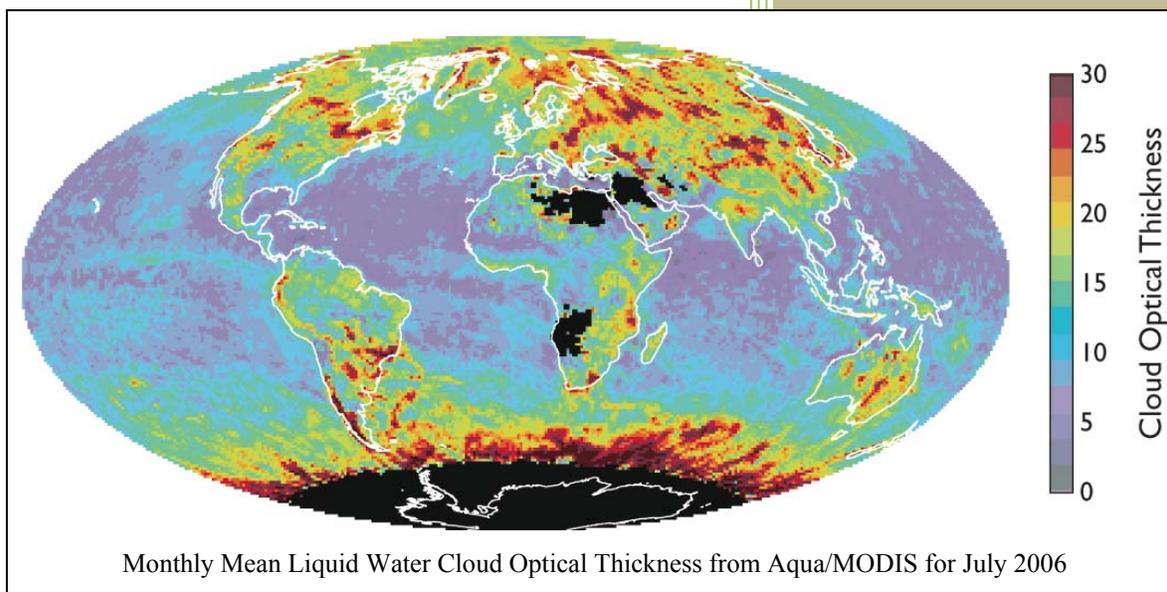


MODIS Atmosphere L3 Gridded Product Algorithm Theoretical Basis Document



PAUL A. HUBANKS¹, MICHAEL D. KING^{2,3}, STEVEN A. PLATNICK², AND ROBERT A. PINCUS⁴

MODIS Algorithm Theoretical Basis Document No. ATBD-MOD-30
for Level-3 Global Gridded Atmosphere Products (08_D3, 08_E3, 08_M3)

(Collection 005 Version 1.1, 4 December 2008)

¹ Wyle Information Systems LLC, Lanham, MD

² Earth Sciences Division, NASA Goddard Space Flight Center, Greenbelt, MD

³ Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO

⁴ Climate Dynamics Center, NOAA Earth System Research Laboratory, Boulder, CO

TABLE OF CONTENTS

1.0. Introduction	1
1.1. Approach.....	3
1.1.1. Time span.....	4
2.0. Gridding.....	5
2.1. Computation of geolocation in L2 atmosphere products.....	12
2.1.1. Dropped pixels in L2 products.....	13
2.2. Sampling	14
2.2.1. Sampling modification to avoid dead detectors.....	15
3.0. Computation of L3 Daily Statistics.....	17
3.1. Aggregation and QA weighting of statistics.....	17
3.1.1. Aggregation of statistics based on physical properties.....	18
3.1.2. QA weighting of statistics based on confidence.....	18
3.2. Types of daily statistics computed.....	20
3.2.1. Simple statistics	21
3.2.2. QA-weighted simple statistics	21
3.2.3. Fraction statistics	22
3.2.4. Pixel count statistics.....	22
3.2.5. Logarithm statistics.....	22
3.2.6. Uncertainty statistics.....	23
3.2.7. Logarithm of uncertainty statistics	24
3.2.8. Histogram.....	24
3.2.9. Histogram of confidence.....	25
3.2.10. Joint histogram.....	25
3.2.11. Joint regression	25
3.3. Special handling logic in daily statistics.....	26

