A CLOUD-AEROSOL DISCRIMINATION ALGORITHM FOR CALIPSO LIDAR OBSERVATIONS: ALGORITHM TESTS

Zhaoyan Liu [(1)], Ralph Kuehn [(2)], Mark Vaughan [(2)], David Winker [(3)], Kathleen Powell [(2)], Chris Hostetler [(3)], Lamont Poole [(3)], and Matthew McGill [(4)]

[(1)] Hampton University, MS 435, NASA Langley Research Center, Hampton, VA 23681 USA,
E-mail: z.liu@larc.nasa.gov

[(2)] SAIC, MS 435, NASA Langley Research Center, Hampton, VA 23681 USA,
E-mail:r.a.kuehn@larc.nasa.gov; m.a.vaughan@larc.nasa.gov

[(3)] NASA Langley Research Center, MS 435, Hampton, VA 23681 USA,
E-mail:David.M.Winker@nasa.gov; Chris.A.Hostetler@nasa.gov; Lamont.R.Poole@nasa.gov

[(4)] NASA Goddard Space Flight Center, Code 912, Greenbelt, Ml 20771 USA,
E-mail: Matthew.J.McGill@nasa.gov

ABSTRACT

An algorithm based on three-dimensional probability distribution functions (PDFs) has been developed for discriminating between clouds and aerosols in the lidar data that will be collected during the Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observations (CALIPSO) mission. The algorithm has been tested with a data set acquired by the Cloud Physics Lidar (CPL) during the THORPEX-PTOST 2003 campaign. Comparisons with feature classifications made by an existing two-dimensional algorithm have also been conducted, and in general the results obtained by the two methods are in good agreement.

Dust presents a special case. Because the intrinsic scattering properties of dust layers can be very similar to those of clouds, additional tests focused on distinguishing dust from clouds are highly desirable. In this paper we briefly introduce the CALIPSO cloud-aerosol discrimination algorithm. We then present a case study conducted using a layer of mixed smoke and dust observed by the CPL during the SAFARI campaign in the Southern Africa region. For this particular layer, a success rate of close to 100% is achieved.

1. INTRODUCTION

When flown in space, backscatter lidars can provide continuous measurements of clouds and aerosols on a global scale and with excellent spatial resolution. This unique ability was first demonstrated by NASA's Lidar In-space Technology Experiment (LITE) [1]. However, the quantitative retrieval of cloud and aerosol optical properties, including backscatter and extinction profiles and layer optical depths, requires knowledge of the lidar ratio. Given a sufficiently accurate measurement of the layer two-way transmittance [2, 3], or simultaneous two-wavelength, high signal-to-noise ratio (SNR) measurements for which the backscatter profiles satisfy some additional similarity requirements [3-5], the lidar ratio can be retrieved from the lidar observation alone. However, for space-based lidar measurements these conditions are only occasionally satisfied, and therefore specific values of lidar ratio must be selected in the data processing according to an informed estimate of the layer type and/or composition. In this latter case, accurate selection of lidar ratio relies on a layer classification scheme that determines layer type based on inferential tests applied to the directly measurable optical and physical properties of the layer [6]. These properties include attenuated backscatter coefficients, attenuated volume color ratios, volume depolarization ratios, layer top, base, and/or center height, geophysical location, season etc.

CALIPSO [7] represents the next generation of space-based lidars. The project is an international effort, being jointly developed by NASA and the French space agency CNES. The CALIPSO payload consists of a polarization-sensitive two-wavelength lidar, an imaging infrared radiometer, and a wide field camera. An enormous amount of data will be acquired during three years of observations that will begin with a launch scheduled in early 2005. This huge volume necessitates the use of a fully automated data analysis system. To this end, a collection of intelligent algorithms for automated CALIPSO lidar data processing is being developed [6, 8-12]. Of fundamental importance in this processing is the accurate discrimination between clouds and aerosols in the backscatter data; the cloud-aerosol discrimination function is crucial to the success of interpreting the CALIPSO lidar observations and to the selection of a lidar ratio that will yield highly reliable data products.

