JPL D-62358

Earth Observing System

Multi-angle
Imaging
Spectro-
Radiometer

Level 3 Cloud Fraction by Altitude Algorithm Theoretical Basis

Larry Di Girolamo[1]
Alex Menzies[2]
Guangyu Zhao[1]
Kevin Mueller[2]
Catherine Moroney[2]
David J. Diner[2]

[1] University of Illinois at Urbana-Champaign
[2] Jet Propulsion Laboratory, California Institute of Technology

Jet Propulsion Laboratory
California Institute of Technology

August 19, 2010

JPL D-62358

Multi-angle Imaging SpectroRadiometer (MISR)

Cloud Fraction by Altitude Algorithm Theoretical Basis

Approval:

David J. Diner
MISR Principal Investigator

The MISR web site should be consulted to determine the latest released version of this document (http://www-misr.jpl.nasa.gov).
Approval signatures are on file with the MISR Project.

Jet Propulsion Laboratory
California Institute of Technology

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

GLOSSARY OF ACRONYMS

A

ASCM (Angular Signature Cloud Mask)
ASDC (Atmospheric Sciences Data Center)
ATB (Algorithm Theoretical Basis)

C

CALIPSO (Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation)
CCD (Charge-Coupled Device)
CFbA (Cloud Fraction by Altitude)
CPR (Cloud Profiling Radar)

E
EOS (Earth Observing System)

G

GEWEX(Global Energy and Water Cycle Experiment)

I

IFOV (Instantaneous Field Of View)
ISCCP (International Satellite Cloud Climatology Project)

L

LaRC (Langley Research Center)

M

MISR (Multi-angle Imaging SpectroRadiometer)

N

NN (Nearest Neighbor)
NISE (Near-real-time Ice and Snow Extent)
NSIDC (National Snow and Ice Data Center)

R

RCCM (Radiometric Camera-by-camera Cloud Mask)

S

SDCM (Stereoscopically Derived Cloud Mask)

T

TASC (Terrestrial Atmosphere and Surface Climatology)
TOA (Top-of-Atmosphere)

1 INTRODUCTION

1.1 PURPOSE

This Algorithm Theoretical Basis (ATB) document describes the algorithms used to produce the MISR Cloud Fraction by Altitude (CFbA) product. The parameters recorded in this product are summarized in Table 1. In particular, this document identifies sources of input data, both MISR and non-MISR, required for parameter retrievals; provides the physical theory and mathematical background underlying the derivation of CFbA; includes implementation details; and describes assumptions and limitations of the adopted approach. It is used by the MISR Science Data System Team to establish requirements and functionality of the data processing software.

Table 1: Parameters in the Cloud Fraction by Altitude Product

Parameter name	Units	Horizontal Sampling and Coverage	Vertical Sampling	Comments
CloudTopHeightFraction _Avg: Mean of cloud fraction	%	0.5×0.5 Lat/Lon Degree (Global)	• 500 m	• Cloud top height ranges from -500 m to 20 km • Also includes data for total cloud fraction and cloud fraction where no height is known
CloudTopHeightFraction _Std: Standard deviation of cloud fraction	%	0.5×0.5 Lat/Lon Degree (Global)	• 500 m	″
CloudTopHeightFraction _Num: Total number of samples	%	0.5×0.5 Lat/Lon Degree (Global)	• 500 m	″
CloudTopHeightFraction _NN_Avg: Mean of cloud fraction with the nearest-neighbor[1] algorithm applied	%	0.5×0.5 Lat/Lon Degree (Global)	• 500 m	″
CloudTopHeightFraction _NN_Std: Standard deviation of cloud fraction with the nearest-neighbor[1] algorithm applied	%	0.5×0.5 Lat/Lon Degree (Global)	• 500 m	″
CloudTopHeightFraction _NN_Num: Total number of samples with the nearest-neighbor[1] algorithm applied	%	0.5×0.5 Lat/Lon Degree (Global)	• 500 m	″

---

[1] See section 3.3.1.2 for a description of the nearest neighbor algorithm

1.2 SCOPE

This document covers the algorithm theoretical basis for the Cloud Fraction by Altitude product that will be routinely generated at the Langley Research Center (LaRC) Atmospheric Sciences Data Center (ASDC).

Chapter 1 describes the purpose and scope of the document. Chapter 2 provides a brief overview. The processing concept and algorithm description are presented in Chapter 3. Chapter 4 summarizes assumptions and limitations. References for publications cited in the text are given in Chapter 5. Literature references are indicated by a number in italicized square brackets, e.g., *[1]*.

1.3 MISR DOCUMENTS

Reference to MISR Project Documents is indicated by a number in italicized square brackets as follows, e.g., *[M-1]*. The MISR web site (http://www-misr.jpl.nasa.gov) should be consulted to determine the latest released version of each of these documents.

[M-1] Experiment Overview, JPL D-13407, Rev. A.

[M-2] Data Product Specification, JPL D-13963, Rev. R.

[M-3] Level 1 Cloud Detection Algorithm Theoretical Basis, JPL D-13397, Rev. B.

[M-4] Level 2 Cloud Detection and Classification Algorithm Theoretical Basis, JPL D-11399, Rev. D.

