THE CALIPSO CLOUD AND AEROSOL DISCRIMINATION: VERSION 3 ALGORITHM AND TEST RESULTS

Z. Liu[1], R. Kuehn[2], M. Vaughan[3], D. Winker[3], A. Omar[3], K. Powell[3], C. Trepte[3], Y. Hu[3], and C. Hostetler[3]

[1]*National Institute of Aerospace, Hampton, VA, USA, Zhaoyan.liu@nasa.gov*
[2]*Science System and Applications, Inc., Hampton, VA, USA*
[3]*NASA Langley Research Center, Hampton, VA, USA*

ABSTRACT

A new five-dimensional (5D) PDF-based cloud and aerosol discrimination (CAD) algorithm has been developed for use in the version 3 CALIPSO lidar data release. Because the separation between clouds and aerosols is better in the new 5D space than in the 3D space previously used in the version 2 algorithm, significant improvements have been achieved in the classification of dense aerosol layers. These improvements are particularly noticeable for very dense dust layers, which were frequently misclassified as cloud by the 3D algorithm in the V2 data release. This paper describes the V3 CAD algorithm and presents test results.

1. INTRODUCTION

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite was launched in April 2006 [1]. The main instrument onboard the CALIPSO payload is a two-wavelength (532 nm and 1064 nm) polarization ratio-sensitive backscatter lidar. Since starting scientific observations in June 2006, the CALIPSO lidar has acquired a global, nearly continuous record of vertical-resolved measurements of clouds and aerosols.

Currently, standard CALIPSO level 1 (L1) and level 2 (L2) lidar products are routinely produced at NASA's Langley Research Center. The L1 data products primarily contain a set of calibrated vertically resolved profiles (i.e., 532 nm total and perpendicular attenuated backscatters and 1064 nm total attenuated backscatter). These L1 profiles are further analyzed in the L2 processing to derive the optical and physical properties of clouds and aerosols. L2 processing includes layer detection, scene classification, and extinction retrieval. First, the layer detection algorithm finds features (clouds, aerosols, surface and subsurface, stratospheric layers, etc.) by searching for regions of enhanced signals in the attenuated backscatter profiles. After finding features, mean values of the 532 nm and 1064 nm attenuated backscatter, attenuated total color ratio, and volume depolarization ratio are computed for each atmospheric layer detected. These layer optical properties, along with physical properties such as top and base heights, latitude, longitude, etc., are reported in the L2 layer products and used by the scene classification algorithms (SCA). A primary function of the SCA is to select an extinction-to-backscatter ratio for the retrieval of particulate extinction and backscatter coefficients and depolarization ratios. Cloud and aerosol discrimination (CAD) is the first step of the SCA, and an accurate classification is critical to the success of all downstream data processes.

The CAD algorithm is a multi-dimensional probability density function (PDF) based approach [2]. The PDFs used in the V2 data release are 3D, including attributes (dimensions) of the mean attenuated backscatter (MAB) at 532 nm, the 1064/532 layer-integrated attenuated backscatter ratio (or total color ratio, TCR), and the midlayer altitude [2]. This set of PDFs was developed prelaunch based on existing airborne and spaceborne lidar data sets. There are a few scenarios which are frequently misclassified in the V2 data release. A typical one is the very dense dust layers associated with dust storms over the source regions. Misclassifications occur because the scattering characteristics of these dense dust layers are nearly identical to those of optically thin clouds in the 3D space [3]. By adding dimensions of volume depolarization ratio (VDR) and latitude, a new set of 5D PDFs has been developed. The scattering characteristics in these new PDFs are based exclusively on measurements made by the CALIPSO lidar. Tests results show that a significant improvement is achieved with the 5D algorithm for very dense dust and smoke layers located over/near the source regions, because these dense aerosol layers are well separated from clouds in the 5D space. In this paper we describe the new 5D CAD algorithm and present some test results.