Classification techniques for distinguishing between two separate classes based on probability distribution functions (PDFs) have been investigated [8]. These classification approaches are based either on a single test or on multiple tests that use a confidence function (*f*-function) constructed from 1-dimension or multiple-dimension (1-D or multiple-D) PDFs to distinguish between two classes. The result of these studies is an operational algorithm that has been developed for the CALIPSO lidar cloud and aerosol discrimination [8]. This algorithm is based on three-dimensional PDFs. The test attributes used are the layer averaged attenuated backscatter coefficient, the volume color ratio, and the layer center altitude. Extensive tests of algorithm performance were conducted using 49 hours of down-looking lidar data obtained by the Cloud Physics Lidar (CPL) [13] during the THORPEX-PTOST 2003 campaign. Comparisons were conducted with the feature type classifications determined by a two-dimensional discrimination algorithm that has been developed for use with the Geoscience Laser Altimetry Satellite (GLAS) lidar observations [14]. In general, good agreement was obtained; only ~5.7% of a total of 228,264 layers analyzed were classified as different types by the two algorithms. Case studies indicated that this disparity came mainly from the incorrect classification of optically thin clouds as aerosol by the two-dimensional algorithm. The CALIPSO algorithm achieves a significant improvement in correctly identifying this type of feature. This improvement is due largely to the use of a three-dimensional approach, because, as pointed out in [8], the separation of cloud and aerosol classes is more complete in a higher dimensional space than in a lower dimensional space. The degree of separation of cloud and aerosol classes is an essential limit on the performance of any scene classification scheme.

Because the intrinsic scattering properties of dust layers can be very similar to those of clouds, additional algorithm testing was performed using an optically dense layer of Saharan dust measured during the Lidar In-space Technology Experiment (LITE). In general, the method was shown to distinguish reliably between dust layers and clouds [8]. Erroneous classifications could occur in those regions of the Saharan dust layer where the optical thickness was the highest. More testing is currently being conducted using the CPL data set acquired during the Southern African Regional Science Initiative (SAFARI). The SAFARI campaign, conducted during August and September of 2000, placed particular emphasis on acquiring measurements of biomass burning and regional emissions [15]. In this paper we will briefly introduce the algorithm developed for the CALIPSO lidar cloud and aerosol discrimination. We will then present a classification case study applying our algorithm to a layer of dust-smoke mixture measured by the CPL during the SAFARI campaign [15].

2. ALGORITHM DESCRIPTION

Fig. 1 presents a flowchart of the algorithm developed for the CALIPSO lidar cloud-aerosol discrimination. To analyze a feature, the inference engine first reads in the feature's layer characteristics, including the mean attenuated backscatter β' and volume color ratio χ' (i.e., the ratio of the mean attenuated backscatters at 1064 nm and 532 nm), together with their uncertainties, $\Delta\beta'$ and $\Delta\chi'$, and the layer center altitude z. All these products are computed in the initial phase of the CALIPSO level 1-B data processing. The confidence function is then computed using [8]

$$f_n(\beta',\chi',z) = \frac{P_{2,n}(\beta',\chi',z) - P_{1,n}(\beta',\chi',z)K_s}{P_{2,n}(\beta',\chi',z) + P_{1,n}(\beta',\chi',z)K_s}. \qquad (1)$$

Here $P_{1,\mathrm{n}}$ and $P_{2,\mathrm{n}}$ are the noise-affected PDFs for clouds and aerosols, respectively. K_s is a scale factor that quantifies the relative occurrence frequency of the cloud and aerosol classes and is, in practice, implicitly included in the scaled, noise-free PDF files. The noise-free PDFs are retrieved from a previously developed database. The noise-affected PDF is the convolution of the noise-free PDF and the noise distribution. Details can be found in [8].

Based on the value of the *f*-function, the inference engine derives both a classification and a measure of the confidence ascribed to that classification. The sign of the *f*-function determines the feature's class; negative values indicate aerosol and positive values cloud. The magnitude of *f* (between 0 and 1) assigns a confidence *Q* to the classification. A value of zero indicates that no classification can be made.

