CALIPSO Quality Statements: Lidar Level 2 Vertical Feature Mask Version Releases: 2.01, 2.02

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 2.4) (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 primary new parameters included in the version 2.0 release of the Vertical Feature Mask (VFM) product are aerosol type and cloud ice/water phase).

Additional Documentation and References

Algorithm Theoretical Basis Documents (ATBDs)

- PC-SCI-202.02 - Feature Detection and Layer Properties Algorithms (PDF)
- PC-SCI-202.03 - Scene Classification (PDF)

General References

- PC-SCI-503 : CALIPSO Data Products Catalog (Version 2.4) (PDF)
- Data analysis overview: Fully automated analysis of space-based lidar data: an overview of the CALIPSO retrieval algorithms and data products (PDF)
- Additional publications (journal articles and conference proceedings about CALIPSO science, algorithms, and data processing)
- CALIPSO Data Read Software

CALIPSO Lidar Level 2 Vertical Feature Mask

This data product describes the vertical and horizontal distribution of cloud and aerosol layers observed by the CALIPSO lidar. Cloud and aerosol discrimination for detected features is reported as a single value, the CAD_Score, which can be found in the Lidar Level 2 Cloud and Aerosol Layer data products. In this data product clouds and aerosols are distinguished by the "feature type" bits, and the CAD_Score is interpreted in the following fashion:

If CAD_Score > 0, feature is a cloud.
If CAD_Score < 0, feature is a aerosol.
Use of the CAD_Score to produce the feature typing QA bits, can be found below.

Latitude
: Latitude, in degrees, of the laser footprint. One value is reported at the temporal midpoint of a 15 shot average for each 5 km chunk of the Feature_Classification_Flag data.

Longitude
: Longitude, in degrees, of the laser footprint. One value is reported at the temporal midpoint of a 15 shot average for each 5 km chunk of the Feature_Classification_Flag data.

Profile Time (TAI)
: 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 a 15 shot average for each 5km chunk of the Feature_Classification_Flag data.

Profile Time (UTC)

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 a 15 shot average for each 5km chunk of the Feature_Classification_Flag data.

Day-Night Flag

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

Land-Water Flag

This is a 30 arc second resolution land/water mask provided by the SDP toolkit. It is an 8-bit integer indicating the surface type at the lidar footprint:

- 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 2.4) (PDF) for more information.

Feature Classification Flags [beta]

For each layer, we report a set of feature classification flags that provide assessments of (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. 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 2.4) (PDF)

Table 45: 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 = ice 2 = water 3 = mixed phase
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 specify the cloud type.	0 = low overcast, transparent 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

List of the data quality summaries and user notes for the feature classification flags.

- Bits 1-3, Feature Type

  Invalid (bad or missing data)

  Features are labeled as invalid when the feature integrated attenuated backscatter, γ'_{532}, is outside the limit, $0 < \gamma'_{532}$. If the limit, $\gamma'_{532} > 1$, is exceeded then the feature will be classified as a cloud with a confidence value of 0, (Bits 4-5). These types of features may be found below some strongly attenuating features. This occurs because the feature finding algorithm has overestimated the attenuation of the overlying feature due to the low SNR of the signal below the previous feature that was found.

  "Clear Air"

  This indicates areas where no features were found. This does not guarantee that there are no features present. Aerosol or cloud layers may be present but they are below the detection threshold.

  Cloud / Aerosol

  When clouds are reported the value of the feature subtype flag (Bits 10-12) will all be zero. For this data release, a value of zero indicates "not reported" rather than that all clouds have been classified as low overcast transparent, as is stated in the data products catalog.

  The cloud aerosol discrimination (CAD) algorithm uses the feature integrated color ratio, χ', and the feature mean attenuated backscatter coefficient, $<\beta'_{532}>$, 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 current probability distribution functions of χ' vs. $<\beta'_{532}>$ for clouds and aerosols that are used by the CAD algorithm were developed based on expert manual classification of all layers detected during one full day of data acquired by CALIOP during August 2006. From these results, a single set of cloud and aerosol PDFs was constructed. This set of PDFs is applied globally for all seasons and at all latitudes. Despite the use of these updated PDFs, which significantly enhance overall performance, the current algorithm (v 2.01) (PDF) continues to have some difficulty correctly classifying optically dense biomass burning layers as aerosol. Users should also be aware that clouds embedded within optically dense aerosols will likely be identified by the feature finder algorithm as one feature and, consequently, these features will likely be classified as clouds.