2. V2 AND V3 CAD ALGORITHMS

The CAD algorithm is driven by the following confidence function (f_{3D} for the V2 release and f_{5D} for the V3 release):

$$f_{3D} = \frac{P_c(\beta'_{532},\chi',z) - P_a(\beta'_{532},\chi',z)}{P_c(\beta'_{532},\chi',z) + P_a(\beta'_{532},\chi',z)}, \qquad (1a)$$

$$f_{5D} = \frac{P_c(\beta'_{532},\chi',\delta',z,\text{lat}) - P_a(\beta'_{532},\chi',\delta',z,\text{lat})}{P_c(\beta'_{532},\chi',\delta',z,\text{lat}) + P_a(\beta'_{532},\chi',\delta',z,\text{lat})}. \qquad (1b)$$

In these equations, P_c and P_a are the PDFs for cloud and aerosol, respectively. β'_{532} is the layer averaged attenuated backscatter (MAB), χ' the total color ratio (TCR), δ' the volume depolarization ratio (VDR), z the altitude, and *lat* is the latitude. The function f is a normalized differential probability that ranges from -1 to 1. The CAD score reported in the CALIPSO L2 products is a percentile (integer) of f ranging from -100 to 100. A feature is classified as cloud when $f \geq 0$ and as aerosol when $f < 0$. The absolute value of the CAD score provides a confidence level for the classification.

The CAD performance is limited ultimately by the degree of separation between clouds and aerosols in the selected attributes space. Better performance can typically be achieved in a space with more dimensions, because clouds and aerosols separate more completely in a higher dimensional space [2]. As an example, Figure 1 shows 2D distributions of feature occurrence numbers as a function of TCR and MAB (upper panels) and a function of VDR and MAB (lower panels). Each plot counts all layers detected between 2 km and 3 km for the entire globe and for four 10° latitude bands, and is derived from a four month test data set (January, May, and August 2007, and January 2008) produced in preparation for the V3 data release. As shown in the upper panels of Figure 1 aerosols generally have relatively small TCR and MAB values, while clouds have large TCR and MAB values. Some overlap is seen between clouds and aerosols in all latitude bands in the TCR-MAB space. For the lower latitudes (20N-40N) shown in Figure 1, the overlap mostly consists of dense dust and polluted aerosols over the source regions and optically thin water clouds for this 2-3 km altitude range. This overlap moves toward smaller TCR and MAB values as the latitude range moves northward toward the Arctic, where a sizable fraction of ice clouds exists in the 2-3 km altitude range, and the occurrence frequency for dense aerosols is small as shown in the lower panels. In a global view, there is a larger overlap region between clouds and aerosols in the TCR-MAB space as seen in the upper left-hand-side panel.

In the V2 algorithm, a single set of altitude-resolved TCR-MAB (3D) PDFs was applied globally. Because the PDFs for dense dust and smoke aerosols overlap with the PDFs for optically thin clouds, these dense aerosols as well as some optically thin clouds could be misclassified when the CAD decision is based solely on the 3D PDFs. Therefore, in the V2 algorithm, an additional test on VDR with a latitude dependent threshold is performed to minimize the misclassification of optically thin ice clouds in the high latitudes [3]. A special CAD score of 101 was then assigned to the cloud layers that were initially classified as aerosol by the 3D PDFs and reclassified as cloud by the VDR test. However, the dense dust layers and thick pollution plumes are both well resolved from clouds in the MAB-TCR-VDR-altitude-latitude 5D space, as indicated by the broken ellipsoid curve in the lower panels of Figure 1. In the V3 data release, these dense aerosol features are classified correctly using the 5D algorithm alone. The special CAD score of 101 is therefore not used in the V3 release, because the additional VDR test used in the 3D algorithm is not required by the 5D algorithm.

