CALIPSO Quality Statements Lidar Level 2 Vertical Feature Mask Product Version Releases: 3.01, 3.02

Introduction | Documentation and References | Standard and Expedited Data Set Definitions | CALIPSO Lidar Level 2 Data Products | Overview | Revision Summary: Version 3.02 and Version 3.01

Introduction

This document provides a high-level quality assessment of the Level 2 lidar vertical feature mask product, as described in section 2.7 of the CALIPSO Data Products Catalog (Version 3.2) (PDF). As such, it represents the minimum information needed by scientists and researchers for appropriate and successful use of these data products. We strongly suggest that all authors, researchers, and reviewers of research papers review this document for the latest status before publishing any scientific papers using these data products.

The purpose of these data quality summaries is to inform users of the accuracy of CALIOP data products as determined by the CALIPSO Science Team and Lidar Science Working Group (LSWG). This document is intended to briefly summarize key validation results; provide cautions in those areas where users might easily misinterpret the data; supply links to further information about the data products and the algorithms used to generate them; and offer information about planned algorithm revisions and data improvements.

The CALIPSO Level 2 Lidar Vertical Feature Mask (VFM) consists of a sequence of bit-mapped integers, with one 16-bit integer being recorded for each range resolution element in the Level 0 lidar data downlinked from the satellite. Decoding the bits in the individual integers yields information on feature type (e.g., cloud, aerosol, or clear air) and subtype (e.g., water cloud or ice cloud) at each location. Information about cloud thermodynamic phase, and the amount of horizontal averaging required for detection is also included, as are quality assessments for all classification decisions. No new parameters have been added for the version 3 VFM product. However, the interpretation assigned to the bits describing cloud ice-water phase has changed slightly from version 2. These changes are described in detail below.

Data Product Maturity

Because validation for different parameters can require different levels of effort, and because the uncertainties inherent in some retrievals can be substantially larger than in others, the maturity levels of the parameters reported in the different data products files are not uniform. Therefore, within this document, maturity levels are provided separately for each scientific data set (SDS) included with the data files. The data product maturity levels for the CALIPSO layer products are defined in the table below.

Maturity Level Definitions

Beta:	Early release products for users to gain familiarity with data formats and parameters. **Users are strongly cautioned against the indiscriminate use of these data products as the basis for research findings, journal publications, and/or presentations.**
Provisional:	Limited comparisons with independent sources have been made and obvious artifacts fixed.
Validated Stage 1:	Uncertainties are estimated from independent measurements at selected locations and times.
Validated Stage 2:	Uncertainties are estimated from more widely distributed independent measurements.
Validated Stage 3:	Uncertainties are estimated from independent measurements representing global conditions.
External:	Data are not CALIPSO measurements, but instead are either obtained from external sources (e.g., the Global Modeling and Assimilation Office (GMAO)) or fixed constants in the CALIPSO retrieval algorithm (e.g., the 532 nm calibration altitude).

Documentation and References

Algorithm Theoretical Basis Documents (ATBDs)

- PC-SCI-202.01 - Mission, Instrument, and Algorithms Overview (PDF)
- PC-SCI-202.02 - Feature Detection and Layer Properties Algorithms (PDF)
- PC-SCI-202.03 - Scene Classification Algorithms (PDF)

General References

- PC-SCI-503 : CALIPSO Data Products Catalog (Version 3.2) (PDF)
- Data analysis overview: Fully automated analysis of space-based lidar data: an overview of the CALIPSO retrieval algorithms and data products (PDF)
- CALIPSO algorithm papers published in a special issue of the Journal of Atmospheric and Oceanic Technology
- Peer-reviewed CALIPSO validation papers
- Additional peer-reviewed publications discussing scientific applications and studies using CALIPSO data
- Recent conference proceedings covering a broad range of CALIPSO-related science and data analysis topics
- CALIPSO Data Read Software

Standard and Expedited Data Set Definitions

Standard Data Sets:
Standard data processing begins immediately upon delivery of all required ancillary data sets. The ancillary data sets used in standard processing (e.g., GMAO meteorological data and data from the National Snow and Ice Data Center) must be spatially and temporally matched to the CALIPSO data acquisition times, and thus the time lag latency between data onboard acquisition and the start of standard processing can be on the order of several days.
The data in each data set are global, but are produced in files by half orbit, with the day portion of an orbit in one file and the night portion of the orbit in another.

Expedited Data Sets:
Expedited data are processed as soon as possible after following downlink from the satellite and delivery to LaRC. Latency between onboard acquisition and analysis expedited processing is typically on the order of 6 to 28 hours. Expedited processing uses the most recently current available set of ancillary data (e.g., GMAO meteorological profiles) and calibration coefficients available, which may lag the CALIPSO data acquisition time/date by several days.
Expedited data files contain at the most, 90 minutes of data. Therefore, each file may contain both day and night data.
NOTE: Users are strongly cautioned against using Expedited data products as the basis for research findings or journal publications. Standard data sets only should be used for these purposes.

The differences between expedited processing and standard processing are explained in more detail in "Adapting CALIPSO Climate Measurements for Near Real Time Analyses and Forecasting" (PDF).