[M-5] Level 2 Ancillary Products and Datasets Algorithm Theoretical Basis, JPL D-13402, Rev. B.

1.4 REVISIONS

This is the original version of the document.

2 EXPERIMENT OVERVIEW

2.1 OBJECTIVES OF MISR CLOUD FRACTION BY ALTITUDE DETERMINATION

As a result of their large areal extent, high albedo, and variability on many timescales, clouds play a major role in governing the Earth's energy balance. Both global and regional studies of the impact of clouds on the energy balance require measurements of the radiation budgets as a function of scene type. The importance of cloud characteristics in global studies of climate has been well documented *[5]*. In the assessment of climate effect of clouds, cloud fraction and cloud top height are two of the most crucial climatological variables. Measurements of these two variables have been widely used to evaluate and parameterize global climate models. A joint distribution of the two variables serves better in constraining parameterization schemes in the climate models than either one. Although long term measurements of cloud fraction and cloud top height can be separately obtained from many metrological satellite cloud products, only the International Satellite Cloud Climatology Project (ISCCP) currently provides a global climatological product that summarizes a joint distribution of them *[6]*. As a result, the ISCCP cloud product has been widely used in the modeling community. However, both the vertical and horizontal resolutions of the ISCCP monthly, gridded product are lower than the grid resolution in current global climate models, making the dataset unsuitable for resolving the mesoscale characteristics of convection. High vertical resolution cloud climatologies can be constructed using datasets from active satellite remote sensors such as the Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) on CALIPSO and the Cloud Profiling Radar (CPR) on CloudSat, but their narrow swath can produce large sampling errors in monthly, gridded cloud cover at the resolution of current climate models *[1]*.

Deriving from its ability to measure any scene from multiple directions, MISR contributes accurate measurements of both cloud fraction and cloud top height at a high horizontal and vertical spatial resolution at a global scale. The joint distribution of cloud fraction and cloud height retrieved from MISR within the CFbA product currently represents the highest horizontal and vertical resolution cloud fraction by altitude product from any passive instrument.

A scientific background and historical perspective on related cloud studies using remote sensing, the unique contributions of MISR, and a scientific rationale for the cloud fraction by altitude parameter contents of the MISR TOA/Cloud Product are presented in *[M-1]*.

2.2 INSTRUMENT CHARACTERISTICS

The MISR instrument consists of nine pushbroom cameras. It is capable of global coverage every nine days, and flies in a 705-km descending polar orbit on the EOS-Terra platform. The cameras are arranged with one camera pointing toward the nadir (designated An), one bank of four cameras pointing in the forward along track direction (designated Af, Bf, Cf, and Df in order of increasing off-nadir angle), and one bank of four cameras pointing in the aftward direction (using the same convention but designated Aa, Ba, Ca, and Da). Images are acquired with nominal view zenith angles, relative to the surface reference ellipsoid, of 0°, 26.1°, 45.6°, 60.0°, and 70.5° for An, Af/Aa, Bf/Ba, Cf/Ca, and Df/Da, respectively. Each camera uses four Charge-Coupled Device (CCD) line arrays in a single focal plane. The line arrays consist of

1504 photoactive pixels plus 16 light-shielded pixels per array, each 21 μm by 18 μm. Each line array is filtered to provide one of four MISR spectral bands. The spectral band shapes are approximately gaussian and centered at 446, 558, 672, and 866 nm.

MISR contains 36 parallel signal chains corresponding to the four spectral bands in each of the nine cameras. The zonal overlap swath width of the MISR imaging data (that is, the swath seen in common by all nine cameras along a line of constant latitude) is 380 km, which provides global multi-angle coverage of the entire Earth in 9 days at the equator, and 2 days near the poles. The cross-track IFOV and sample spacing of each pixel is 275 m for all of the off-nadir cameras, and 250 m for the nadir camera. Along-track IFOVs depend on view zenith angle, ranging from 214 m in the nadir to 707 m at the most oblique angles. Sample spacing in the downtrack direction is 275 m in all cameras. The instrument is capable of buffering the data to provide 4 sample x 4 line, 2 sample x 2 line, or 1 sample x 4 line averages, in addition to the mode in which pixels are sent with no averaging. The averaging capability is individually selectable within each of the 36 channels, and there are several observational modes of the MISR instrument. The MISR TOA/Cloud Product is generated from Global Mode data, which provides pole-to-pole imagery with 1 x 1 resolution in all bands of the nadir camera, and in the red band at all angles. The remaining channels are averaged to 4 x 4 mode.

Most of the highest resolution observations are acquired in the red (672-nm) band, as this is the wavelength where the imagery has the highest contrast, based upon considerations of atmospheric haze, land and ocean reflectivity, and instrument performance. These observations are central to the stereoscopic and texture-based approaches used as part of MISR cloud classification *[M-4]*, which is used as input to the CFbA product.

Additional background on the instrument design is provided in *[M-1]*.