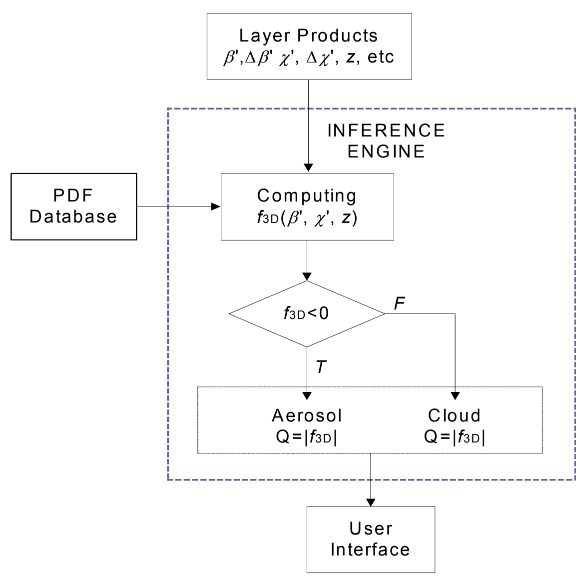

Fig.1. Flowchart of the algorithm for the CALIPSO lidar cloud-aerosol discrimination.

In general, if a layer has smaller values of both mean attenuated backscatter and mean attenuated volume color ratio, that layer is very likely to be classified as an aerosol. Similarly, higher values are likely to result in the layer being classified as a cloud.

3. CASE STUDY

As expected by the campaign coordinators, during SAFARI CPL observed a large amount of smoke and dust aerosols. The CPL SAFARI data set consequently provides a good data source for testing the ability of the CALIPSO discrimination algorithm to identify these particular types of aerosols. Fig.2 presents an example of the algorithm test with the CPL SAFARI data set. Fig.2(a) shows 532-nm attenuated backscatter profiles measured during the flight on 14 September 2000 passing over the coast of Namibia. An aerosol layer transporting from the continent over to the ocean is seen. This layer is dominated by dust and smoke aerosols [15]. The presence of smoke was due largely to the heavy biomass burning that frequently occurs in and to the north of Zambia in August and September.

Figs. 2(c) and (d) show mean attenuated volume color ratios and mean backscatter coefficients that were derived by averaging the attenuated backscatter measurements between the upper and lower boundary of the aerosol layer (including the cloud layer on the top of the aerosol layer). A feature finder [14] has been applied to detect the boundaries of features. In general, the mean attenuated volume color ratio increases from the continent to the ocean; a maximum increment of over 2 times is seen. Conversely, the mean attenuated backscatter decreases. The opposing trend of these parameters implies that the mixing ratio of smoke and dust (possibly contaminated somewhat with other type aerosols) is changing along the track.

This smoke-dust layer has then been chosen for the case study to assess the performance of the CALIPSO algorithm to discriminate smoke and dust aerosols from clouds. The input to the algorithm testing includes the 532 nm mean attenuated backscatter, the attenuated volume color ratio (e.g., values in Figs. 2(c) and (d) for the mixed aerosol layer and the upper cloud layer), and the layer center altitude. Feature locations together with the feature types classified by the CALIPSO algorithm are presented in Fig. 2(b). Color is used to denote different feature type: blue indicates cloud, red aerosol. Over 10,000 profiles were acquired in the aerosol layer, and only a few of these profiles have been classified as cloud. The upper dense cloud has also been correctly identified for the most part. We note however that, in situations such as shown here, where the cloud layer is adjacent to the aerosol layer at the optically thin part of the cloud, features detected by the feature finder are not guaranteed to be homogeneous: they can instead be a mixture of cloud and aerosol. Furthermore, the cloud-aerosol discrimination algorithm does not consider mixed layers as a separate class; the layer is classified as either cloud or aerosol depending on what type predominates the layer. However, for the aerosol-only part of the test scene, a success rate of close to 100% has been achieved using the CALIPSO algorithm.