CALIPSO Vertical Feature Mask

- Overview
- Spatial and Temporal Parameters
- Metadata Parameters

Overview

The CALIPSO Level 2 lidar vertical feature mask data product describes the vertical and horizontal distribution of cloud and aerosol layers observed by the CALIPSO lidar. Each range bin in the Lidar Level 0 data is characterized by a single 16-bit integer, with the various bits in the integer representing flags that describe some aspect of the data measured within the bin. Instructions on how to decode these integer data are given in the sections below. Also provided are quality summaries for each classification reported. The data are recorded in nominal increments of 15 consecutive laser pulses, which is nominally equivalent to a distance of 5-km along the laser ground-track.

Spatial and Temporal Parameters

Latitude (ValStage1)
: Latitude, in degrees, of the laser footprint. One value is reported at the temporal midpoint of each 15 shot data segment.

Longitude (ValStage1)
: Longitude, in degrees, of the laser footprint. One value is reported at the temporal midpoint of each 15 shot data segment.

Profile Time (TAI) (external)
: Time expressed in International Atomic Time (TAI). Units are in seconds, starting from January 1, 1993. One value is reported at the temporal midpoint of each 15 shot data segment.

Profile Time (UTC) (external)

Time expressed in Coordinated Universal Time (UTC), and formatted as 'yymmdd.ffffffff', where 'yy' represents the last two digits of year, 'mm' and 'dd' represent month and day, respectively, and 'ffffffff' is the fractional part of the day. One value is reported at the temporal midpoint of each 15 shot data segment.

Day-Night Flag (external)

Indicates the lighting conditions at an altitude of ~24 km above mean sea level: 0 = day, 1 = night.

Land-Water Flag (external)

This is a 30 arc second resolution land/water mask provided by the SDP toolkit. The data is stored as an 8-bit integer. The values indicate the surface type at the lidar footprint, and are interpreted as follows:

- 0 = shallow ocean
- 1 = land
- 2 = coastlines
- 3 = shallow inland water
- 4 = intermittent water
- 5 = deep inland water
- 6 = continental ocean
- 7 = deep ocean

Please see section 4.5 in PC-SCI-503: CALIPSO Data Products Catalog (Version 3.2) (PDF) for more information.

Feature Classification Flags (ValStage1)

For each layer detected in the CALIPSO backscatter data, we derive a set of feature classification flags that report (a) feature type (e.g., cloud vs. aerosol vs. stratospheric layer); (b) feature subtype; (c) layer ice-water phase (clouds only); and (d) the amount of horizontal averaging required for layer detection. Quality assessments are also provided for all layer classifications. The complete set of flags is stored as a single 16-bit integer. The following table is reproduced from the CALIPSO Data Products Catalog.

Reproduced from PC-SCI-503: CALIPSO Data Products Catalog (Version 3.2) (PDF)

Table 44: Feature Classification Flag Definition

Bits	Field Description	Bit Interpretation
1-3	Feature Type	0 = invalid (bad or missing data) 1 = "clear air" 2 = cloud 3 = aerosol 4 = stratospheric feature 5 = surface 6 = subsurface 7 = no signal (totally attenuated)
4-5	Feature Type QA	0 = none 1 = low 2 = medium 3 = high
6-7	Ice/Water Phase	0 = unknown / not determined 1 = randomly oriented ice 2 = water 3 = horizontally oriented ice
8-9	Ice/Water Phase QA	0 = none 1 = low 2 = medium 3 = high
10-12	Feature Sub-type	
	If feature type = aerosol, bits 10-12 will specify the aerosol type	0 = not determined 1 = clean marine 2 = dust 3 = polluted continental 4 = clean continental 5 = polluted dust 6 = smoke 7 = other
	If feature type = cloud, bits 10-12 will	0 = low overcast, transparent

	specify the cloud type.	1 = low overcast, opaque 2 = transition stratocumulus 3 = low, broken cumulus 4 = altocumulus (transparent) 5 = altostratus (opaque) 6 = cirrus (transparent) 7 = deep convective (opaque)
	If feature type = Polar Stratospheric Cloud, bits 10-12 will specify PSC classification.	0 = not determined 1 = non-depolarizing PSC 2 = depolarizing PSC 3 = non-depolarizing aerosol 4 = depolarizing aerosol 5 = spare 6 = spare 7 = other
13	Cloud / Aerosol /PSC Type QA	0 = not confident 1 = confident
14-16	Horizontal averaging required for detection (provides a coarse measure of feature backscatter intensity)	0 = not applicable 1 = 1/3 km 2 = 1 km 3 = 5 km 4 = 20 km 5 = 80 km

User notes for the feature classification flags.

- Bits 1-3, Feature Type

  Invalid (bad or missing data)

  Invalid features are defined as those layers for which the 532 nm integrated attenuated backscatter, γ'_{532}, is less than zero. In version 3 of the CALIPSO Level 2 data products, invalid layers will only occur in those very rare cases when (a) a layer is detected at one of the coarser averaging resolutions (i.e., 20 km or 80 km) and (b) above this layer there are also overlying layers detected at finer resolutions (typically 5 km) that have large and widely varying optical depths. The non-uniform attenuation corrections applied to the backscatter signal within the lower layer can result in an unfavorable redistribution of signal magnitudes; e.g., large corrections applied to profiles with a substantial fraction of negative attenuated backscatter values, and smaller corrections applied to profiles with all positive values. In especially perverse cases, γ'_{532} for the rescaled and reaveraged data can be negative, even though prior to rescaling γ'_{532} was comfortably above zero.

