allow improvements to the method and improved accuracy and confidence in the selection of S_a. The aerosol type confidence flags are functions of uncertainties in the 532 nm and 1064 nm integrated backscatter values and are discussed in the QA/QC section of the ATBD.

9. Selection of Cloud Model and Cloud Typing

The Cloud Model Algorithm (CMA) will determine consistent values of the cloud lidar ratio, S_c, and the multiple scattering function, $h_c(z)$, for all cloud layers. S_c is assumed to be constant throughout the cloud layer (except within mixed-phase clouds). As mentioned in Section 2, whenever possible, S_c will be derived directly from the measured data via the transmittance-constraint technique (see Section 7). Otherwise, a model-based lidar ratio for a given cloud layer will be selected. In either case, the selected value will be passed to the extinction algorithm for use in retrieving extinction and backscatter coefficients.

For water clouds, the SCA chooses a constant value of lidar ratio of 18 sr, based on theoretical studies [*Pinnick et al.*, 1983]. The approach used in determining a model-based value of S_c for cirrus clouds is to use an empirical relation between temperature and S_c, shown in Figure 9.1:, which is based on observations in the tropics and at mid-latitudes (Platt, private communication). According to Figure 9.1:, S_c is selected according to the simple rule:

$$S_c = -1.2591T - 6.698. \qquad (9.1)$$

where T is mid-cloud temperature (in Celsius) derived from the observed cloud layer heights using the ancillary met data product. For each determination of S_c, the CMA will estimate the uncertainty in this value based on the observed spread in values around the regression line.

It has been suggested that a correlation may exist between S_c and depolarization [*Reichardt et al.*, 2002, in press], in which case it would be beneficial to select S_c using both temperature and depolarization. However, the relationship proposed in *Reichardt et al.* [2002] is contradicted by other observations [*Eloranta*, 2000] and requires further study to determine if it is general.