3.3.1. Partial retrieval fix for cloud optical thickness.....	26
4.0. Computation of L3 Multiday (8-Day & Monthly) Statistics.....	27
4.1. Computational weighting schemes for multiday statistics	27
4.2. Types of multiday statistics computed.....	28
4.2.1. Simple statistics based on the daily mean.....	28
4.2.2. Simple statistics based on the daily standard deviation.....	30
4.2.3. QA-weighted statistics based on the daily mean	30
4.2.4. Fraction statistics	31
4.2.5. Pixel count statistics.....	31
4.2.6. Logarithm statistics.....	32
4.2.7. Uncertainty statistics.....	32
4.2.8. Logarithm of uncertainty statistics	33
4.2.9. Histogram.....	33
4.2.10. Histogram of confidence.....	34
4.2.11. Joint histogram.....	34
4.2.12. Joint regression	35
5.0. Reading and Unpacking HDF File Data.....	35
5.1. Descaling the SDS data.....	36
5.2. Definitions of local attributes	37
5.2.1. Local attributes used in all L3 files.....	37
5.2.2. Local attributes used in the L3 daily file only	39
5.2.3. Local attributes used in the L3 eight day and monthly file only	39
6.0. Using the MODIS-Atmosphere Web Site.....	40
7.0. Interpretation of Data: Frequently Asked Questions.....	44
7.1. How do the cloud mask and optical properties ‘cloud fractions’ differ?.....	44
7.1.1. Cloud mask cloud fraction.....	44
7.1.2. Cloud optical properties cloud fraction	45

7.1.3. Comparing the cloud mask with the optical properties cloud fraction ...	47
7.2. What do the ‘undetermined’ and ‘combined’ cloud phases mean?	50
7.3. What is the minimum illumination requirement for a MODIS retrieval?	52
7.4. What is the best way to display (marginal) histogram data?	52
7.5. What is the best way to display joint histogram data?.....	54
7.6. How is the aerosol cloud mask computed and how should it be interpreted?	60
7.7. How should the various multilayer cloud fractions be interpreted?	62
7.8. Is there a minimum L2 pixel count requirement to compute L3 daily statistics?63	
7.9. Are statistics within a L3 Daily file a single-time snapshot or a daily average?65	
8.0. References	68
Appendix A: Scientific Data Sets found in the L3 Daily File only	69
Appendix B: Scientific Data Sets common to both the L3 Daily and L3 Multiday Files	74
Appendix C: Scientific Data Sets found in the L3 Multiday File only.....	79
Appendix D: Histogram Parameters & Bin Boundaries	84
Appendix E: Joint Histogram Parameters & Bin Boundaries.....	88

1.0. Introduction

This document describes the theoretical basis of MODIS Atmosphere Level 3 (L3) products. Topics covered include computational approaches and scenarios, subsampling and gridding characteristics and caveats, the impact of Level 2 (L2) confidence flags on L3, and finally a breakdown of the parameters and statistics computed in L3.

The “Level” terminology is used to denote broad categories of MODIS data products: Level 0 (L0) denotes raw spectral channel counts, Level 1B (L1B) denotes calibrated and geolocated radiances, Level 2 (L2) denotes orbital-swath science products, and finally Level 3 (L3) denotes global-gridded science products.

All L3 MODIS Atmosphere products, retrieved from MODIS instrument and ancillary data, are grouped into user-friendly Hierarchical Data Format (HDF) files based upon the platform (Terra or Aqua) and temporal period of collection (Daily (08_D3), Eight Day (08_E3), and Monthly (08_M3)).