  "Clear Air"

  Regions are where the signal has not been totally attenuated by overlying layers and in which no feature is detected are classified as being clear air. Because the determination of clear air depends not only on the cloud/aerosol content of the atmosphere, but also on the minimum detectable backscatter of the CALIOP layer detection scheme, regions containing especially weak layers can, on occasion, be misclassified as clear air. Detection threshold issues are addressed in detail in the CALIPSO layer detection ATBD (PDF).

  Cloud / Aerosol

  The V3 cloud aerosol discrimination (CAD) algorithm uses the altitude-and-latitude-dependent feature integrated color ratio, χ', the layer-integrated volume depolarization ratio, δ_v, and the feature mean attenuated backscatter coefficient, $\langle\beta'_{532}\rangle$, to compute the CAD score. These parameters depend on the quality of the 532 nm and 1064 nm channel calibrations. Significant errors in the calibration of either channel may result in the misclassification of a particular feature.

  The probability distribution functions (PDFs) of χ' vs. δ_v vs. $\langle\beta'_{532}\rangle$ for clouds and aerosols that are used by the V3 CAD algorithm were developed based on a four month test data set. These PDFs are binned by altitude (1 km increments between 0 km and 20 km) and latitude (between 90°S and 90°N in 10° increments). Despite the use of these updated 5D PDFs, which significantly enhance overall performance, the V3 CAD algorithm (PDF) may continue to have some difficulty correctly classifying moderately dense dust and smoke layers presented in high latitudes and/or high altitudes as aerosol. Users should also be aware that clouds embedded within optically dense aerosols may be identified by the feature finder algorithm as a single layer. While this happens less in version 3 than in the earlier releases, these misidentified heterogeneous features will likely be classified as clouds.

  Stratospheric

  All layers with base altitudes lying above the tropopause are classified as stratospheric features. The correctness of this classification thus depends critically on the accuracy of the tropopause heights obtained from the GMAO GEOS-5 meteorological data set. Because base height is the sole criterion used in determining stratospheric features, there are occasions when high altitude cirrus clouds and/or overshooting tops from deep convection are incorrectly classified as being stratospheric layers. In uncertain situations, users are advised to query the feature type QA bits (see below) and refer to the measured optical properties reported in the aerosol layer products.

In the polar regions where polar stratospheric clouds (PSCs) are observed, there may be times when stratospheric layers are misclassified as cloud. This typically happens when the base of a PSC drops below the GMAO-reported tropopause or when a PSC is vertically adjacent to a cloud system in the troposphere.

Surface/Subsurface

All regions lying directly below the lidar surface elevation (i.e., surface echoes detected by the CALIOP layer detection algorithm) are classified as subsurface.

No signal

The backscatter signal in those regions lying directly beneath opaque clouds, aerosols, and/or stratospheric layers is totally attenuated by the overlying features, and thus these regions are classified as having "no signal".

- Bits 4-5, Feature Type QA

Invalid

Not applicable. Always 0.

"Clear Air"

Not applicable. Always 0.

Cloud / Aerosol

For cloud and aerosol layers, feature type QA is directly related to the cloud-aerosol discrimination (CAD) score, as follows:

high confidence	=	|CAD score| >= 70
medium confidence	=	50 <= |CAD score| < 70
low confidence	=	20 <= |CAD score| < 50
no confidence	=	|CAD score| < 20

Stratospheric

Let H_t indicate the local tropopause altitude, in kilometers; then

high confidence	=	layer base >= (H_t + 2.5 km)
medium confidence	=	(H_t + 1.0) km <= layer base < (H_t + 2.5 km)
low confidence	=	layer base < (H_t + 1.0 km)

Surface / Subsurface

Always high.

No signal

Not applicable; always 0.

- Bits 6-7, Ice/Water Phase

Cloud phase is determined from relations between depolarization ratio, backscatter intensity, temperature, and attenuated backscatter color ratio. The cloud phase algorithm used in Version 3 is new and completely different from that used in Version 2. The Version 3 algorithm classifies cloud layers as water, randomly-oriented ice, or horizontally-oriented ice. In those cases where the classification is ambiguous, the phase is reported as "unknown/not determined". The classification of "mixed phase cloud" has been eliminated; the version 3 algorithm does not attempt to determine if more than one phase is present within a layer. The version 3 algorithm distinguishes between two separate classes of ice clouds: those dominated by randomly oriented particles, and those containing a substantial fraction of horizontally oriented crystals.

- Bits 8-9, Ice/Water Phase Quality Assessment Changes in the interpretation of the ice-water phase QA flags between version 2 and version 3 are described in the table below.

Value	Version 2 Interpretation	Version 2 Interpretation
0	no confidence	no/low confidence
1	low confidence	phase based on temperature only
2	medium confidence	medium confidence
3	high confidence	high confidence

- Bits 10-12, Feature Sub-type

  Aerosols

  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 volume depolarization is directly related to the hydration state of the aerosol. The backscatter and volume depolarization are not sufficient to fully constrain the model selection, however. Therefore, the selection algorithm uses surface type to aid in the type identification. The input parameters - the magnitude of attenuated backscatter, altitude, location, surface type, depolarization ratio, and mean attenuated backscatter coefficient measurements - are used to identify the type following one of several pathways. The volume 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. The integrated attenuated backscatter (γ') 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 the ocean. For aerosols in polar regions, the algorithm takes into consideration the high aerosol loading events caused by arctic haze. Once the type is identified, the aerosol lidar ratio, S_a is chosen from a lookup table that currently consists of six pairs of 532 nm and 1064 nm values.