  Stratospheric

  The current classification algorithm calls any feature with its base above the tropopause as a stratospheric feature. While there exist bonafide observations of relatively dense stratospheric aerosol layers (plume from the Soufriere Hills, Montserrat eruption on May 20, 2006), some high altitude cirrus clouds and deep convective clouds have been misclassified as stratospheric features. We believe that the misclassification is due to an incorrect location of the tropopause that is provided by the GMAO GEOS-5 meteorological data set. Please query the feature type QA bits when using this data product.

In the Antarctic region where polar stratospheric clouds (PSCs) have been observed, there may be times when a vertical strip of the PSC may be classified as cloud. In many situations this happens because the base of the PSC drops below the GMAO- reported tropopause or because the PSC is vertically adjacent to a cloud system in the troposphere. The current version of the feature finding algorithm reports only a single feature even if its vertical extent spans the tropopause.

Surface/Subsurface

Please see the comments on the Lidar_Surface_Elevation Lidar Level 2 Cloud and Aerosol Layer.

No signal

This is the value given to any region below a fully attenuating atmospheric feature.

- Bits 4-5, Feature Type QA

  Invalid

  Not applicable. Always 0.

  "Clear Air"

  Not applicable. Always 0.

  Cloud / Aerosol

  For the feature, if abs(CAD_Score) > 0.70, then the confidence value is high.
  If 0.5 < abs(CAD_Score) < 0.7, then the confidence value is medium.
  If 0.2 < abs(CAD_Score) < 0.5, then the confidence value is low.
  If abs(CAD_Score) < 0.2, then the confidence value is none.

  Stratospheric

  For this description use H_t to indicate the tropopause altitude.
  For the feature, if Base_Altitude > H_t+ 2.5km, then the confidence value is high.
  If H_t + 2.5km < Base_Altitude < H_t + 1.0km, then the confidence value is medium.
  If H_t + 1.0km < Base_Altitude < H_t, then the confidence value is low.

  Surface / Subsurface

  Always high.

  No signal

  Not applicable. Always 0.

- Bits 6-7, Ice/Water Phase

  Cloud phase is determined using a depolarization/backscatter relation, together with temperature and backscatter thresholds. The algorithm used in the version 2.01 release identifies obvious water and ice clouds and clear cases of oriented ice crystals. Improvements for recognizing mixed phase clouds are planned for future release.

- Bits 8-9, Ice/Water Phase Quality Assessment
- 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.

  Clouds

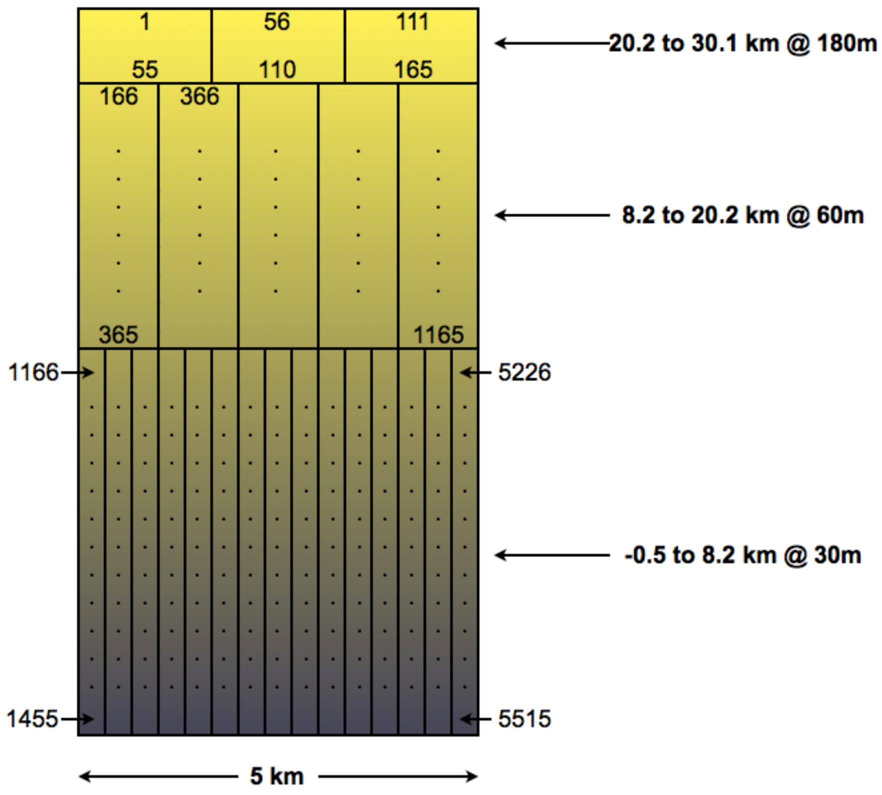
  This flag attempts to classify observed cloud layers into the standard meteorological cloud types (e.g., as defined by the ISCCP). The classification decisions are based on CALIOP measurements of cloud ice/water phase, cloud altitude, cloud fraction along an 80 km segment, and whether the cloud is opaque or transparent.