2.3 MISR CLOUD FRACTION DETERMINATION BY ALTITUDE STRATEGY

The MISR Cloud Fraction by Altitude product provides the frequency of cloud occurrence partitioned into different cloud top height bins at a global and monthly scale with a spatial resolution of 0.5° × 0.5° latitude/longitude and vertical resolution of 500 m. For each height bin, the frequency of cloud occurrence of a region over a time period is represented by the temporal mean of the spatial coverage of cloud tops. The spatial coverage of clouds is referred to as cloud fraction, which is defined in this document as the ratio of the number of cloudy pixels to the total number of cloudy and cloud-free pixels observed by the instrument. Clouds are assigned to height bins based on their top height as retrieved by the MISR stereoscopic technique *[M-4]*.

The accuracy of the reported cloud fraction relies on the validity of cloud/cloud-free classification. Cloud fractions are calculated using three MISR cloud masks, namely the Radiometric Camera-by-camera Cloud Mask (RCCM)*[M-3]*, Stereoscopically Derived Cloud Mask (SDCM)*[M-4]*, and Angular Signature Cloud Mask (ASCM)*[M-4]*, whose cloud detecting abilities differ from one another over different underlying surfaces. For example, the RCCM performs the best over dark ocean, while the ASCM performs the best over snow or ice covered surfaces. The three cloud masks are optimally combined within the Cloud Classifier product *[M-4]* to achieve the best cloud detection performance over all the underlying surfaces.

In order to provide statistically robust representation of cloud occurrence, the following stages of processing are performed: 1) Cloud fraction is first computed on an orbit-by-orbit basis within each grid box. 2) The daily mean cloud fraction of each grid box is calculated as the arithmetic mean of cloud fractions produced from stage 1 processing. 3) The monthly mean cloud fraction of each grid box is represented by the arithmetic mean of all the daily mean cloud fractions from step 2 processing. 4) The seasonal mean cloud fraction of each grid box is represented by the arithmetic mean of all the monthly mean cloud fractions from step 3 processing. 5) The annual mean cloud fraction of each grid box is represented by the arithmetic mean of all the seasonal mean cloud fractions from step 4 processing.

A large number of cloudy pixels may not have cloud top height retrievals, because of the nature of the stereoscopic technique used by MISR *[M-4]*. Many factors, including cloud type and underlying surface albedo, may contribute to a cloudy pixel having no retrieved height. This non-random, no-retrieval pattern may systematically bias the statistical results summarized for the CFbA product. To reduce this bias, an additional cloud fraction summary is calculated in which a cloudy pixel with no-height is assigned the height of the nearest-neighbor[1] cloudy pixel that has a valid height retrieval and is within 200 km. This assumes that cloud top height variations over the 200 km scale are well correlated. This is a reasonable assumption based on several studies of cloud correlation length statistics *[2] [3] [4]*.

---

[1] See section 3.3.1.2 for a description of the nearest neighbor algorithm

3 ALGORITHM DESCRIPTION

3.1 PROCESSING OUTLINE

Routine in-flight standard processing at the LaRC ASDC to generate the MISR Cloud Fraction by Altitude product occurs in several stages, described below.

(1) Calculate cloud fractions for each height bin at 0.5° × 0.5° latitude/longitude using the MISR Level 2 RCCM and Cloud Classifier products **for each orbit**

(2) Aggregate orbit data from step 1 into **daily** summaries

(3) Aggregate daily data from step 2 into **monthly** summaries

(4) Aggregate monthly data from step 3 into **seasonal**[1] summaries

(5) Aggregate seasonal data from step 4 into **annual**[2] summaries

Processing flow concepts are shown diagrammatically throughout the document. The convention for the various elements displayed in these diagrams is shown in Figure 1.

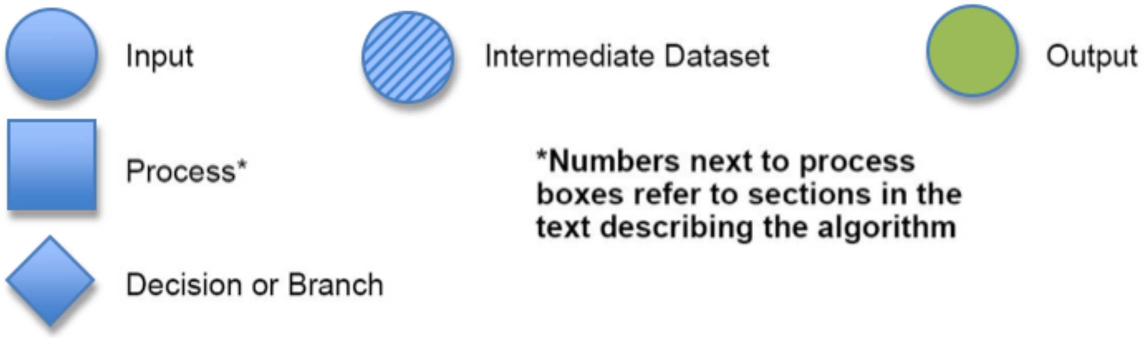

Figure 1. Conventions used in processing flow diagrams

---

[1] MISR defines each season as a time period consisting of exactly 3 months as described in Table 6

[2] MISR defines a year as 12 months beginning on the December 1st and running through the following November. Year 2009, for example, begins on Dec. 1, 2008 and runs through Nov. 30, 2009