  In summary, the algorithm classifies aerosol layers that have volume depolarization ratio (δ_v) greater than 0.2 as desert dust and $0.075 < \delta_v < 0.2$ as polluted dust. Note that polluted dust could be a component of urban pollution, i.e., it is not confined to desert regions but is any type of aerosol composed of some dust-like particles. Of the non-depolarizing aerosols, layers lofted above 1 km are assumed to be smoke, and layers less than 1 km above the surface are either clean continental if the layer IAB is small or polluted continental if the layer IAB is large.

  An assessment of the CALIOP aerosol subtyping scheme can be found in Mielonen et al., 2010.

  Clouds

  This flag classifies the cloud layers detected by CALIPSO into the standard meteorological cloud types defined by the ISCCP). The classification decisions are based on CALIOP measurements of cloud top altitude, assessments of cloud ice/water phase, cloud opacity, and cloud fraction within an 80 km segment.

  Stratospheric

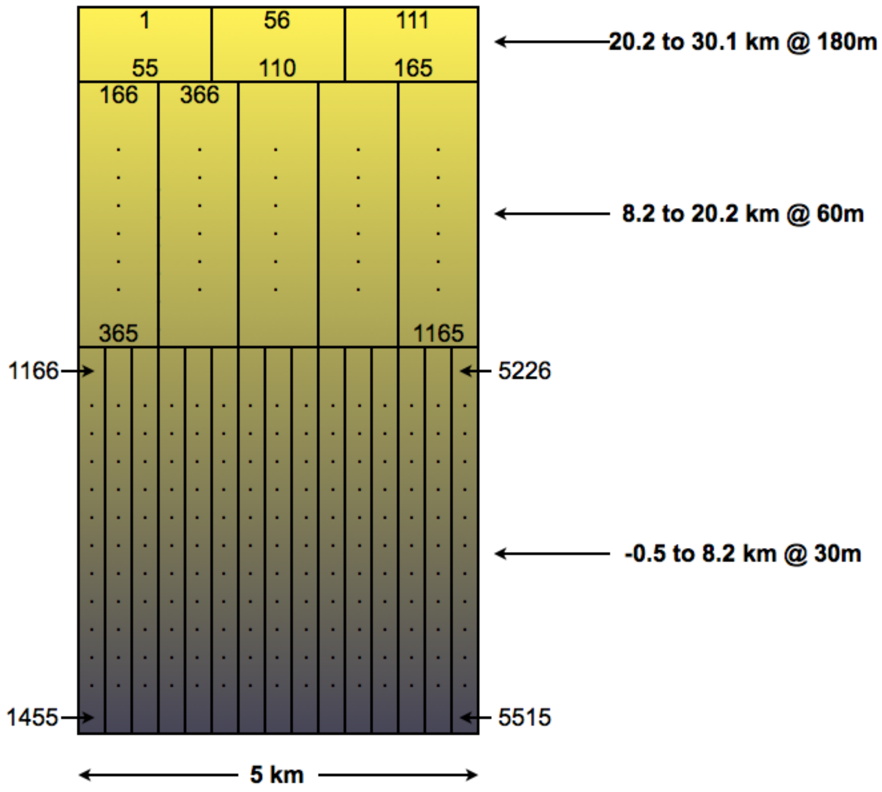
  N/A

- Bit 13, Cloud / Aerosol / PSC Subtype Quality Assessment

- Bits 14-16, Horizontal Averaging

  Specifies the amount of horizontal averaging required for a feature to be detected. For all data versions up to and including 3.01 release, the values decoded from the bits in this field will be either 1/3 km, 1 km, 5 km, 20 km, or 80 km.

Layout of the Feature Classification Flag data block



The Feature_Classification_Flag values are stored as a sequence of 5515 element arrays (i.e., as an N x 5515 matrix, where N is the number of separate records in the file). Each array represents a 5 km "chunk" of data, and each array element contains the feature classification information for a single range resolution element in the Level 0 lidar data downlinked from the satellite. As shown in the table below, the vertical and horizontal resolution of the CALIPSO data varies as a function of altitude above mean sea level (see Hunt et al., 2009). The image above provides a pictorial representation of the mapping of the one-dimensional array of Feature_Classification_Flag values into a two-dimensional array of range-resolved lidar data samples. The numbers in each block of the image indicate the 1-D array indices associated with each spatial averaging regime in the 2-D lidar backscatter data. Only the starting and ending indices are shown.

Example code for transforming a 1-D array of feature classification flags into a rectangular, altitude-registered matrix is available for both IDL and MatLab.

- IDL example code: extractVFM.pro (.txt)
- MatLab example code: vfm_row2block.zip (.zip)

NOTE: When saving the IDL code, remove the ".txt" extension from each of the files before running them.