It is also seen that most low broken clouds and the PBL aerosols over the ocean have been correctly identified, although an exact success rate is hard to quantify because it is difficult to label these low features accurately. The misclassifications however happen mostly at the edges of clouds, where clouds and aerosols are very likely to be detected by the feature finder as a single (mixed) layer.

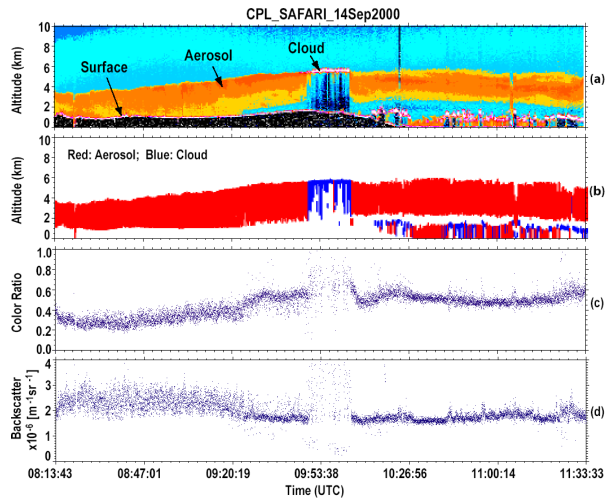

Fig.2. 532 nm attenuated backscatter profiles (a), detected feature layers (b), and layer-averaged attenuated volume color ratio (c) and backscatter (d) obtained from the CPL observation on 14 September 2000. The color in (b) denotes different feature types classified by the CALIPSO cloud-aerosol discrimination algorithm: blue indicates cloud, red aerosol.

4. DISCUSSION

This case study shows that the CALIPSO algorithm can correctly classify the chosen mixed smoke-dust aerosol layer, even though this layer changed its optical scattering properties significantly along the transport path. An extensive algorithm test reported separately in [8] demonstrated that the CALIPSO algorithm could also discriminate other type aerosols from clouds with high success rates. When classifying high, optically thin clouds, the three-dimensional approach based CALIPSO algorithm performs better than the two-dimensional algorithm originally developed for GLAS lidar observations. This is because a three-dimensional space provides a better separation of cloud and aerosol classes than in a two-dimensional space [8]. Fig. 3 shows scatter plots of (a) mean attenuated backscatter at 532 nm and (b) mean attenuated volume color ratio (1064 nm/532 nm) of all features found during the

THORPEX-PTOST 2003 campaign. The color is once again used to denote the feature type as classified by the CALIPSO algorithm: blue indicates cloud, red aerosol. A very good separation of cloud and aerosol classes is seen above ~1.5 km in the mean attenuated color ratio-altitude space. Introducing tests based on mean attenuated volume color ratio in the CALIPSO algorithm greatly helps the classification of clouds and aerosols in this region. However, the separation seen below ~1.5 km is somewhat less encouraging, and the misclassifications made by the CALIPSO algorithm were concentrated in the lowest altitudes. Although the actual false rate in this region is still low (few percent), additional testing and algorithm refinement is required to improve the classifications made at low altitudes. This task will occupy the authors of this study in the immediate future.