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 2.01 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 an 5515 element array (as rows in the HDF file) for a 5 km "chunk" of data. The numbers in this image indicate the column indices for the array. Only start and end indices are shown.

Reproduced from PC-SCI-503 : CALIPSO Data Products Catalog (Version 2.4) (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

Data Version Releases

Lidar Level 2 Vertical Feature Mask (VFM) Information Half orbit (Day) geolocated data radiances			
Release Date	**Version**	**Data Date Range**	**Maturity Level**
October 2008	2.02	September 14, 2008 to February 16, 2009; March 17, 2009 to March 28, 2010	• Layer Heights - Provisional • Aerosol/Cloud/ Stratospheric Classifications - Beta
January 25, 2008	2.01	June 13, 2006 to September 13, 2008	• Layer Heights - Provisional • Aerosol/Cloud/ Stratospheric Classifications - Beta

Data Quality Statement for the release of the CALIPSO Lidar Level 2 Vertical Feature Mask Products Version 2.02, October 2008

Version 2.02 of the Level 2 data products is a maintenance release that implements the following changes.

- Corrections were made to the code used to interpolate the GMAO meteorological data products to the CALIPSO orbit tracks.
- The Cabannes backscattering cross-sections used to derive the molecular scattering models used for the Level 1 and Level 2 analyses were revised downward by ~0.8%.
- A typographical error was identified in the runtime script that controls the behavior of the aerosol subtyping algorithm in the Level 2 analyses.

The impacts of these changes on the Level 2 data products are as follows:

- Layer detection: As a result of the first two changes, the 532 nm and 1064 nm calibration constants are larger, on average, by ~1%, resulting in corresponding decreases in the magnitudes of the attenuated backscatter coefficients at both wavelengths. These changes in the level 1 data result only small changes to the layer detection statistics. For example, the difference in the total number of layers detected by the two different versions on August 12, 2006 was 4: 9680 layers were detected by the version 2.01 code, versus 9676 layers by the version 2.02 code.

- Cloud-aerosol discrimination: with one exception, there were only minimal changes in cloud-aerosol discrimination results. The exception occurs in the polar regions when PSCs are present. For the August 12, 1006 test case, corrections to the interpolation algorithms applied to the GMAO data result in a slight upward shift in the tropopause heights, and as a consequence, more clouds and fewer stratospheric layers are identified in the version 2.02 results.

- Ice-water phase determination: because this classification is based on depolarization ratio and temperature no substantial changes, there were no substantial changes in the assessments of cloud thermodynamic state.

- Aerosol subtype identification: Correcting the level 2 runtime script error will reduce the number of layers identified as smoke, and increase the number of layers identified as sea salt.

Cloud and aerosol extinction profiles and optical properties: changes in backscatter and extinction coefficients at the tops of layers are small, and proportional to the changes in the calibration coefficients ... however, due to the cumulative nature of error propagation in the extinction retrieval, differences increase with increasing penetration depths, and can grow large when the optical depths of the clouds are large (i.e., > 3).

Data Quality Statement for the release of the CALIPSO Lidar Level 2 Vertical Feature Mask products Version 2.01, January 25, 2008

The CALIPSO vertical feature mask (VFM) data product reports a single 16-bit integer for each lidar altitude resolution element in the data stream downlinked from the satellite. Upon decoding each of these bit-mapped integers, users will obtain information describing layer location (both vertically and horizontally), layer type, and the amount of horizontal averaging required for the layer to be detected.