Reproduced from PC-SCI-503 : CALIPSO Data Products Catalog (Version 3.2) (PDF) section 2.7

Profile Spatial Resolution

Altitude Region		Vertical Resolution (meters)	Horizontal Resolution (meters)	Profiles per 5 km	Samples per Profile
Base (km)	Top (km)				
-0.5	8.2	30	333	15	290
8.2	20.2	60	1000	5	200
20.2	30.1	180	1667	3	55
Total					545

Metadata Parameters

Product ID

an 80-byte (max) character string specifying the data product name. For all CALIPSO Level 2 lidar data products, the value of this string will be "L2_Lidar".

Date Time at Granule Start

a 27-byte character string that reports the date and time at the start of the file orbit segment (i.e., granule). The format is yyyy-mm-ddThh:mm:ss.ffffffZ.

Date Time at Granule End

a 27-byte character string that reports the date and time at the end of the file orbit segment (i.e., granule). The format is yyyy-mm-ddThh:mm:ss.ffffffZ.

Date Time at Granule Production

This is a 27-byte character string that defines the date at granule production. The format is yyyy-mm-ddThh:mm:ss.ffffffZ.

Lidar L1 Production Date Time

For each CALIOP Lidar Level 2 data product, the Lidar Level 1 Production Date Time field reports the file creation time and date for the CALIOP Level 1 lidar data file that provided the source data used in the Level 2 analyses.

Number of Good Profiles

This is a 32-bit integer specifying the number of good attenuated backscatter profiles contained in the granule.

Number of Bad Profiles

This is a 32-bit integer specifying the number of bad attenuated backscatter profiles contained in the granule.

Initial Subsatellite Latitude

This field reports the first subsatellite latitude of the granule.

Initial Subsatellite Longitude

This field reports the first subsatellite longitude of the granule.

Final Subsatellite Latitude

This field reports the last subsatellite latitude of the granule.

Final Subsatellite Longitude

This field reports the last subsatellite longitude of the granule.

Orbit Number at Granule Start

This field reports the orbit number at the granule start time.

Orbit Number at Granule End

This field reports the orbit number at the granule stop time.

Orbit Number Change Time

This field reports the time at which the orbit number changes in the granule.

Path Number at Granule Start

This field reports the path number at the granule start time.

Path Number at Granule End

This field reports the path number at the granule stop time.

Path Number Change Time

This field reports the time at which the path number changes in the granule.

Lidar Data Altitude

This field defines the lidar data altitudes (583 range bins) to which lidar Level 1 profile products are registered.

GEOS Version

This is a 64-bit character that reports the version of the GEOS data product provided by the GMAO.

Classifier Coefficients Version Number

Version number of the classifier coefficients file that stores the five-dimensional probability distribution functions used by the cloud-aerosol discrimination (CAD) algorithm

Classifier Coefficients Version Date

Creation date of the classifier coefficients file that stores the five-dimensional probability distribution functions used by the cloud-aerosol discrimination (CAD) algorithm

Production Script

Provides the configuration information and command sequences that were executed during the processing of the CALIOP Lidar Level 2 data products. Documentation for many of the control constants found within this field is contained in the CALIPSO Lidar Level 2 Algorithm Theoretical Basis Documents

---

Data Release Versions

Lidar Level 2 Vertical Feature Mask (VFM) Information *Half orbit (Day) geolocated data radiances*			
Release Date	**Version**	**Data Date Range**	**Maturity Level**
December 2011	3.02	November 1, 2011 to present	Validated Stage 1
May 2010	3.01	June 13, 2006 to February 16, 2009; March 17, 2009 to October 31, 2011	Validated Stage 1

Data Quality Statement for the release of the CALIPSO Lidar Level 2 Vertical Feature Mask (VFM) Products Version 3.02, December 2011

The CALIPSO Team is releasing Version 3.02 which represents a transition of the Lidar, IIR, and WFC processing and browse code to a new cluster computing system. No algorithm changes were introduced and very minor changes were observed between V 3.01 and V 3.02 as a result of the compiler and computer architecture differences. Version 3.02 is being released in a forward processing mode beginning November 1, 2011.

Data Quality Statement for the release of the CALIPSO Lidar Level 2 Vertical Feature Mask (VFM) Products Version 3.01, May 2010

Version 3.01 of the Lidar Level 2 data products is a significant improvement over previous versions. Major code and algorithm improvements include

- the elimination of a vicious, vile, and pernicious bug in the cloud clearing code that caused a substantial overestimate of low cloud fraction in earlier data releases (details given in Vaughan et al., 2010 (PDF));
- enhancements to the cloud-aerosol discrimination algorithm that increase the number of diagnostic parameters used to make classification decisions (details given in Liu et al., 2010 (PDF));
- improved daytime calibration procedures, resulting in more accurate estimates of layer spatial and optical properties (details given in Powell et al., 2010 (PDF)); and
- an entirely new algorithm for assessing cloud thermodynamic phase (details given in Hu et al., 2009).

The sections below highlight important changes to the layer detection, scene classification, and extinction algorithms that have implications for the overall quality of the Lidar Level 2 data products.