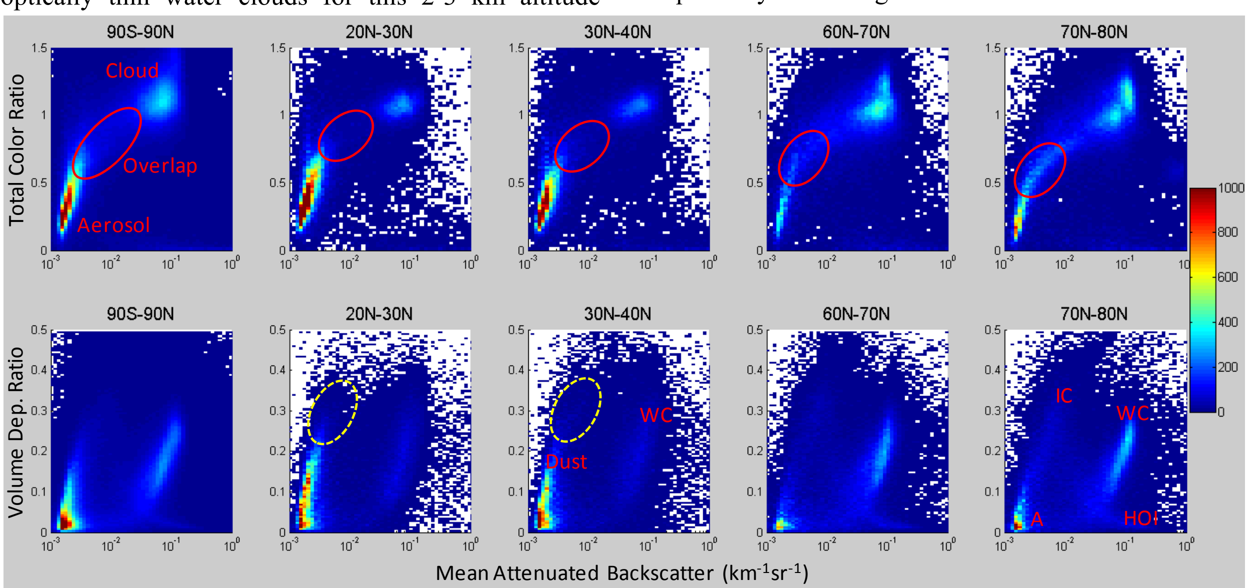

Figure 1. 2D distribution of feature occurrence numbers as a function of layer integrated total color ratio (TCR) and mean attenuated backscatter (MAB) (upper panels) and a function of layer integrated volume depolarization ratio (VDR) and MAB (lower panels) at 2-3 km altitudes for the entire globe (90S-90N) and for four latitude bands (20N-30N, 30N-40N, 60N-70N, and 70N-80N) from a four month V3 test data. Ellipsoid curves in the upper panels indicate an overlap region of cloud and aerosol PDFs in TCR-MAB space. Dashed ellipsoids in the lower panel denote the region in VDR-MAB space where dense dust layers generally present. A = aerosol, WC = water cloud, IC = ice cloud, and HOI = horizontally oriented ice.

3. TEST RESULTS AND DISCUSSIONS

The 5D CAD algorithm has been tested with the four month test data shown in Figure 1, and the results show that significant improvements in the discrimination of clouds from aerosols have been achieved. Figure 2 compares the V2.01 production and V3 test results using an example of the very dense dust layers that were often misclassified in the V2.01 data release. Extensive dust layers were observed south of ~31°N below 2 km over North Africa by the CALIPSO lidar on January 1, 2007. The densest parts of this dust layer have large backscatter values (grayish colors) that are similar in magnitude to those found in the cirrus cloud north of 38°N above 5 km. While these dust regions were misclassified as cloud by the 3D algorithm used in the V2.01 release (middle panel), they are now classified correctly as aerosol by the 5D algorithm (bottom panel). Note that some very dense parts appear to be still misclassified in the V3 test. It should be noted, however, that these parts of the layer were detected at single shot resolution, and currently all layers detected at single shot resolution are automatically classified as clouds, and are never analyzed by the standard CAD algorithm. Additional improvements were also seen for dense polluted/smoky aerosols over and near the source regions, although we do not show examples in this paper.