The primary new parameters included in the version 2.01 release of the vertical feature mask product are aerosol type classification and cloud ice/water phase discrimination. These are not independently derived results, but instead are the end products derived from a fully automated scene classification process. As such, the quality and correctness of the individual classifications depends not only on the specific input data and its associated pattern recognition algorithm, but also on the accuracy of several prior decisions made in other parts of the processing stream.

Layer Detection

Given the accuracy of the CALIPSO altitude registration, the layer heights reported in the Lidar Level 2 Cloud and Aerosol Layer Products appear to be quite accurate. In optically dense 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. 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

Based on the initial CALIOP measurements, an improved version of the cloud-aerosol discrimination algorithm has been implemented for this release. Overall, the updated algorithm works well in most cases; manual verification of the classifications for a full day of data suggest that the success rate is in the neighborhood of 90% or better. Nevertheless, several types of misclassifications are still occur with some frequency. Among these, the most prevalent are:

1. Dense aerosol layers (primarily very dense dust and smoke over and close to the source regions) are sometimes labeled as cloud. Because the CAD algorithm operates on individual layers, without a contextual awareness of any surrounding features, it can happen that small but strongly scattering regions within an extended aerosol layer can occasionally be labeled as cloud. This occurs because the optical properties (backscatter and color ratio) within the region are similar to what would be expected for the relatively faint clouds that fall within the PDF overlap region. These misclassifications are often apparent from studying the Level 1 browse images. Based on the initial analysis of the CALIOP measurements, the cloud and aerosol distributions show variabilities that depend on season and on geophysical location. The globally averaged PDFs used in the current release will have a larger overlap between the cloud and aerosol than would occur for more regionally specific statistics. For future versions of the CAD algorithm, we expect to develop and deploy PDFs that will correctly reflect both seasonal and latitudinal variations.

2. Many optically thin clouds, both ice and water, are encountered in the polar regions. The current CAD PDFs do not work as well in the polar regions as at lower latitudes and misclassifications of clouds as aerosol are more common. In particular, thin ice clouds which can extend from the surface to several kilometers in altitude, are sometimes misclassified as aerosol.

3. Correct classification of heterogeneous layers is always difficult, and the process can easily go awry. 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. By convention, heterogeneous layers should be classified as clouds. However, depending on the relative strengths of the components, these layers are sometimes erroneously identified as aerosol.

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 artifical 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; however, because they are not properly interdicted in the processing stream, the CAD algorithm nonetheless attempts to assign them to one class or the other. Very frequently these layers can be identified by their very low CAD scores (typically less than 20).

Aerosol Type Identification

While this set of aerosol subtypes does not cover all possible aerosol mixing scenarios it accounts for a majority of mesoscale aerosol layers. In essence, the classification algorithm trades off complex multi-component mixtures of aerosols for relatively stable layers under the assumption that the former are transient phenomena while the latter are stable, long-lived aerosol layers with considerable spatial extent (10-1000's km).

The algorithm assumes that in the ocean aerosol layers whose base height is greater than 1 km, (i.e., above the mean marine boundary layer height) are not marine aerosols. The algorithm assigns a biomass burning smoke type to such layers if the volume depolarization ratio is less than 0.075.

In Arctic and Antarctic regions, the airmasses can either be clean Arctic and Antarctic or Arctic haze. The aerosol types in the first two cases are modeled as clean continental types. Arctic haze results mainly from emissions of industrial pollution throughout Europe and Russia and is therefore modeled as polluted continental. The subtyping scheme does not allow the identification of dust in polar regions because the presence of low altitude ice particles (e.g., diamond dust) does not allow unambiguous attribution of high depolarization to dust.

Over the ocean, the algorithm classifies all aerosol elevated above 1 km and with volume depolarization ratios less than 0.075 as smoke. Since not all these are smoke layers, the frequency of smoke layers over the ocean will be somewhat biased high.

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

Cloud phase is determined using a depolarization/backscatter relation, together with temperature and backscatter thresholds. Complete descriptions of the algorithm mechanics and underlying theory are given in Section 6 of the CALIPSO Scene Classification ATBD (PDF). The algorithm implemented for the version 2.01 release identifies obvious water and ice clouds and clear cases of oriented ice crystals. Improvements for recognizing mixed phase clouds are planned for future release.