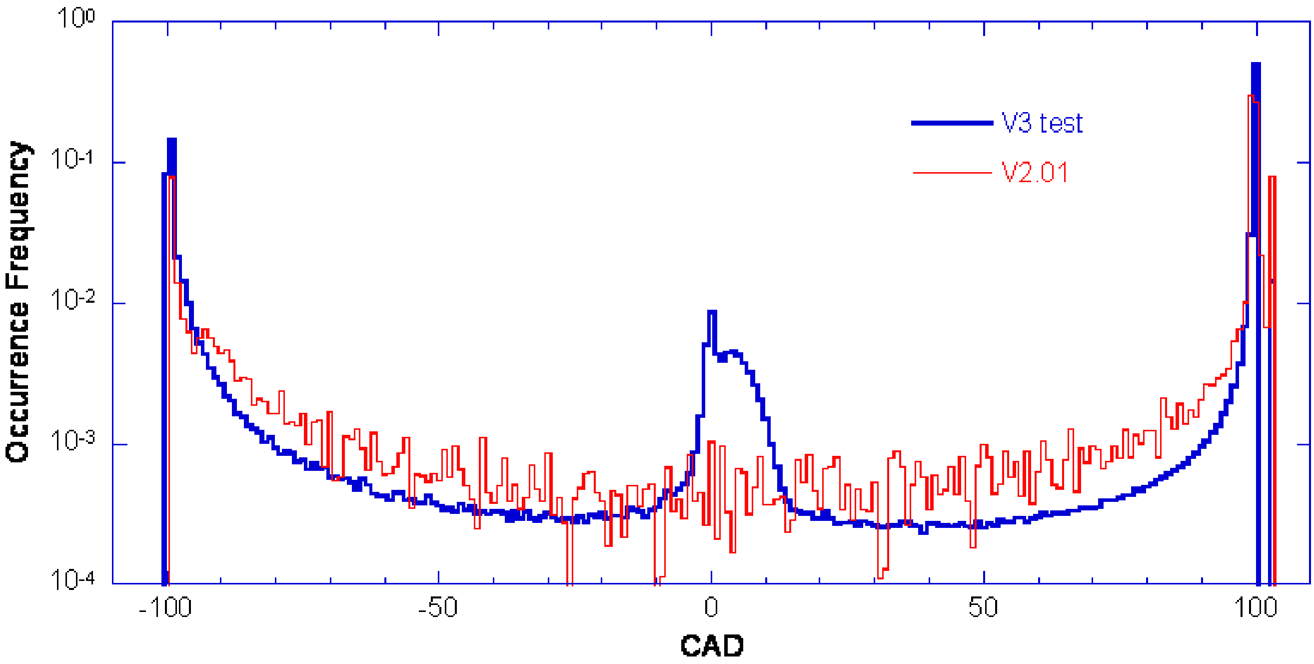
Layer Detection

As in previous versions, the layer boundaries reported in the Lidar Level 2 Cloud and Aerosol Layer Products appear to be quite accurate. Some false positives are still found beneath optically thick layers; these, however, can generally be identified by their very low CAD scores (e.g., |CAD score| <= 20). In opaque layers, the lowest altitude where signal is reliably observed is reported as the base. In actuality, this reported base may lie well above the true base. Opaque layers are denoted by an opacity flag. In this release, the layers which are reported represent a choice in favor of high reliability over maximum sensitivity. Weakly scattering layers sometimes will go unreported, in the interest of minimizing the number of false positives.

Cloud-Aerosol Discrimination

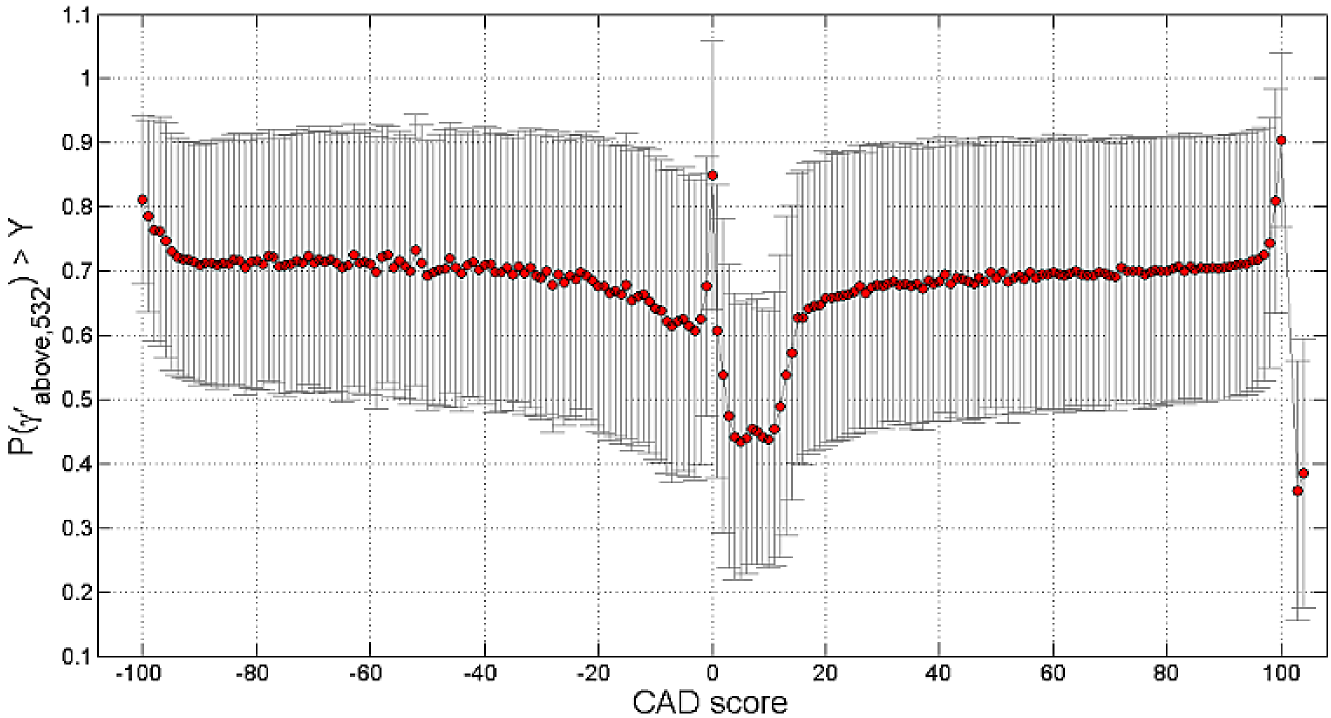
The figure below compares the distributions of CAD scores derived from four months of version 3 test data to the corresponding version 2.01 data. The V3 curve shows a smoother distribution and generally has fewer low CAD values (i.e., values less than ~|95|), reflecting the better separation of clouds and aerosols in the 5D space than in the 3D space. One notable exception to this observation is the bump between -10 and 20 in the V3 test curve, which accounts for ~ 6% of the total features. The CAD scores in this region identify both outlier features whose optical/physical properties are not correctly measured or derived, and the those features whose attributes fall within the overlap region between the cloud and aerosol PDFs. These outliers are populated over the entire CAD span in the V2 release.