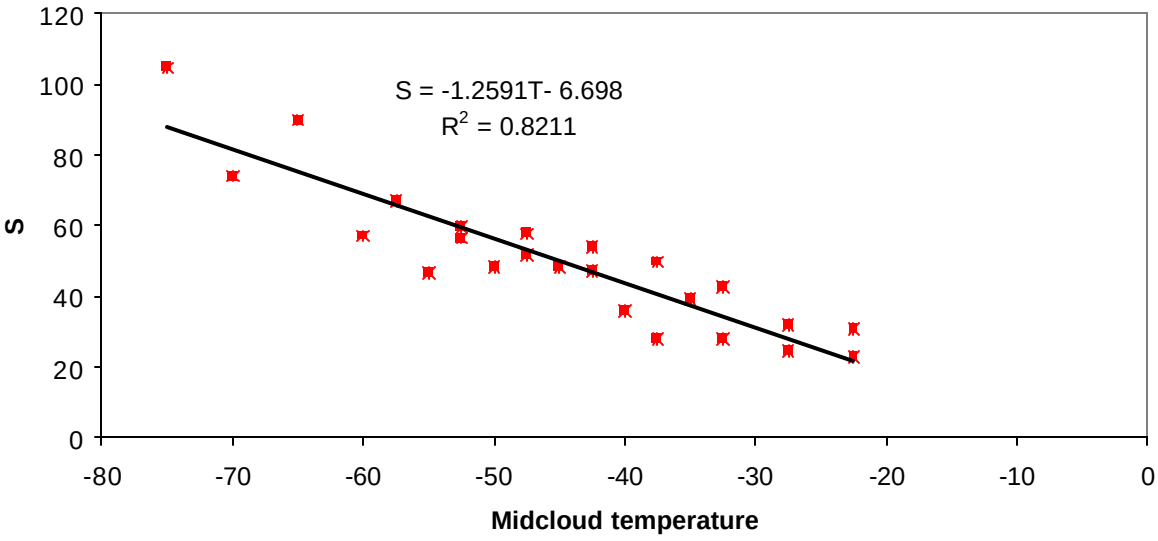
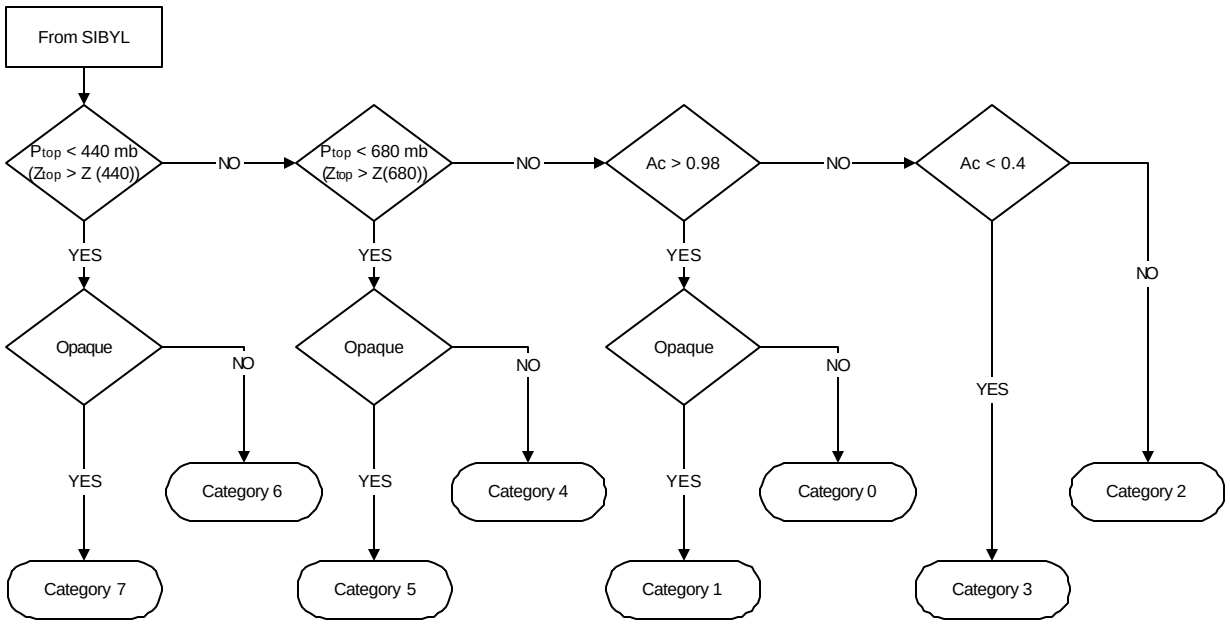


Figure 9.1: Empirical relation between lidar ratio and mid-cloud temperature (Platt, private comm.)

The multiple scattering function selected by the CMA should be consistent with the selected value of the lidar ratio. However, both the modeled and the observationally-derived S_c values

have been determined from direct measurements, without knowledge of multiple scattering effects or the ice crystal phase function. The definition of multiple scattering functions which are consistent with S_c values is provided in the multiple scattering section of this ATBD.

The SCA also performs cloud sub-typing based on lidar measurements. The ISCCP cloud-top pressure/optical depth classification scheme [*Rossow and Schiffer*, 1991] has been widely used for satellite cloud climatologies. Because CALIOP cannot measure optical depths greater than about 3, the ISCCP scheme cannot be used directly in the CALIOP cloud sub-typing. The ISCCP pressure boundaries at 680 mb and 440 mb are adopted, however, to classify low/middle/high clouds. The flow of the CALIOP cloud sub-typing algorithm is illustrated in Figure 9.2:. Cloud subtyping is based on cloud-top pressure derived from lidar cloud-top height and cloud fraction A_c. Clouds are classified as transparent or opaque by checking whether or not the surface or a lower layer can be seen (refer to the CALIOP Feature Finder ATBD). If the surface or a lower layer can be seen, then the layer is transparent; otherwise, the layer is opaque. We expect that cloud layers with optical depths greater than 3-5, depending on conditions, will appear opaque.