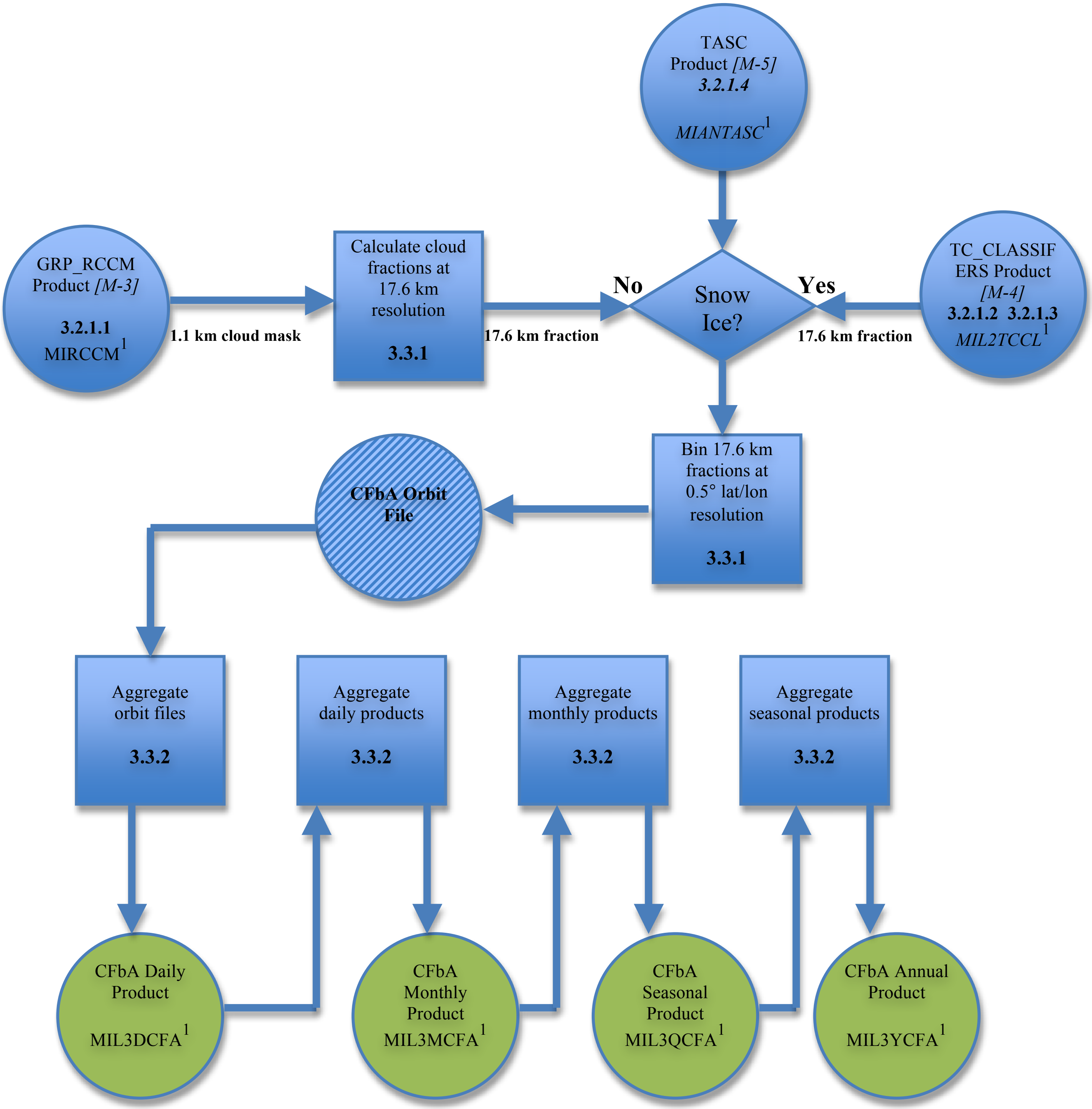

Figure 3. CFbA processing overview

[1] Data products available from the LaRC ASDC at *http://eosweb.larc.nasa.gov/*

3.2 ALGORITHM INPUT

The required inputs for the Cloud Fraction by Altitude product come from MISR and non-MISR sources and are summarized individually in the following paragraphs.

3.2.1 MISR data

Required inputs for the Cloud Fraction by Altitude product to be obtained from other MISR data products are summarized in Table 3. Further information on each of the inputs is provided below.

Table 3: Cloud Fraction by Altitude Inputs (MISR Data)

Input data	Source of data	Reference
Radiometric Camera-by-camera Cloud Mask - *Cloud* - *Glitter*	Level 1B2 RCCM product (*GRP_RCCM)* - 9 files (1 per camera) per orbit	*[M-3]*
Combined Cloud Fraction - *CombinedFractionCloudBestEstimate*	Level 2 TOA/Cloud Classifier product *(TC_CLASSIFIERS)*	*[M-4]*
Cloud Top Height - *MedianPrelimCloudHeight*	Level 2 TOA/Cloud Classifier product *(TC_CLASSIFIERS)*	*[M-4]*
Snow Ice Mask - *SnowIceMask*	Ancillary Terrestrial Atmosphere and Surface Climatology *(TASC)*	*[M-5]*

3.2.1.1 Radiometric Camera-by-camera Cloud Mask (RCCM)

The *Cloud* field of the MISR Level 1B2 RCCM product *[M-3]* provides 1.1 km cloud masks calculated per camera using radiometric information collected within each camera. It also contains a *Glitter* mask to flag 1.1 km pixels as potentially contaminated by sun glitter, derived on the basis that a particular view direction may be within a certain cone angle of the direction corresponding to specular reflection. Note that this calculation is based solely on scattering angle and does not take into account surface type. In CFbA processing, the RCCM is used to calculate cloud fractions over surfaces not covered by snow or ice.

