Scene Classification for the CALIPSO Lidar

Zhaoyan Liu

Hampton University, 23 Tyler St, Hampton, Virginia 23668
Tel: +1-757-728-6368 • Fax: +1-757-727-5090

Mark A. Vaughan

SAIC/NASA LaRC, Hampton, Virginia 23681

Lamont R. Poole, Chris A. Hostetler, and David M. Winker

NASA LaRC, Hampton, Virginia 23681

ABSTRACT

CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations) has been scheduled to launch in April 2004 to provide key measurements of aerosols and clouds needed to improve the current climate predictions. A key component of the lidar data processing algorithms for CALIPSO is the scene classification algorithm that determines whether particulate scatterers are cloud or aerosol particles and, for clouds, whether they are composed of water or ice. The scene classification algorithm makes use of the magnitude and backscatter color ratio of the signals at the 532 and 1064 nm and the polarization ratio at 532 nm. This paper introduces the theoretical bases that form the foundation for the development of the CALIPSO lidar scene classification algorithms.

1. Introduction

Because of significant uncertainties in the modeled radiative effects of aerosols and clouds, the discrepancies among climate change simulations reported by different climate models are very large. This results in low confidence in the current generation of climate predictions. The CALIPSO lidar, together with passive instruments onboard the same satellite and the Aqua satellite, which flies in formation with CALIPSO, will provide the comprehensive measurements of aerosol and cloud properties that are needed to validate present climate models and improve climate prediction. [1]

The CALIPSO lidar is a two-wavelength polarization-sensitive system. It will provide simultaneous atmospheric backscatter profiles at 532 nm and 1064 nm and depolarization profiles at 532 nm. With excellent profiling capability and global coverage the CALIPSO lidar will enable three-dimensional mapping of aerosols and clouds on a global scale. By applying scene classification techniques, discrimination between aerosols and clouds and, for clouds, determination of water/ice phase will also be possible. This classification is crucial for the accurate retrieval of aerosol and cloud optical properties and other instrument data processing. In particular, an aerosol vs. cloud determination is required to properly constrain the selection of the optical models used to determine the extinction-to-backscatter ratio (otherwise known as the lidar ratio) for extinction retrievals. Scene classification strategies for the CALIPSO lidar are currently under study. This paper discusses the theoretical bases for the CALIPSO scene classification algorithms.

2. Theoretical Basis for Scene Classification

Scene classification is performed based on the differences in the physical and scattering properties of aerosols and clouds that are reflected in lidar signals. To illustrate the general characteristics of aerosol and cloud scattering, Fig. 1 shows a plot of the 1064 to 532-nm backscatter color ratio vs. 532-nm backscatter coefficient derived using the OPAC software package.[2, 3] The scattering models used are maritime, continental, urban, desert, Arctic and Antarctic type aerosols, and stratus, cumulus and cirrus type clouds and fog. Three values of relative humidity (0%, 50%, and 90 %) are considered for the aerosols. The cirrus model consists of randomly oriented perfect or somewhat distorted hexagonal columns and ray-tracing code is used for the cirrus scattering computation.[3] Mie code is used for all other computations.[2] In general, as seen in Fig. 1, clouds have larger backscatter coefficients and higher color ratios (~ 1). Aerosols have smaller backscatter coefficients and lower color ratios. 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. Desert dusts have color ratio larger than one. This may not well represent the real value of dust aerosols, because in general the Mie theory does not well describe scattering properties of irregular dust particles. Those scattering features can be used to distinguish aerosols from clouds. In addition, depolarization ratio is a useful indicator to identify irregular particles,[4] and it provides the means to discriminate ice clouds and dust aerosols.

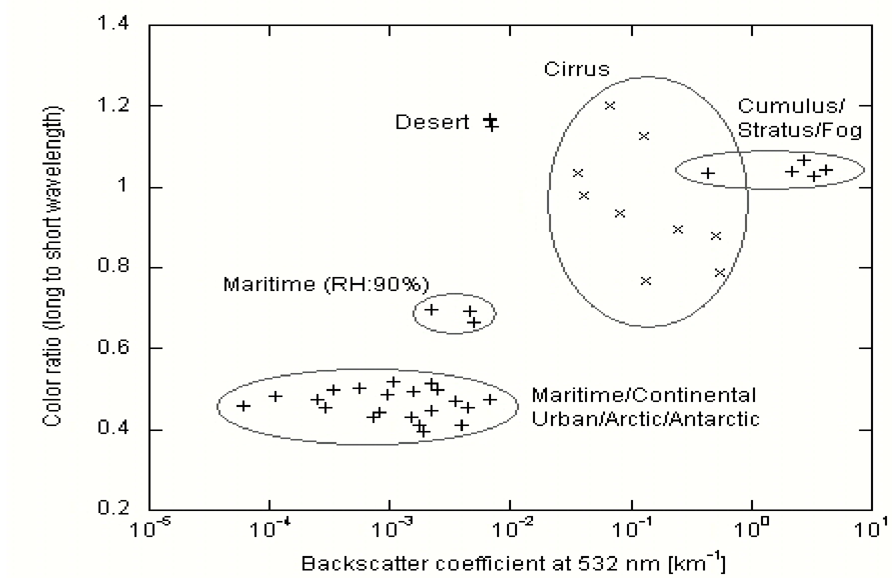

Fig. 1 Example of modeled aerosol and cloud scattering properties calculated using the OPAC software package.