Terra data use a MOD prefix in the HDF file name. Aqua data use a MYD prefix. So, for example, a Terra L3 Daily file would have an HDF filename prefix of MOD08_D3 and an Aqua L3 Daily file would have an HDF filename prefix of MYD08_D3.

The Terra overpass time at the equator is around 1030 local solar time in its descending (daytime) mode and 2230 local solar time in its ascending (nighttime) mode. The Aqua overpass time is around 1330 local solar time in ascending (daytime) mode and 0130 local solar time in descending (nighttime) mode.

L3 Atmosphere products contain hundreds of $1^\circ \times 1^\circ$ global gridded Scientific Data Sets (SDSs) or statistics derived from the L2 products of Aerosol (04_L2), Water Vapor (05_L2), Cloud (06_L2), and Atmospheric Profile (07_L2). L3 SDSs derived from Cloud Mask (35_L2) related data are actually read from (duplicated in) the Cloud (06_L2) product, therefore the L2 Cloud Mask (35_L2) product is never read (directly)

into L3.

MODIS atmosphere products are grouped as follows. Earth Science Data Type (ESDT) names, used to identify each product by their HDF filename, are provided in parentheses.

Level 2 Products

- Aerosol Properties (04_L2)
- Water Vapor (or Precipitable Water) (05_L2)
- Cloud Properties (06_L2)
- Atmospheric Profiles (07_L2)
- Cloud Mask (35_L2)
- Joint Atmosphere (ATML2)

Level 3 Products

- Daily Global (08_D3)
- Eight-day Global (08_E3)
- Monthly Global (08_M3)

Each Level 2 atmospheric parameter is retrieved at a spatial resolution determined by the sensitivity of retrieval, not necessarily on a single field of view (FOV) basis of the MODIS spectral band retrieval needed to perform the retrieval. Resolutions of L2 science products are at 1×1 , 5×5 , and 10×10 km. Level 3 atmospheric parameters are computed at $1^\circ \times 1^\circ$ spatial resolution.

MODIS Level-3 Hierarchical Data Format (HDF) product files have standardized filenames that are detailed below.

Terra (AM) Platform:

MOD08_D3.AYYYYDD.D.VVV.YYYYDDHHMMSS.hdf

Aqua (PM) Platform:

MYD08_D3.AYYYYDD.D.VVV.YYYYDDHHMMSS.hdf

Definitions:
MOD08_D3 = Earth Science Data Type Name
A = Acquisition Date
YYYYDDD = Data Year and Day of Year
VVV = Collection Version
YYYYDDHHMMSS = Processing Date & Time
hdf = Suffix denoting HDF file

Note that: (i) all times are UTC times, not local times, and (ii) the MOD08_D3 prefix represents a daily L3 Terra platform file. Daily Aqua platform files have the prefix MYD08_D3. Eight day files have the prefix MOD08_E3 and MYD08_E3. Monthly files have the prefix MOD08_M3 and MYD08_M3.

The Collection “version” denotes a complete set of MODIS files that are earmarked by comprehensive changes (updates) to the science algorithms. Collection 5 was completed in calendar year 2006. Collection 6 is expected to begin in early 2009. Details on the changes implemented in each Collection are available in the “products” section of the MODIS-Atmosphere web site (modis-atmos.gsfc.nasa.gov). Occasionally significant updates are implemented in the middle of a Collection. This is only done when a L2 or L3 operational algorithm software bug is discovered that seriously impacts one or more of the Scientific Data Sets (SDSs) contained within a L2 or L3 HDF file. Scientists working with MODIS data should always be aware of updates applied to the operational software (especially those applied in the middle of a collection) by visiting the ‘PGE History’ page in the “products” section of the MODIS-Atmosphere web site (see link above).

1.1. Approach

The Collection 5 L3 MODIS Atmosphere Daily Global Product contains nearly 700 statistical Scientific Data Sets (SDSs) that are derived from 106 scientific parameters read from four Level-2 MODIS Atmosphere Products: Aerosol, Water Vapor, Cloud, and Atmosphere Profile.