3.2.1.2 Combined Cloud fraction

Cloud fraction best estimate is read from the *CombinedFractionCloudBestEstimate* field of the MISR Level 2 Cloud Classifier product and provides cloud fractions at a resolution of 17.6 km. Cloud fractions best estimate is derived in Level 2 using an optimal combination of several 1.1 km cloud masks *[M-4]*, namely the RCCM, SDCM, and ASCM. A cloud fraction is defined as the ratio of the number of 1.1 km cloudy pixels to the number pixels for which there was a valid retrieval within the containing 17.6 km region. The *CombinedFractionCloudBestEstimate* field is used only over snow and ice covered surfaces.

3.2.1.3 Cloud Top Height

This input comes directly from the *MedianPrelimCloudHeight* field of the MISR Level 2 Cloud Classifier product at 17.6 km resolution. In the Level 2 algorithm, cloud top heights are retrieved at a spatial resolution of 1.1 km using MISR's stereoscopic technique and further corrected using wind retrievals from the best-quality category. The wind retrieval technique assumes that there is no vertical cloud motion and that horizontal cloud motion is constant for a given altitude over distances of 70.4 km. The *MedianPrelimCloudHeight* field is calculated by taking the median value for all valid 1.1 km best-wind height retrievals within the enclosing 17.6 km region. Cloud top heights are reported relative to the WGS84 Surface Ellipsoid value, which is roughly equivalent to sea level. It should be noted that negative heights are possible in regions where the surface dips below the WGS84 Ellipsoid.

3.2.1.4 Snow Ice mask

The snow-ice mask is read from the *SnowIceMask* field of the MISR Ancillary TASC dataset [M-5]. The snow-ice mask has a temporal resolution of 1 month and a spatial resolution of 1 degree. For each 1-degree region, the snow-ice mask contains a flag indicating if the region was covered by snow or ice. The CFbA product uses this flag to determine which cloud fraction to use for each 0.5° × 0.5° grid cell. Over snow/ice covered grid cells, the combined cloud fraction from the MISR Cloud Classifier product is used; otherwise, the cloud fraction calculated from the RCCM is used.

The snow-ice mask contained in the TASC is supplied on a monthly basis from the NSIDC. The criteria for flagging a 1-degree region as covered by snow or ice is if 5% or more of the region is covered by snow/ice for more than 4 days out of the month based on NSIDC/NISE measurements.

3.3 THEORETICAL DESCRIPTION:

3.3.1 Cloud Fraction by Altitude produced for each orbit

3.3.1.1 Physics of the problem

The purpose of stage 1 processing is to calculate, on a per orbit basis, the cloud fraction of each 500 m height bin for each 0.5° × 0.5° latitude/longitude grid cell. The output files are generated on an orbit-by-orbit basis and are intended for JPL/DAAC internal use only.

3.3.1.2 Mathematical description of the algorithm

The first operation performed by the algorithm is to divide a given orbit into 17.6 km × 17.6 km regions and to calculate a cloud fraction for each region from the RCCM product *[M-3]*. The cloud fraction of each 17.6 km region, f^r, is calculated as the ratio of the number of 1.1 km cloudy pixels to the total number of cloudy and cloud-free pixels within the region.

$$f^r = \frac{\text{the number of 1.1 km cloudy pixels}}{\text{the number of 1.1 km cloudy and cloud-free pixels}} \tag{1}$$

A 1.1 km pixel is considered cloudy if the AN-RCCM *Cloud* field flags it as either high-confidence cloudy or low-confidence cloudy. However, if the An-RCCM for the pixel is contaminated with sunglint, as determined by the *Glitter* field in the An-RCCM product, the *Cloud* field in the Af-RCCM will be used. If the Af-RCCM is also contaminated with sunglint, the Aa-RCCM *Cloud* field will be used instead. If the Aa-RCCM is contaminated as well, then the Bf, Ba, Cf, Ca, Df, and Da cameras will be checked in order until a glint free pixel is found. If no glint free camera can be found, the 1.1 km pixel is treated as "no retrieval". The 17.6 km cloud fractions calculated from the RCCM will later be used over snow/ice free regions.