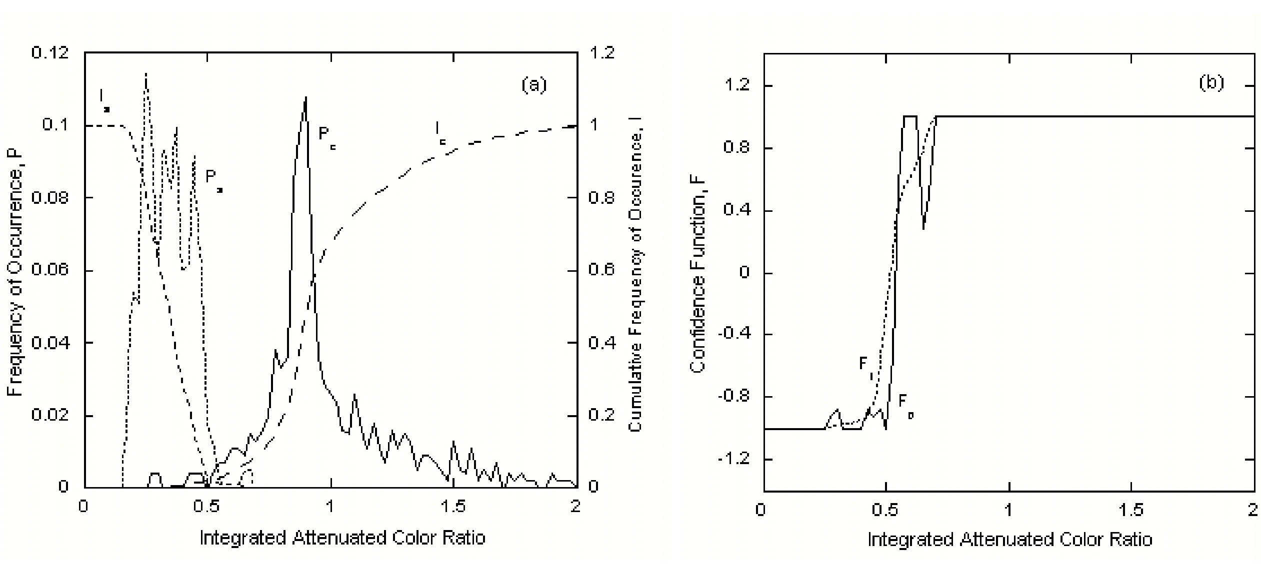

Fig. 2 Example of aerosol and cloud scattering properties obtained from the LITE observation (a), and confidence curves (b) as defined by Eqs (1a) and (1b). P_a and P_c are occurrence frequency, and I_a and I_c are cumulative occurrence frequency, respectively, for aerosols and clouds.

Figure 2(a) presents an example of the (cumulative) frequency of occurrence of integrated attenuated aerosol and cloud total color ratios derived from a small subset of the LITE observations. For about 98% of aerosols, the value of integrated attenuated total color ratio[§] is smaller than 0.54, meanwhile, for about 98% of clouds larger than 0.54. Such significant difference in aerosol and cloud scattering properties provides the basis for the lidar scene classification. A simple scheme, for example, is to set a threshold at 0.54. When the value of integrated attenuated color ratio is smaller than 0.54, the feature is interpreted as aerosol, and when larger than 0.54 the feature is interpreted as cloud.

However, as in the example in Fig. 2(a), there exists an ambiguous region where the aerosol and cloud frequency occurrence curves overlap. When the color ratio of an observed feature is within this region, the feature then cannot be identified definitely, i.e., although a decision can be made, there is the possibility that the decision is wrong. To quantify the confidence of a decision, the CALIPSO scene classification data product will introduce a confidence flag. The confidence function can be defined, for example, by

$$F_P(X) = \frac{n_2(X) - n_1(X)}{n_2(X) + n_1(X)} = \frac{P_2(X) - P_1(X)N_1/N_2}{P_2(X) + P_1(X)N_1/N_2} \tag{1a}$$

or

$$F_I(X) = \frac{I_2(X) - I_1(X)N_1/N_2}{I_2(X) + I_1(X)N_1/N_2} = \frac{\int_0^X P_2(X')dX' - N_1/N_2\int_X^\infty P_1(X')dX'}{\int_0^X P_2(X')dX' + N_1/N_2\int_X^\infty P_1(X')dX'} \tag{1b}$$

where X is the value of a given spatial or optical attribute used to classify the feature. For the CALIPSO scene classification algorithms X can be the (integrated) attenuated backscatter, (integrated) attenuated total color ratio, aspect ratio, depolarization ratio, etc. $n_k(X)$ and N_k are the number of occurrences of class k having attribute X and the total number of events for the k^{th} class, respectively. $P(X)$ is the probability distribution function (PDF) and $I(X)$ is the corresponding cumulative probability distribution function. Subscripts 1 and 2 refer to the aerosol and cloud in the case of aerosol and cloud discrimination, and water and ice in the case of cloud phase discrimination, respectively.