The Collection 5 L3 MODIS Atmosphere Eight-Day and Monthly Global Products contain over 900 statistical Scientific Data Sets (SDSs) that are derived from statistics within the Daily L3 Global product only.

All statistics are sorted into $1^\circ \times 1^\circ$ cells on an equal-angle global grid (see section 2.0). A number of statistical summaries are computed for each of these L3 products, de-

pending on the parameter being considered. Statistics for a given measurement might include:

- Simple (mean, minimum, maximum, standard deviation) statistics
- Parameters of normal and log-normal distributions
- Fraction of pixels that satisfy some condition (e.g., cloudy, clear)
- Histograms of the quantity within each grid box
- Histograms of the confidence placed in each measurement
- Joint Histograms and/or regressions derived from comparing one science parameter to another, statistics may be computed for a subset that satisfies some condition.

1.1.1. Time span

L2 MODIS data granules (input to L3 Daily) are always time-stamped using Coordinated Universal Time. Coordinated Universal Time is International Atomic Time (TAI) with leap seconds added at irregular intervals to compensate for the Earth's slowing rotation. These added leap seconds allow Coordinated Universal Time to track the mean solar time at the Royal Observatory, Greenwich (GMT).

Coordinated Universal Time is abbreviated UTC. The International Telecommunication Union wanted Coordinated Universal Time to have a single abbreviation for all languages. English speakers and French speakers each wanted the initials of their respective languages' terms to be used internationally: "CUT" for "coordinated universal time" and "TUC" for "*temps universel coordonné*." This resulted in the final compromise of using "UTC."

Each L2 MODIS granule (stored in a single HDF file) contains 5 minutes of data. The time stamp (which is part of the HDF file name) shows the start of the 5-minute data collection period. For example, a L2 MODIS granule (contained within a single HDF file) might have a time stamp of 1430. This means the start time of data acquisition was

14:30:00 UTC. The subsequent HDF granule would have a time stamp of 1435, which corresponds to an acquisition start time of 14:35:00 UTC.

The Daily L3 product contains statistics computed from a set of L2 MODIS granules (HDF files) that span a 24-hour (00:00:00 to 23:59:59 UTC) interval. In the case where a L2 parameter is only computed during the daytime, then only daytime files are read to compute the L3 statistics.

The Eight-Day L3 product is computed from 8 consecutive Daily L3 product files. The start date for the 8-day interval is always reset to January 1st at the beginning of each calendar year, however the last 8-day interval in each year is allowed to span over the first few days of January. For example, the last eight-day interval for 2005 runs from 12/27/05 to 1/3/06, while the first eight-day interval for 2006 runs from 1/1/06 to 1/8/06. Users should note these days of overlap when evaluating a set of 8-day product files that span several years.

The Monthly L3 product is computed from the complete set of daily files that span a particular month. There are never any days of overlap (like that which exists in the eight-day product).

2.0. Gridding

All MODIS Atmosphere L3 SDSs are stored on an equal-angle latitude-longitude grid (cf. Figure 1).

The grid cells are $1^\circ \times 1^\circ$, which means the MODIS Atmosphere L3 output grid is always 360 pixels in width and 180 pixels in height. An example of a MODIS Atmosphere L3 Scientific Data Set (SDS) mapped on a $1^\circ \times 1^\circ$ latitude-longitude grid is shown in Figure 2.

The L3 grid cell indexed (1,1) in the SDS is located at the upper left corner of the map and corresponds to a grid box with boundaries of 89° to 90° N latitude and 179° to 180° W longitude.