Figure 6: Histograms of CAD scores for Version 2 (red) and Version 3 (blue)



This figure in below presents the relationship between the CAD score and the IAB QA (= 1- Column IAB cumulative probability). The IAB QA is a measure of the backscatter overlying a cloud or an aerosol layer. A value of 1 corresponds to a situation where the atmosphere above the cloud or aerosol layer under consideration is clear. Values smaller than 1 indicate the presence of overlying features (aerosol and/or cloud); the smaller of the IAB QA value, the more of the overlying features. The IAB QA is highest for high magnitude CAD scores and slopes down gradually for small CAD score magnitudes. This relationship reflects the fact that the presence of overlying features tends to add difficulty to the cloud and aerosol classification and therefore reduce the confidence of a classification made. The dip between -10 and 20 corresponds to the outlier features, indicating that these outliers are mostly likely overlaid by some relatively dense features. The cloud layers with special CAD scores (103 and 104) have the smallest IAB QA values. The relatively big value at CAD = 0 corresponds to the features having zero CAD values at high altitudes where the probability of the presence of overlying features is low. At high altitudes the separation of clouds and aerosols is not as good as at low altitudes because of the presence of cirrus clouds.

Figure 7: Relation between CAD score and Layer IAB QA Factor



Overall, because of the better separation between clouds and aerosols in the 5D space, the 5D algorithm significantly improves the reliability of the CAD scores. The improvements include:

1. Dense aerosol layers (primarily very dense dust and smoke over and close to the source regions), which are sometimes labeled as cloud in the V2 release, are be correctly identified as aerosol, largely because of the addition of the integrated volume depolarization ratio. In addition, in the open oceans, dense aerosols that were previously classified as clouds are now frequently observed in the marine boundary layer. Improvements are also seen for these maritime aerosols. Note, however, dense dust/smoke layers found at single-shot (0.333 km) resolution will be classified as cloud by default. This issue will be revisited for post-V3 releases.

2. Because the V2 CAD algorithm used a latitude-independent set of 3D PDFs, a class of optically thin clouds encountered in the polar regions that can extend from the surface to several kilometers were sometimes misclassified as aerosols. In version 3, these features are now correctly classified as cloud.

3. Correct classification of heterogeneous layers is always difficult. An example of a heterogeneous layer would be an aerosol layer that is vertically adjacent to a cloud or contains an embedded cloud, but which is nonetheless detected by the feature finder as a single entity in the V2 release. By convention, heterogeneous layers should be classified as clouds. The version 3 feature finding algorithm has also been improved greatly, and can now much better separate the embedded or adjacent single-shot cloud layers from the surrounding aerosol. This improvement in layer detection contributes significantly to the improvement of the CAD performance.

4. Some so-called features identified by the layer detection scheme are not legitimate layers, but instead are artifacts due to the noise in the signal, multiple scattering effects, or to artificial signal enhancements caused by non-ideal detector transient response or an over estimate of the attenuation due to overlying layers. These erroneous "pseudo-features" are neither cloud nor aerosol and are distributed outside of the cloud and aerosol clusters in the PDF space. The V3 CAD algorithm can better identify these outlier features by assigning a small CAD score (the bump between -10 and 20 in the V3 CAD histogram) and classify most of them as cloud by convention. A CAD threshold of 20 can effectively filter these outliers out.

Some misclassification may still occur with the 5D algorithm. For example, dust aerosols can be transported long distance to the Arctic. When moderately dense dust layers are occasionally transported to high latitudes, where cirrus clouds can present even in the low altitudes, they may be misclassified. This is also the case for moderately dense smoke aerosols occasionally transported to the high latitudes. Smoke can be mixed with ice particles during the long range transport, which makes the smoke identification even more difficult. When moderately dense dust and smoke are transported vertically to high altitudes even at low latitudes, they may be misclassified, because of the presence of cirrus clouds. Volcanic aerosol that is newly injected into the high altitudes may have a large cross-polarized backscatter signal and thus may be misclassified as cloud.