Figure 2(b) presents confidence curves derived from Fig. 2(a) using Eqs. (1a) and (1b). We note that $n_1/(n_1+n_2)$ and $n_2/(n_1+n_2)$ are the occurrence probability of feature class 1 and 2, respectively, and therefore F defined by (1a) is the differential occurrence probability. Likewise, F defined by (1b) is the weighted differential occurrence probability. The definition (1b) is insensitive to the possible variations in P due to the limited observation number as shown in Fig. (2b). For both definitions the value returned by the confidence function is bounded on [-1, 1]. The sign of a confidence function evaluation determines a feature's classification; for example, as in Fig. (2b), negative values could indicate aerosols and positive values clouds. The magnitude of the function (between 0 and 1) assigns a confidence to the classification. The value of zero gives the lowest confidence for any decision. The overlap region of two PDFs corresponds to low confidence.

One approach to resolving the ambiguity introduced into a single-test scene classification by the overlap of PDFs is simply to perform a number of different tests. When the value of one test is in a low confidence region, the value of other tests may be in a high region. In such cases an unambiguous decision can be made (for example) by appealing to the tests that have the highest confidence.

[§] The integrated attenuated color ratio, χ', is defined for any layer as $\chi' = \int_{top}^{base} B_{1064}(r)/B_{532}(r)\,dr$ where $B_\lambda(r) = \left(\beta_{\lambda,m} + \beta_{\lambda,p}\right)\cdot T^2_{\lambda,p}(r)$ and can be obtained directly from the calibrated backscatter data and a molecular model. Note that in addition to being modified by the two-way transmittance due to particulates (i.e., clouds or aerosols) both the molecular and the particulate backscatter coefficients contribute to the value of χ'. The color ratios shown in Fig. 1 represent the pure particulate backscatter color ratios, $\chi_p = \beta_{1064,p}/\beta_{532,p}$

A single classification decision can be synthesized from the (intermediate) results obtained from multiple tests in a number of different ways. For example, one very straightforward decision rule is given by

$$F = \max[|F_i|] \tag{2}$$

where subscript i =1,2,…,n refers to the number of the test. The sign of the confidence function of the test with maximum value indicates the feature class. The magnitude once again assigns a classification confidence level. Alternately, multiple test results can be combined using

$$F = \sum_{i=1}^{n} w_i F_i \tag{3}$$

where w_i is a weighting coefficient for the test X_i. Uncertainties in individual test results induced by noise in the measurements are an important factor to consider in choosing the weighting coefficient. For example, tests made on a high SNR measurement would have a larger weighting coefficient than similar tests made on a low SNR measurement. As with the previous formulations, the sign of the combined confidence function indicates the feature class and the magnitude assigns the classification confidence level. The relative merits of these combination approaches are currently under active investigation.

Preliminary studies of lidar observational data suggest however that a more effective method for reducing classification ambiguity is to employ multi-dimensional histogram analyses. Using this approach, the multiple-test scene classification may be improved by modifying Eq. (1a) using multiple-dimensional PDFs

$$F_p(X) = \frac{P_2(X_1,\cdots,X_n) - P_1(X_1,\cdots,X_n)N_1/N_2}{P_2(X_1,\cdots,X_n) + P_1(X_1,\cdots,X_n)N_1/N_2}\,. \tag{4}$$

To optimize test performance, other parameters such as altitude, latitude, temperature and season can be added to constrain the problem. Ultimately, lookup tables can be constructed based on observations and models using Eqs. (1) and (4).

Noise in the data can further reduce the confidence of a decision. Its effect should be estimated statistically and taken into consideration in developing the confidence flag for the decision. The classification confidence levels assigned by the CALIPSO scene classification algorithm will incorporate any additional uncertainties due to noise. A detailed assessment of the effects of noise on the classification process is currently being conducted.

3. Summary

The CALIPSO lidar will provide two-wavelength depolarization-sensitive observation of aerosols and clouds. This will make the discrimination between aerosol and cloud and cloud ice and water phase possible. In this paper, the theoretical bases for the CALIPSO scene classification algorithms have been described. The problem is still under study to develop robust algorithms. A key step is to derive the probability distribution functions for the spatial and optical properties of aerosols and clouds.

REFERENCES

1. D. Winker, "The CALIPSO Mission, " the Proceeding of 21st ILRC, 2002.
2. M. Hess, P. Koepke, and I. Schult, "Optical Properties of Aerosols and clouds: The software package OPAC," *Bull. Am. Met. Soc.*, 79, 831-844, 1998.
3. M. Hess, R. B. A. Koelemeijer, and P. Stammes, "Scattering matrices of imperfect hexagonal ice crystals," *J. Quant. Spectros. Radiat. Transfer*, 60, 301-308, 1998.
4. K. Sassen, "The polarization lidar technique for cloud research: A review and current assessment," *Bull. Am. Met. Soc.*, 72, 1848-1866, 1991.