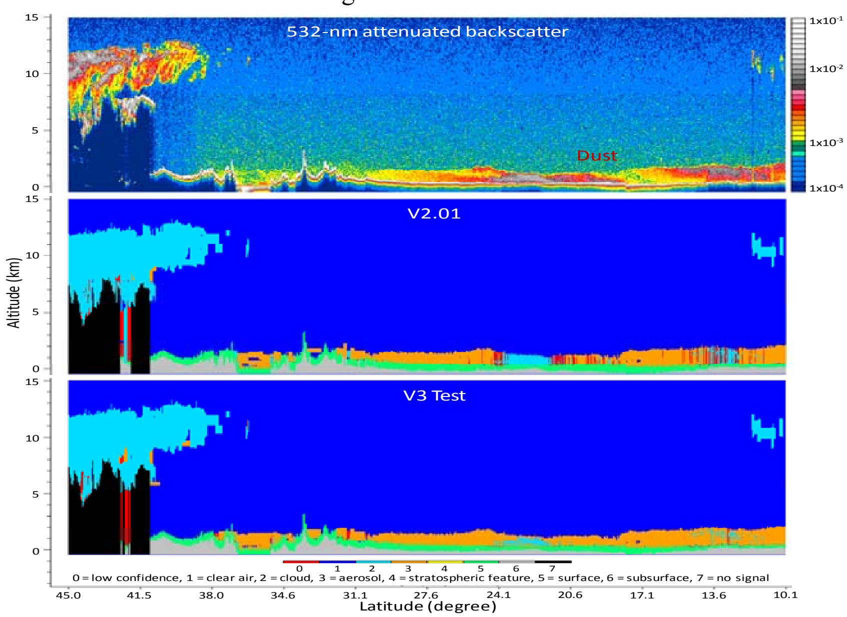

Figure 2. 532-nm attenuated backscatter (upper) and vertical feature mask (VFM) from the V2.01 data products (middle) and a V3 test data (bottom). The measurement was made by CALIPSO on Jan.1, 2007 during a nighttime orbit passing over Africa. In VFM (middle and lower panels), aerosol layers are color coded in orange, clouds in light blue, and the areas in red are mostly features with |CAD| < 20 and small number of invalid features (<0.01%).

Another notable improvement is that the new 5D algorithm can better identify features located far from the main aerosol and cloud clusters, and assign them a suitably low CAD value (< 20). An example would be the features that populate at lower-right hand corner of the upper panels in Figure 1. These outlier features generally have physical and/or optical properties which are highly uncertain, and most of them are classified as cloud (refer to Figure 3). An example is the red stripes in the V3 VFM north of 41°N extending vertically from the surface to the base of the overlying cirrus deck at ~5 km. These low-confidence features are likely to be detected in error due to overestimate of the optical depth of the overlying cirrus, and so their measured optical properties are not representative of either clouds or aerosols. Using the new 5D scheme, most of these artifacts can be screened out by rejecting layers with |CAD| ≤ 20.

Figure 3 presents a comparison of histograms of CAD scores for the V2.01 data and V3 test data. The V3 curve shows a smoother distribution and, except for a small region near zero, has generally smaller occurrence frequencies for low CAD values (< 95), reflecting the better separation of clouds and aerosols in the 5D space versus the 3D space. Boundary-layer cloud clearing in the CALIPSO feature finding algorithm has also been improved in Version 3. This improvement contributes significantly to the improved CAD performance. The bump between -10 and 20 in the V3 test curve, which accounts for ~ 6% of the total features, includes the outlier features mentioned above and other features falling within the deep PDF overlap region. In contrast, these outliers were distributed over the entire CAD range (from -100 to 100) in the V2.01 data.

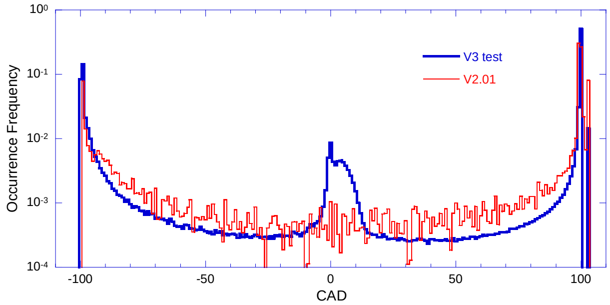

Figure 3. Distributions of occurrence frequency as a function of CAD derived from four months of the V2.01 production data using the 3D CAD algorithm and the V3 test data using the 5D algorithm. The bump between -10 and 20 in the V3 test curve, accounting for ~6% of the total features, corresponds mostly to outlier features whose physical/optical properties are not correctly derived.