Category 0: Low, overcast, thin (transparent St, StCu, and fog)
Category 1: Low, overcast, thick (opaque St, StCu, and fog)
Category 2: Transition stratocumulus
Category 3: Low, broken (trade Cu and shallow Cu)
Category 4: Altocumulus (transparent)
Category 5: Altostratus (opaque, As, Ns, Ac)
Category 6: Cirrus (tranparaent)
Category 7: Deep convective (opaque As, Cb, Ns)

Figure 9.2: Flow of the sub-typing of clouds.

The cloud fraction, A_c, (used only for clouds with tops < 3 km) is computed along a line over an 80 km horizontal segment. For the baseline version of the code, this parameter is simply estimated from the ratio of the number of 1-km profiles with clouds found below 3 km to the number of the total 1-km profiles in the 80-km segment (i.e., 80). Further considerations are however necessary for more complicated cases. For example, within an 80-km segment there

may be overlying opaque clouds above 3 km. In this case, the detection of lower features is not possible due to the absorption of the overlying opaque cloud layers and hence the true cloud fraction cannot be observed. Treatment of these complicated cases will be addressed in future algorithm improvements The cloud sub-typing algorithm will be modified and improved using observational data acquired after the launch.

10. References

Ackermann, J., 1998: "The extinction-to-backscatter ratio of tropospheric aerosol: A numerical study", *Journal of Atmospheric and Oceanic Technology*, **15**, pp 1043-1050.

Anderson, T.L., S.J. Masonis, D.S. Covert, and R.J. Charlson, 2000: "In situ measurements of the aerosol extinction-to-backscatter ratio at a polluted continental site", *J. of Geophys. Res.*, **105** (D22), pp 26907-26915.

Ansmann, A., F. Wagner, D. Althausen, D. Müller, A. Herber, and U. Wandinger, 2001: "European pollution outbreaks during ACE 2: Lofted aerosol plumes observed with Raman lidar at the Portuguese coast", *J. Geophys. Res.*, **106** (D18), pp 20725-20733.

Bevington, P. R., and D. K. Robinson, 1992: Data Reduction and Error Analysis for the Physical Sciences, McGraw-Hill, 328 pp.

Beyerle, G., et al., 2001: "A lidar and backscatter sonde measurement campaign at Table Mountain during February –March 1997: Observations of cirrus clouds", *J. Atmos. Sci.*, **58**, pp 1275–1287.

Bosenberg, J., et al., 2003: "EARLINET: A European Aerosol Research Lidar Network to establish an aerosol climatology", MPI-Rep. 348, Max- Planck-Inst. fu¨r Meteorol., Hamburg, Germany.

Browell, E.V., et al., 1990: "Airborne lidar observations in the wintertime arctic stratosphere: polar stratospheric clouds", *Geophys. Res., Lett.*, **17**, pp 385-388.

Cattrall, C., J. Reagan, K. Thome, O. Dubovik, 2005: "Variability of aerosol lidar, backscatter and extinction ratios of key aerosol types derived from selected AERONET locations", *J. Geophys. Res.*, **110**, D10S11, doi:10.1029/2004JD005124.

Dubovik, O., and M.D. King, 2000: "A flexible inversion algorithm for retrieval of aerosol optical properties from Sun and sky radiance measurements", *J. Geophys. Res.*, **105**, pp 9791-9806.

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

Eloranta, E., R. E. Kuehn, and R. E. Holz, 2001: Measurements of backscatter phase function and depolarization in cirrus clouds made with the University of Wisconsin High Spectral Resolution Lidar. In Advances in Laser Remote Sensing, Proceedings of the 20th International Laser Radar Conference, A. Dabas, C. Loth, and J. Pelon, eds. (Editions de l'Ecole polytechnique, Palaiseau, France), 492 pp.

Fernald, F.G., B. M. Herman and J. A. Reagan, 1972: "Determination of aerosol height distributions with lidar", *Journal of Applied Meteorology*, **11**, pp 482-489.

Ferrare, R.A., D.D. Turner, L.H. Brasseur, W.F. Feltz, O. Dubovik, and T.P. Tooman, 2001: "Raman lidar measurements of the aerosol extinction-to-backscatter ratio over the Southern Great Plains", *J. Geophys. Res.*, **106** (D17), pp 20333-20347.