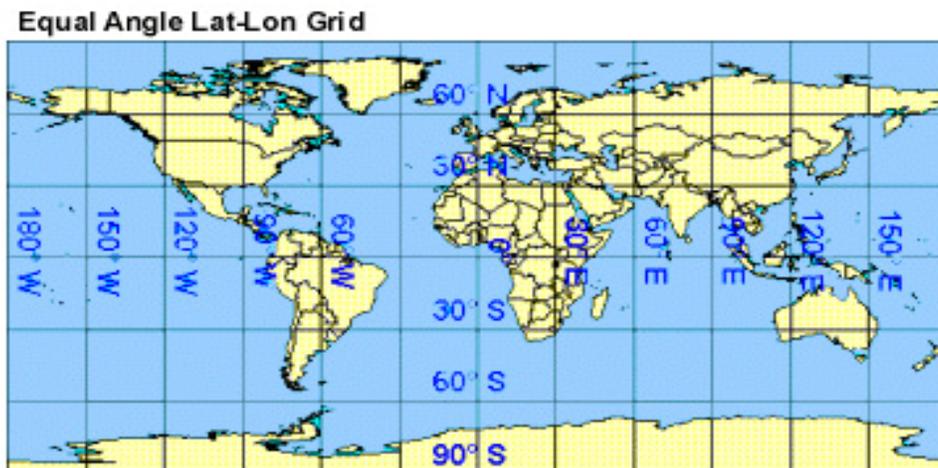


Figure 1. A latitude-longitude rectangular grid.

The assignment of L2 pixels (to a L3 grid cell) that fall exactly on a L3 grid boundary is performed using the following convention: L2 pixels that fall exactly on the first whole degree boundary 90°N ($+90.0$) latitude and 180°W (-180.0) longitude are binned in the L3 grid column and row #1. L2 pixels that fall exactly on the second whole degree boundary 89°N ($+89.0$) and 179°W (-179.0) are binned in L3 grid column and row #2. The exception to this logic occurs in the last L3 grid row (89°S to 90°S), which contains both whole degree latitude boundary pixels (-89.0 as well as -90.0). There is no exception for the last L3 grid column (179°E to 180°E) since $+180.0$ and -180.0 represent the same physical location (these L2 pixels are binned in the first L3 grid column).

An important property to note when considering L3 gridding occurs due to distortion

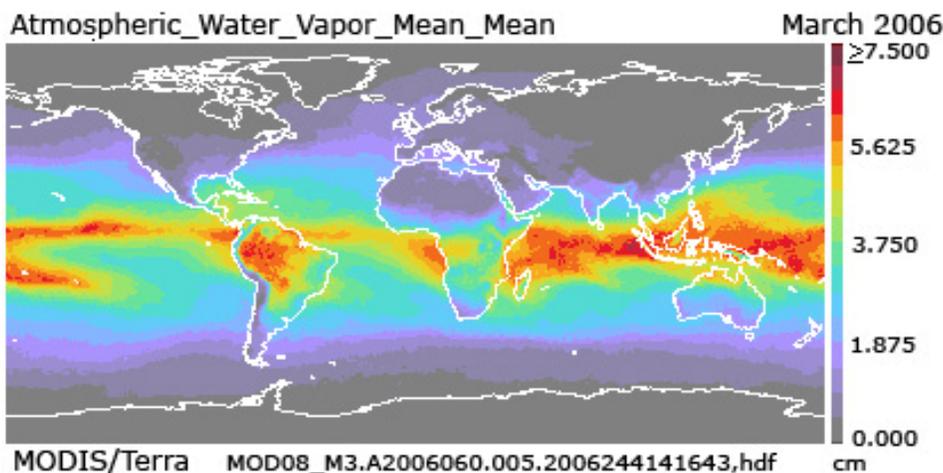


Figure 2. An example L3 SDS shown in its native latitude-longitude projection format.