Next, the algorithm assigns a cloud top height to each 17.6 km × 17.6 km region. The heights are read in from the *MedianPrelimCloudHeight* field in the MISR L2 Cloud Classifiers product *[M-4]*. Heights are recorded in meters at 17.6 km resolution and are relative to the WGS84 elliposid. Both the cloud fractions and cloud top heights of each region for each orbit are stored as an intermediate dataset for further processing. Because the *MedianPrelim-CloudHeight* field was derived from the MISR stereo technique, some regions may not have valid height retrievals. The stereo algorithm requires high quality wind measurements to correct cloud height and will fail when such measurements are not available, such as over highly variable wind conditions or large homogenous cloud fields that lack the contrast necessary for accurate wind retrieval. When height retrievals are not available, a **nearest-neighbor** search of valid height retrievals within 200 km from the center location of a region will be conducted. The first returned nearest-neighbor valid height at 17.6 km resolution will be assigned to the region. If no valid height retrieval is returned, the region will be flagged as no retrieval. Two fields for the Cloud Fraction by Altitude outputs are stored in the intermediate product; one set is generated using the nearest-neighbor algorithm and the other without. By convention, both versions have the same name except the one utilizing the nearest-neighbor algorithm has a suffix of "NN" as shown in Table 1.

Next, the 17.6 km cloud fractions derived from the RCCM are projected into 0.5° × 0.5° latitude/longitude grid boxes and binned by altitude using the previously computed 17.6 km cloud top heights.. The grids are defined to break latitude and longitude into half-degree increments ranging from 90° N to 90° S and 180° W to 180° E, respectively. A region belongs to a grid box if its center is within the boundaries of the grid box. For each half-degree box, cloud fractions are further binned by their heights. Height is broken up into 500 m increments and ranges from -500 m to 20 km. There are 4 additional height bins that represent: values less than -500 m, values above 20 km, negative infinity to infinity representing total cloud fraction, and no height retrieval. The convention used for the height bins is given in Table 3.

Table 5: Height bin convention

h (bin number)	1	2	3	…	42	43	44	45
Height range [km]	(-∞, -0.5)	[-0.5, 0)	[0, 0.5)	…	[19.5, 20)	[20, ∞)	(-∞, ∞)	no retrieval

The mean, standard deviation, and number of cloud fractions of the regions sampled in each bin are calculated and reported at each grid box. The formulas used to calculate the mean and standard deviation are shown below, respectively, and examples are given in Table 4.

$$f^{o}_{lat,lon,h} = \begin{cases} -9999, N_{lat,lon,h} = 0 \\ \dfrac{1}{N_{lat,lon,h}} \sum_{i=1}^{N_{lat,lon,h}} f_i^r, N_{lat,lon,h} > 0 \end{cases} \tag{2}$$

$$\sigma(f^{o}_{lat,lon,h}) = \begin{cases} -9999, N_{lat,lon,h} = 0 \\ 0, N_{lat,lon,h} = 1 \\ \sqrt{\dfrac{\sum_{i=1}^{N_{lat,lon,h}} (f_i^r - \bar{f}^r)^2}{N_{lat,lon,h} - 1}}, N_{lat,lon,h} > 1 \end{cases} \tag{3}$$

where f^o represents the cloud fraction for a certain *lat,lon* grid box and height bin, *h*; *lat, lon*, and *h* are indices to latitude, longitude grids and height bins, respectively; *N* is the total number of the 17.6 km cloud fractions (f^r) sampled for height bin, *h*, but only with $f^r \geq 0$; and $\bar{f}^r = \dfrac{\sum_{i=1}^{N_{lat,lon,h}} (f_i^r)}{N_{lat,lon,h}}$. "-9999" represents "no retrieval" value.

Table 6: An example of the calculation of mean cloud fraction for each height bin

Cloud height bin	Input cloud fractions at 17.6 km	Output cloud fraction of each height bin
Bin #1	0.50;0.25;0.75	(0.50 + 0.25 + 0.75) / 3 = 0.5
Bin #2	1.0;-9999	1.0/1 =1.0
Bin #3	0.3;0.2;-9999	(0.3 + 0 .2) / 2 = 0.25

Finally, the half-degree cloud fraction summary based on the RCCM is written out to an intermediate, orbit-based dataset. An additional half-degree cloud fraction summary derived from the MISR Level 2 Could Classifier product is also written to the intermediate file. This second summary is generated in a manner similar to the RCCM summary, the only difference being that cloud fractions for each 17.6 km region are read directly from the *CombinedFraction-CloudBestEstimate* field of the Cloud Classifier product (Section 3.2.1.2) instead of calculated from the RCCM as in Equation (1).

3.3.2 Generation of monthly, seasonal, and annual CFbA

3.3.2.1 Physics of the problem

This section describes the process used to aggregate the intermediate orbit-based datasets into publicly available[1] monthly, seasonal, and annual CFbA products.

3.3.2.2 Mathematical description of the algorithm

Due to the nature of the MISR Level 2 product, some orbits are poorly registered and contain no valid heights in the *MedianPrelimCloudHeight* field from the Cloud Classifiers product. Before processing continues, intermediate datasets associated with these orbits are excluded from the set of inputs used in generating the public product. The intermediate files to exclude are identified by checking to see if the sample count is 0 for all height bins except the "no retrieval" bin.