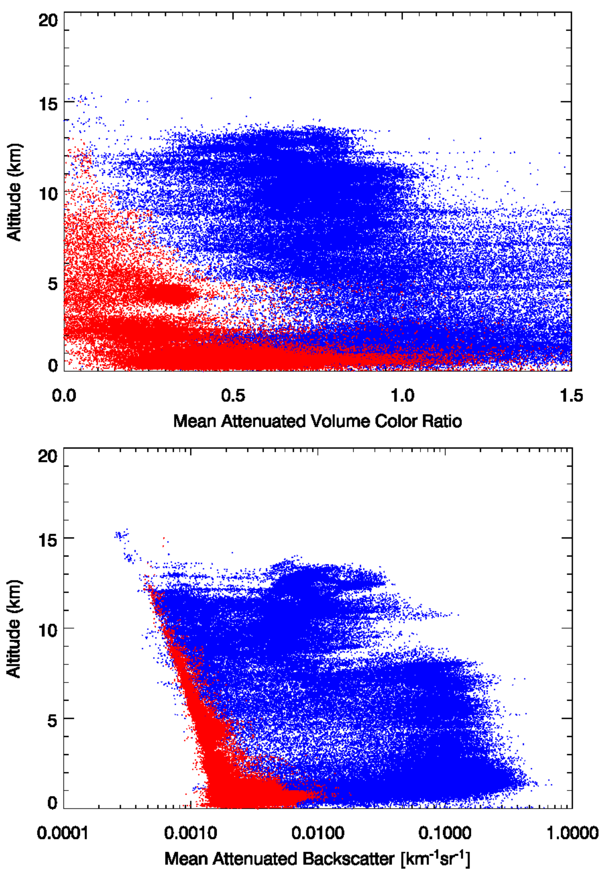

Fig.3. Scatter plots of layer-averaged, attenuated volume color ratio (upper panel) and backscatter [km^{-1} sr^{-1}] (lower panel) as a function of altitude of 228,264 features found from all ten CPL flights during the THORPEX-PTOST 2003 campaign. Blue denotes cloud type, red aerosol type, as classified by the CALIPSO algorithm.

REFERENCES

1. Winker D., et al., An Overview of LITE: NASA's Lidar In-space Technology Experiment, *Proc. IEEE*, 84, p.164, 1996.
2. Young S., Analysis of Lidar Backscatter Profiles in Optically Thin Clouds, *Appl. Opt.*, 34, p.7019, 1995.
3. Liu Z., et al., Simulations of the Observation of Clouds and Aerosols With the Experimental Lidar in Space Equipment system, *Appl. Opt.*, 39, pp.3120-3137, 2000.
4. Sasano Y. and Browell E., Light Scattering Characteristics of Various Aerosol Types Derived From Multiple Wavelength Lidar Observations, *Appl. Opt.*, 28, p.1670, 1989.
5. Vaughan M., Algorithm for Retrieving Lidar Ratios at 1064 nm From Space-Based Lidar Backscatter Data, *Proc. of SPIE*, 5240, p.104, 2004.
6. Liu Z., et al., Scene Classification for the CALIPSO Lidar, *Proc. of the 21st ILRC*, p.785, 2002.
7. Winker D., et al., The CALIPSO mission: spaceborne lidar for observation of aerosols and clouds, *Proc. of SPIE*, 4893, p.1, 2003.
8. Liu Z., et al., Use of Probability Distribution Functions for Discriminating Between Cloud and Aerosol in Lidar Backscatter Data, submitted to *JGR*, 2004.
9. Reagan J., et al., Spaceborne lidar calibration from cirrus and molecular backscatter returns, *IEEE Trans. Geosci. Remote Sens.*, 40, p.2285, 2002.
10. Vaughan M., et al., SIBYL: A selective iterated boundary location algorithm for finding cloud and aerosol layers in CALIPSO lidar data, *Proc. of the 21st ILRC*, p.791, 2002.
11. Omar A., et al., Estimation of aerosol extinction-to-backscatter ratios using AERONET measurements and cluster analysis, *Proc. of the 21st ILRC*, p.373, 2002.
12. Hu Y., et al., Multiple scattering effect and its impact on water/ice determination for PICASSO lidar, *6th International Congress on Optical Particle Characterization*, 3-5 April 2001.
13. McGill M., et al., The Cloud Physics Lidar: Instrument description and initial measurement results, *Appl. Opt.,* 41, p.3725, 2002.
14. Palm S., et al, Geoscience Laser Altimeter System (GLAS) Algorithm Theoretical Basis Document Version 4.2: GLAS Atmospheric Data Products, NASA Goddard Space Flight Center, 2002.
15. McGill M., et al., Airborne Lidar Measurements of Aerosol Optical Properties During SAFARI-2000, *J. Geophys. Res.*, 108, No. 8493, 2003.