As shown in Table 1, the V3 processing produced ~14% more aerosol layers when compared to the V2.01 data. We attribute ~10% of this to the improvement of the cloud clearing as mentioned above. Mixed cloud and aerosol layers identified by Version 2 algorithms were most often classified as cloud, because the optical properties of these mixed layers are dominated by the cloud contributions. After the improved clearing of the embedded cloud signals by the V3 feature finding algorithm, the surrounding aerosol layers can now be correctly classified. The remaining ~4% of the aerosol increase in the V3 test is attributable to the use of 5D CAD algorithm.

Table 1: CAD distributions for a four month test data

CAD	-100–0 aerosol	0–100 cloud	101	102	103
V3	36.2%	63.3%	none	none	1.5%
V2.01	21.9%	67.3%	2.2%	0.7%	8.0%

Figure 4 shows the differences in aerosol occurrence frequencies between the 5-km V3 test and V2.01 layer products as a function of latitude and altitude. The V3 processing produces more aerosols primarily in the boundary layer over the geographic areas where dense dust and/or dense smoke and polluted layers frequently occur. In addition, relatively dense maritime aerosols are more frequently observed, especially over the southern hemisphere open oceans. The 5D algorithm also improves the classification of dense aerosols, including dust layers over/near the source regions (0°N-50°N), and smoke layers (30°S-10°N).

As mentioned earlier, because the additional VDR test needed in the 3D algorithm is no longer necessary in the 5D algorithm, the special CAD score of 101 no longer exist. Improvements have also been made to the cloud sub-typing algorithm with a better identification of horizontally oriented ice (HOI) crystals which were mostly labeled with the special CAD score of 102 in the V2 data [3]. Therefore, the special CAD score of 102 is eliminated, and the number of features with the special score of 103, a mixture of HOI and other types of cloud, is reduced significantly.

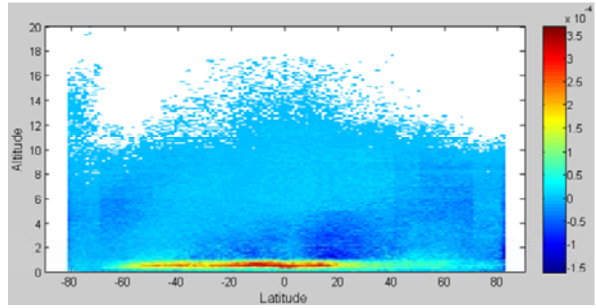

Figure 4. Vertical-latitudinal distribution of differential aerosol occurrence frequencies of the 5 km V3 test data and V2.01 production data (V3-V2.01). The vertical axis is the mid-layer altitude.

Even when using five dimensions, there may still be some unavoidable overlap between aerosol PDFs and cloud PDFs. One such example is at high latitudes where cirrus clouds can occur at low altitudes, as shown in the lower two most right-hand-side panels in Figure 1. If moderately dense dust is transported into these high latitude regions, it can still be misclassified as cloud, even though its occurrence frequency should be very low.

Continuous efforts will be made to improve the CAD algorithm for post Version 3 releases. Among the strategies currently being considered is the addition of a time dimension to the PDFs (i.e., monthly or seasonal PDFs). Assessing the horizontal correlations between the optical properties of neighboring layers should also improve our ability to distinguish between clouds and aerosols. Tests based on spatial correlations could be particularly useful in the boundary layer, where aerosol layers can extend up to several hundred kilometers horizontally, whereas clouds in this region typical exist on much smaller spatial scales.

REFERENCES

[1] Winker, D., W. Hunt, and M. McGill, 2007: Initial performance assessment of CALIOP, *Geophys. Res. Lett.*, **34**, doi:10.1029/2007GL030135.

[2] Liu, Z., M. Vaughan, D. Winker, et al., 2004: Use of probability distribution functions for discriminating between cloud and aerosol in lidar backscatter data, *J. Geophys. Res.*, **109**, doi:10.1029/2004JD004732.

[3] Liu, Z., M. Vaughan, D. Winker, et al., 2009: The CALIPSO Lidar Cloud and Aerosol Discrimination: Version 2 Algorithm and Initial Assessment of Performance, *J. Atmos. Oceanic Technol.*, **26**, pp.1198-1213.