Figure 3. Convergence of longitude lines produces shrinking of $1^\circ \times 1^\circ$ grid cells toward the poles.

in the latitude-longitude map projection as one moves poleward. The actual (real world) size of each $1^\circ \times 1^\circ$ grid box physically shrinks when moving from the equator toward the pole due to the convergence of longitude lines on the globe (see Figure 3). At the equator each $1^\circ \times 1^\circ$ grid cell is roughly $12,321 \text{ km}^2$ in size, at the pole each $1^\circ \times 1^\circ$ grid cell is only 107 km^2 or less than $1/100^{\text{th}}$ the size.

It should be noted that there is also a variation of pixel size in L2 (input) products due to viewing (scan angle) distortion. For example, in 1 km (nadir) resolution L2 data, the L2 pixels expand due to view angle distortion when moving from nadir towards the limb (high scan angles) of the MODIS scan (see Figures 4 and 5). These same distortion factors apply to 5 and 10 km native resolution L2 pixels.

This viewing distortion of L2 input pixels means that there are fewer L2 pixels to average in the computation of L3 daily statistics for L3 $1^\circ \times 1^\circ$ grid boxes that cover regions of high scan angles in the L2 orbital products than for those that cover L2 orbital products computed closer to nadir view. For multiday products, this effect is averaged out. It should be noted that this “growth” of the L2 pixel also causes some pixel (scene)

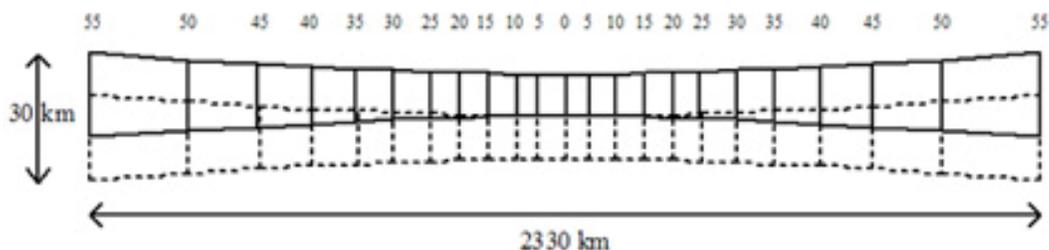


Figure 4. L2 pixel overlap due to scan width growth as a function of scan angle.