Aerosol Type Identification

The main objective of the aerosol subtyping scheme is to estimate the appropriate value of the aerosol extinction-to-backscatter ratio (S_a) to within 30% of the true value. S_a is an important parameter used in the determination of the aerosol extinction and subsequently the optical depth from CALIOP backscatter measurements. S_a is an intensive aerosol property, i.e., a property that does not depend on the number density of the aerosol but rather on such physical and chemical properties as size distribution, shape and composition. These properties depend primarily on the source of the aerosol and such factors as mixing, transport, and in the case of hygroscopic aerosols, hydration.

The extinction products are produced by first identifying an aerosol type and then using the appropriate values of S_a and the multiple scattering factor, $\eta(z)$. Note that multiple scattering corrections have not yet been implemented for the current data release, so that $\eta(z) = 1$ for all aerosol types. The accuracy of the S_a value used in the lidar inversions depends on the correct identification of the type of aerosol. In turn, the accuracy of the subsequent optical depth estimate depends on the accuracy of S_a.

The underlying paradigm of the type classification is that a 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 aerosol subtype product is generated downstream of the cloud-aerosol discrimination (CAD) scheme and, therefore, depends on the cloud-aerosol classification scheme in a very fundamental way. If a cloud feature is misclassified as aerosol, the aerosol subtype algorithm will identify this 'aerosol' as one of the aerosol subtypes. The user must exercise caution where the aerosol subtype looks suspicious or unreasonable. Such situations can occur with some frequency in the southern oceans and the polar regions.

Cloud Ice/Water Phase Discrimination

The cloud phase algorithm used in Version 2 has been replaced with a new, completely different algorithm. The Version 3 algorithm classifies detected cloud layers as water, randomly-oriented ice (ROI), or horizontally-oriented ice (HOI) based on relations between depolarization, backscatter, and color ratio (Hu et al. 2009). These classifications have not yet been rigorously validated, which is difficult, but many of the obvious artifacts found in the Version 2 data have been eliminated.

The version 2 algorithm included a rudimentary ability to identify a specific subset of high confidence instances of HOI. These clouds were classified as ice clouds, and flagged with a 'special CAD score' of 102, indicating that they had been further classified as HOI. The new version 3 algorithm implements a much more sophisticated scheme for recognizing HOI that correctly identifies many more instances of these sorts of ice clouds. The special CAD score of 102 is no longer used to identify these layers. Instead, the "ice cloud" and "mixed phase cloud" classifications have been eliminated, and replaced as shown in the table below.

Value	Version 2 Interpretation	Version 3 Interpretation
0	unknown/not determined	unknown/not determined
1	ice	randomly oriented ice (ROI)
2	water	water
3	mixed phase	horizontally oriented ice (HOI)

The Ice/water Phase QA flags have also been redefined slightly for Version 3, as follows:

Value	Version 2 Interpretation	Version 3 Interpretation
0	no confidence	no/low confidence
1	low confidence	phase based on temperature only
2	medium confidence	medium confidence
3	high confidence	high confidence

A confidence flag of QA=1 indicates the phase classification is based on temperature. Initial classification tests are based on layer depolarization, layer-integrated backscatter, and layer-average color ratio. Layers classified as water with temperature less than -40 C are forced to ROI and given a confidence flag of QA=1. Layers classified as ROI or HOI with temperature greater than 0 C are forced to water and also given a confidence flag of QA=1. Clouds for which the phase is 'unknown/not determined' are assigned a confidence value of 0 (no/low confidence).

Layers classified as HOI based on anomalously high backscatter and low depolarization are assigned QA=3. These layers are rare after the CALIOP viewing angle was changed to 3° in November 2007. The Version 3 algorithm computes the spatial correlation of depolarization and integrated backscatter and uses this as an additional test of cloud phase. Layers classified as HOI using this test are assigned QA=2. The majority of the layers classified as HOI result from this test and typically have higher backscatter than ROI but similar depolarization. These layers are common even at a viewing angle of 3°. We interpret this as clouds with significant perpendicular backscatter from ROI but containing enough HOI to produce enhanced backscatter. These layers tend to be found at much colder temperatures than the high confidence HOI (see Hu et al. 2009).

Cloud and Aerosol Optical Depths

The reliability of cloud and aerosol optical depths reported in the version 3 data products is considerably improved over the version 2 release. Whereas the version 2 optical depths were designated as a beta quality product, and not yet suitable for use in scientific publications, the maturity level of the version 3 optical depths has been upgraded to provisional. Several algorithm improvements and bugs fixes factored into the decision to upgrade the maturity level. Among these were the addition of the aerosol layer base extension algorithm, which greatly improves the estimates of PBL aerosols, and significant improvements to the code responsible for rescaling the attenuated backscatter coefficients in lower layers to compensate for the beam attenuation the occurs when traversing transparent upper layers.

PLEASE NOTE: Users of the CALIOP optical depths should read and thoroughly understand the information provided in the Profile Products Data Quality Summary. This summary contains an expanded description of the extinction retrieval process from which the layer optical depths are derived, and provides essential guidance in the appropriate use of all CALIOP extinction-related data products. Validation and improvements to the profile products QA are ongoing efforts, and additional data quality information will be included with future releases.