Flamant, C., V. Trouillet, P. Chazette, and J. Pelon, 1998: "Wind speed dependence of atmospheric boundary layer optical properties and ocean surface reflectance as observed by airborne backscatter lidar", *J. Geophys. Res.*, **103** (C11), pp 25137-25158.

Franke, K., A. Ansmann, D. Müller, and D. Althausen, 2001: "One-year observations of particle lidar ratio over the tropical Indian Ocean with Raman lidar", *Geophys. Res., Lett., In press.*

Gobbi, G. P., F. Barnaba, R. Giorgi, and A. Santacasa, 2000: "Altitude resolved properties of a Saharan dust event over the Mediterranean", *Atmos. Environ.*, **34**, pp 5119– 5127.

Hess, M., P. Koepke, and I. Schult, 1998b: "Optical properties of aerosols and clouds: The software package OPAC", *Bull. Am. Meteorol. Soc.*, **79**, pp 831– 844.

Hess, M., R. Koelemeijer, and P. Stammes, 1998a: "Scattering matrices of imperfect hexagonal ice crystals", *J. Quant. Spectrosc. Radiat. Transfer*, **60**, pp 301– 308.

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

Hu, Yong-X., D. Winker, P. Yang, B. Baum, L. Poole and L. Vann, 2001: "Identification of cloud phase from PICASSO-CENA lidar depolarization: A multiple scattering sensitivity study", *J. Quant. Spectros. Radiat. Trans.*, **70**, pp 569-579.

Kahn, R., P. Banerjee, and D. McDonald, 2001: "Sensitivity of multiangle imaging to natural mixtures of aerosols over ocean", *J. Geophys. Res.*, **106** (D16), pp 18219-18238.

Kalashnikova, O.V., and I.N. Sokolik, 2002: "Importance of shapes and compositions of wind-blown dust particles for remote sensing at solar wavelengths", *Geophys. Res. Lett.*, **29** (10), doi:10.1029/2002GL014947.

Liu, Z., M. A. Vaughan, D. M. Winker, C. A. Hostetler, L. R. Poole, D. Hlavka, W. Hart, and M. McGill, 2004: "Use of probability distribution functions for discriminating between cloud and aerosol in lidar backscatter data", *J. Geophys. Res.*, **109**, D15202, doi:10.1029/2004JD004732.

Liu, Z., N. Sugimoto, and T. Murayama, 2002: "Extinction-to-backscatter ratio of Asian dust observed by high-spectral-resolution lidar and raman lidar", *Appl. Opt.*, **41** (15), pp 2760-2767.

Masonis, S.J., T.L. Anderson, D.S. Covert, V. Kapustin, A.D. Clarke, S. Howell, and K. Moore, 2002: "A study of the extinction-to-backscatter ratio of marine aerosol during the Shoreline Environmental Aerosol Study", *Journal of Atmospheric and Oceanic Technology*, **20** (10), pp 1388–1402.

Matthias, V., and J. Bosenberg, 2002: "Aerosol climatology for the planetary layer derived from regular lidar measurements", *Atmos. Res.*, **63**, pp 221– 245.

McGill, M. J., D. L. Hlavka, W. D. Hart, J. D. Spinhirne, V. S. Scott, and B. Schmid, 2002: "The cloud physics lidar: Instrument description and initial measurement results", *Appl. Opt.*, **41**, pp 3725–3734.

McGill, M.J., D.L. Hlavka, W.D. Hart, E.J. Welton, and J.R. Campbell, 2003: "Airborne lidar measurements of aerosol optical properties during SAFARI-2000", *J. Geophys. Res.*, **108** (D13), doi:10.1029/2002JD002370.

Müller, D., F. Wagner, D. Althausen, U. Wandinger, and A. Ansmann, 2000: "Physical properties of the Indian aerosol plume derived from six-wavelength lidar observations on 25 March 1999 of the Indian Ocean Experiment", *Geophys. Res., Lett.*, **27** (9), pp 1403-1406.