The aggregation process begins by combining multiple intermediate orbit files to generate a daily CFbA summary. However, before the intermediate files can be averaged together they must each be normalized so that the sum across all height bins (excluding the total cloud fraction bin; i.e., Bin #44 in Table 3) for any 0.5° × 0.5° grid box is equal to the total cloud fraction of the grid box. This is done by multiplying each bin by the number of values it contains to obtain the sum of the cloud fractions in that bin, and then dividing each bin by the total number of samples across all heights for a given grid box. The formula used to do this normalization is shown below and an example is provided in Table 5.

$$\hat{f}^{o}_{lat,lon,h} = \begin{cases} -9999, N_{lat,lon,h} = 0 \\ \dfrac{f^{o}_{lat,lon,h} \bullet N_{lat,lon,h}}{\sum_{h=1}^{43} N_{lat,lon,h} + N_{lat,lon,45}}, N_{lat,lon,h} > 0 \end{cases} \tag{4}$$

where $\hat{f}^{o}$ is the renormalized cloud fraction, and $N_{lat,lon,45}$ is the number of observations in the no

---

[1] Data products available from the Langley DAAC at *http://eosweb.larc.nasa.gov/*

height retrieval bin.

Table 7: An example of renormalization of CFbA intermediate file. In this example, only three bins (bin #1, #2, #3) have cloud fractions larger than zero.

Cloud height bin	**Before normalization**		**Cloud fraction after normalization**
	Cloud Fraction	**# of samples**	
Bin #1	0.50	3	0.5*3 /(3+1+2) = 0.25
Bin #2	1.0	1	1.0*1/(3+1+2)=0.17
Bin #3	0.25	2	0.25 *2/(3+1+2) = 0.08

Note that the no-height-retrieval bin is included when doing the renormalization. In other words, the count in this bin is taken as part of the total number of samples. The normalization is done for both the fields RCCM and Cloud Classifier based summaries **with and without using the nearest-neighbor algorithm** as described in section 3.3.1.2. After renormalization, the two summaries are then combined to form the field *CloudTopHeightFraction*. The logic in the combination is given as follows: If a grid box is covered by snow or ice, then *CloudTopHeightFraction* uses the Cloud Classifier derived fraction; Otherwise, *CloudTopHeightFraction* uses the RCCM derived cloud fraction. Whether a grid box is over snow or ice is determined by the *SnowIceMask* field in the TASC ancillary dataset.

Next, the *CloudTopHeightFraction* fields from all the orbits for a single day are averaged together with each orbit assigned equal weight. Orbits are given equal weight because each one is considered to be an independent estimate of cloud fractions. The number of samples for each grid box and height bin will now represent the number of orbits that went into each average as opposed to the number of 17.6 km×17.6 km regions that went into the intermediate orbit file. Likewise, the standard deviation will now represent the variance among orbits instead of 17.6 km samples. However, before the average takes place for a certain *lat,lon* grid box, if there exists any of $\{\hat{f}^{o(j)}_{lat,lon,1}, \hat{f}^{o(j)}_{lat,lon,2} \ldots \hat{f}^{o(j)}_{lat,lon,43}\} \geq 0$ for orbit j, then any of $\hat{f}^{o(j)}_{lat,lon,h}\ (h = 1,2,..43) = -9999$ will be reassigned a value of 0. However, if there does not exist any of $\{\hat{f}^{o(j)}_{lat,lon,1}, \hat{f}^{o(j)}_{lat,lon,2} \ldots \hat{f}^{o(j)}_{lat,lon,43}\} \geq 0$, then orbit j will be excluded from the calculation for this *lat, lon* grid box. In other words, if none of the height bins for the grid box contain a valid cloud fraction then they are considered to be outside of the orbit's swath and are not included in the average. The formulas used to calculate the mean and standard deviation are given below.

$$f^{d}_{lat,lon,h} = \begin{cases} -9999, M_{lat,lon,h} = 0 \\ \dfrac{1}{M_{lat,lon,h}} \sum_{j=1}^{M_{lat,lon,h}} \hat{f}^{o(j)}_{lat,lon,h}, M_{lat,lon,h} > 0 \end{cases} \tag{5}$$

$$\sigma(f^{d}_{lat,lon,h}) = \begin{cases} -9999, M_{lat,lon,h} = 0 \\ 0, M_{lat,lon,h} = 1 \\ \sqrt{\dfrac{\sum_{j=1}^{M_{lat,lon,h}} (\hat{f}^{O(j)}_{lat,lon.h} - \overline{\hat{f}}^{O(j)}_{lat,lon.h})^2}{M_{lat,lon,h} - 1}} M_{lat,lon,h} > 1 \end{cases} \tag{6}$$

where, f^{d} is the daily cloud fraction, the index j is associated with an orbit number, and M is the total number of orbits sampled in a day, only for those with valid $\hat{f}^{o(j)}_{lat,lon,h}$. ; $\overline{\hat{f}}^{o}_{lat,lon,h} = \dfrac{\sum_{j=1}^{M_{lat,lon,h}} (\hat{f}^{o(j)}_{lat,lon,h})}{M_{lat,lon,h}}$

.