Murayama, T., et al., 2001: "Ground-based network observation of Asian dust events of April 1998 in East Asia", *J. Geophys. Res.*, **106**, pp 18,345–18,359.

Murayama, T., H. Okamoto, N. Kaneyasu, H. Kamataki, and K. Miura, 1999: "Application of lidar depolarization measurement in the atmospheric boundary layer: Effects of dust and sea-salt particles", *J. Geophys. Res.*, **104** (D24), pp 31781-31792.

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

Palm, S., W. Hart, D. Hlavka, E. Welton, A. Mahesh, and J. Spinhirne, 2002: Geoscience Laser Altimeter System (GLAS) Algorithm Theoretical Basis Document Version 4.2: GLAS atmospheric data products, NASA Goddard Space Flight Center, Greenbelt, Md., Oct.

Pinnick, R. G., S. G. Jennings, P. Chylek, C. Ham, and W. T. Grandy, Jr., 1983, "Backscatter and extinction in water clouds", *J.Geophys. Res*. **88**, pp 6787–6796.

Platt, C. M. R., 1973: "Lidar and radiometer observations of cirrus clouds", *J. Atmos. Sci.,* **30** (6), pp 1191-1204.

Post, M. J., F. F. Hall, R. A. Richter, and T. R. Lawrence, 1982: "Aerosol backscattering profiles at $\lambda = 10.6$ μm", Appl. Opt. 21, 2442-2446.

Pruppacher, H. R. and J. D. Klett, 1997: Microphysics of clouds and precipitation, Kluwer Academic Publisher, London.

Pruppacher, H., 1995: "A new look at homogeneous ice nucleation in supercooled water drops", *J. Atmos. Sci.*, **52**, pp 1924-1933.

Rauber, R.M. and A. Tokay, 1991: "An explanation for the existence of supercooled water at the top of cold clouds", *J. Atmos. Sci.*, **55**, pp 2016-2038.

Reagan, J.A., K.J. Thome, and D.M. Powell, 2001: "Lidar Aerosol Ratio: Measurements and Models", in *International Geosciences and Remote Sensing Symposium (IGARSS)*, Sidney, Australia.

Reichardt J, S. Reidchardt, M. Hess, T. J. McGee, 2002: "Correlations among the optical properties of cirrus-cloud particles: Microphysical interpretation", *J. Geophys. Res.*, **107** (D21), 4562, doi:10.1029/2002JD002589.

Rossow, W. B. and R. A. Schiffer, 1991: "ISCPP Cloud Data Products", *Bull. Amer. Meteor. Soc*., **72**, pp 2-20.

Sakai, T., T. Shibata, K. Hara, M. Kido, K. Osada, M. Hayashi, K. Matsunaga, and Y. Iwasaka, 2003: "Raman lidar and aircraft measurements of tropospheric aerosol particles during the

Asian dust event over central Japan: Case study on 23 April 1996", *J. Geophys. Res.*, **108** (20), pp 4349.

Sassen K, S. Benson, 2001: "A midlatitude cirrus cloud climatology from the facility for atmospheric remote sensing. Part II: Microphysical properties derived from lidar depolarization", *J. Atmos. Sci.*, **58**, pp 2103-2112.

Sassen, K., 1991: "The polarization lidar technique for cloud research: A review and current assessment", *Bull. Am. Meteorol. Soc.*, **72**, pp 1848–1866.

Shupe, M.D., et al., 2001: "Cloud water contents and hydrometeror size during the FIRE Arctic Clouds Experiment", *J. Geophys. Res.*, **106**, pp 15015-15028.

Voss, K.J., E.J. Welton, P.K. Quinn, J. Johnson, A.M. Thompson, and H.R. Gordon, 2001: "Lidar measurements during Aerosols99, *Journal of Geophysical Research*, *106* (D18)", 20821-20831.

Young, S. A., 1995: "Lidar analysis of lidar backscatter profiles in optically thin clouds", *Appl. Opt.*, **34**, pp 7019-7031.