To generate a monthly summary, all the daily summaries for the month are read and averaged together with equal weight assigned to each day.

$$f^{m}_{lat,lon,h} = \begin{cases} -9999, O_{lat,lon,h} = 0 \\ \dfrac{1}{O_{lat,lon,h}} \sum_{j=1}^{O_{lat,lon,h}} f^{d}_{lat,lon,h}, O_{lat,lon,h} > 0 \end{cases} \tag{7}$$

$$\sigma(f^{m}_{lat,lon,h}) = \begin{cases} -9999, O_{lat,lon,h} = 0 \\ 0, O_{lat,lon,h} = 1 \\ \sqrt{\dfrac{\sum_{j=1}^{O_{lat,lon,h}} (f^{d}_{lat,lon,h} - \overline{f^{d}_{lat,lon,h}})^2}{O_{lat,lon,h} - 1}}, O_{lat,lon,h} > 0 \end{cases} \tag{8}$$

where f^m is the monthly cloud fraction; and O is the number of days with $f^d \geq 0$. In the final product, $f^m_{lat,lon,h}$, $\sigma(f^m_{lat,lon,h})$, and $O_{lat,lon,h}$ are stored in fields *CloudTopHeightFraction_Av*e, *CloudTopHeightFraction_Std*, and *CloudTopHeightFraction_Num*, respectively. When producing a monthly mean estimate of cloud fraction, daily estimates are treated as independent samples of cloud fraction and thus given equal weight. This ensures that that the estimate of the mean cloud fraction is unbiased, unlike weighting by the number of samples in an orbit or day, which would bias the estimated mean cloud fraction towards the cloud fraction of a particular orbit or day with the most samples.

The generation of seasonal and annual products is similar to that of monthly products. To generate a seasonal summary, the cloud fractions from 3 monthly summaries are averaged together with equal weight assigned to each month. Each quarter consists of exactly 3 months and is named after the season it most closely resembles as shown in Table 6. When generating a annual summary, 12 monthly summaries are averaged together with equal weight assigned to each quarter.

Table 8: MISR season definition

Season	Months
Winter	December (previous year), January, February
Spring	March, April, May
Summer	June, July, August
Autumn	September, October, November

3.4 ALGORITHM VALIDATION

A variety of unit tests have been developed to verify the correct operation of the algorithm when run on both contrived and real world data. In addition, the algorithm has a number of built-in error detection mechanisms that ensure specific inequalities and invariants are maintained throughout processing.

The RCCM, SDCM, ASCM, and stereo height products used to produce CFbA have been validated. The evaluation of their performances can be found at http://eosweb.larc.nasa.gov/PRODOCS/misr/Quality_Summaries/misr_qual_stmts.html. Intercomparison of the CFbA product and other cloud products retrieved for other satellite sensors within the Global Energy and Water Cycle Experiment (GEWEX) Cloud Assessment is being conducted.

4 ASSUMPTIONS AND LIMITATIONS

4.1 ASSUMPTIONS

The following assumptions are made with respect to cloud fraction determination described in this document:

(1) An accurate snow-ice mask is provided on a monthly basis in the MISR TASC dataset.

(2) In the nearest-neighbor search window of 200 km, cloud top heights are assumed to be correlated.

4.2 LIMITATIONS

The following limitations apply to the cloud fractions described in this document:

(1) Cloud fractions are for the highest cloud tops and this does not preclude the presence of additional lower-level clouds that may not have been observed due to obscuration by the higher cloud.

(2) Thick aerosols may be flagged as clouds and contaminate the calculated cloud fractions.

(3) Cloud fraction may be overestimated over the regions dominated by cumulus clouds. *[7]*

(4) When multilayer clouds occur, cloud top heights may be retrieved for the lower level clouds, if they exhibit stronger feature contrast than the upper-level clouds. As a result, high-level cloud fraction may be underestimated.

5 REFERENCES

[1] Astin, I., and L. Di Girolamo (2001). Baysian confidence intervals for true fractional coverage from finite transect measurements: implications for cloud studies from space. *J. Geophys. Res*., **116** (D15), 17,303-17,310.

[2] Cahalan, R.F., D.A. Short, and G.R. North (1982). Cloud fluctuation statistics. *Mon. Weath. Rev*., **110**, 26-43.

[3] Davies, R., (1994). Spatial autocorrelation of radiation measured by the Earth Radiation Budget Experiment: Scene inhomogeneity and reciprocity violation. *J. Geophys. Res*., **99** (D10), 20,879-20,887.

[4] Jones, P.A., 1992: Cloud-cover distributions and correlations. *J. Appl. Meteor*., **31**, 732-741.

[5] Rossow, W.B., A.W. Walker, D.E. Beuschel, and M.D. Roiter (1996). International Satellite Cloud Climatology Project (ISCCP) Documentation of New Cloud Datasets. WMO/TD-No. 737, World Meteorological Organization, 115 pp.

[6] Schiffer, R. A. and W. B. Rossow (1983). The International Satellite Cloud Climatology Project (ISCCP): The First Project of the World Climate Research Program (WCRP). *Bull. Amer. Met. Soc*. **64**, 779.

[7] Zhao, G. and L. Di Girolamo (2006). Cloud fraction errors for trade wind cumuli from EOS-Terra instruments, *Geophys. Res. Lett*., **33**, L20802, doi:10.1029/2006GL027088.