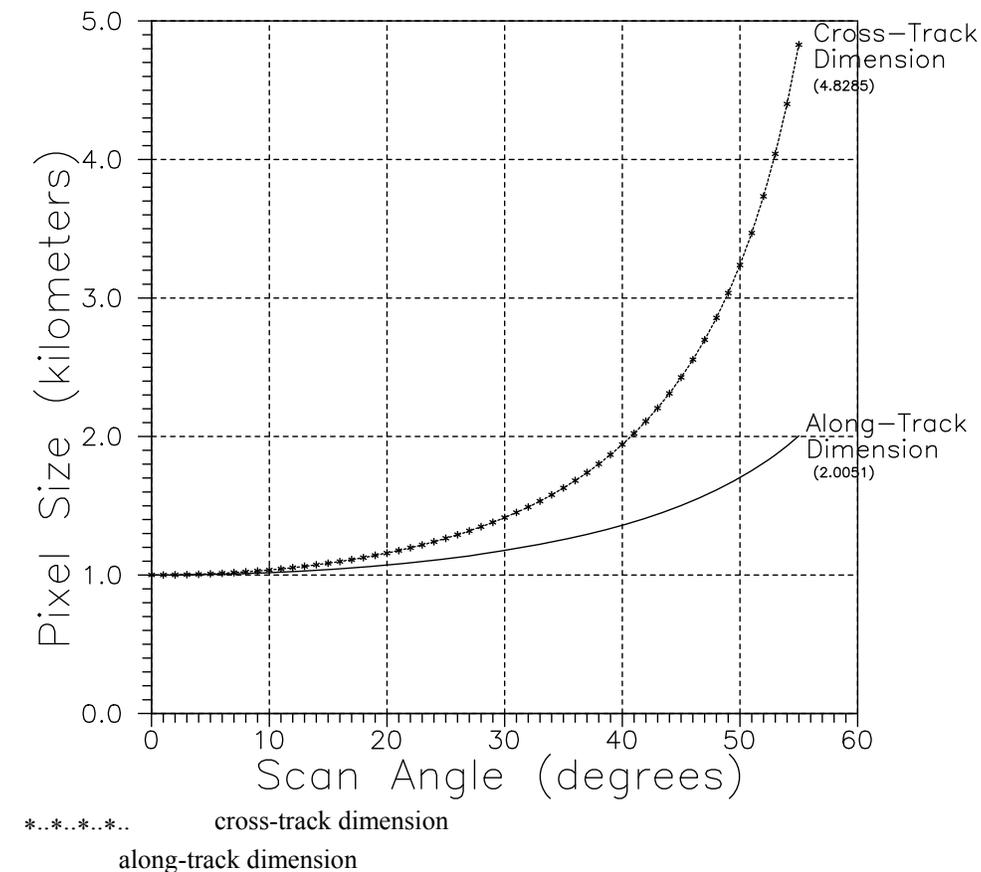


Figure 5. Growth of a 1-km resolution L2 pixel as a function of scan angle.

overlap at high scan angles – meaning that the L2 pixels used to compute L3 daily statistics in these “orbit edge” regions are not as statistically independent as those computed with near nadir view pixels. So not only are there fewer L2 pixels to average at high MODIS scan angles, but they are less independent. A third impact is these high scan angle L2 pixels may be less reliable in general as they are viewed by the sensor through more atmosphere, which complicates the retrieval process. A fourth impact is that they could have more inherent geolocation error depending on the local terrain.

Figure 6 shows the predicted orbital track (white lines) for the MODIS Terra instrument for a given day in 2007. L3 gridding issues at the poles (due to very small (area-wise) L3 grid boxes) are exacerbated by MODIS orbital tracks that go no further north or south than roughly 82° – which means the pixels at the poles are always viewed at higher scan angles. Also there are 16 overlapping orbits near 82° latitude (each one roughly 98

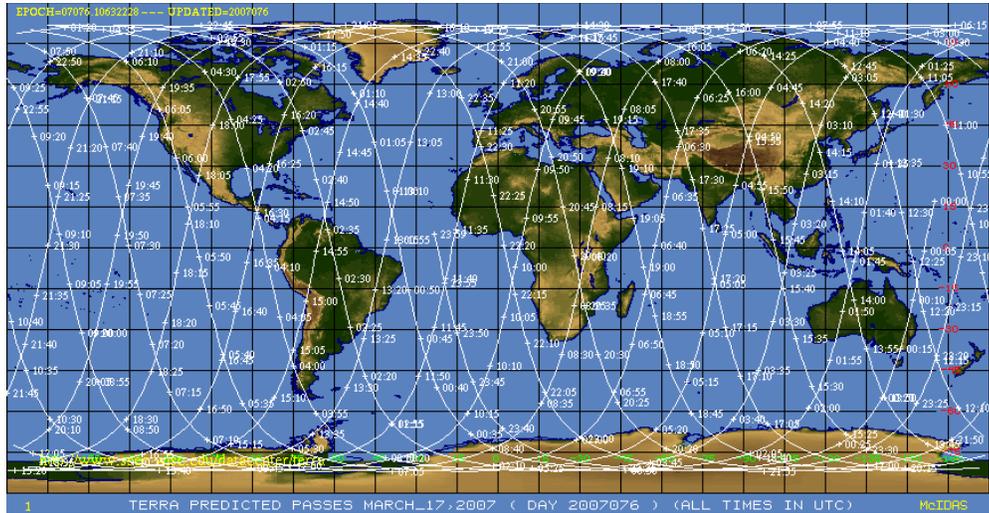


Figure 6. Predicted Terra MODIS orbit tracks for a particular day (March 17, 2007).

minutes apart) that cause “time averaging” to occur for daily statistics computed poleward of about 77° (that is, they tend to be daily average statistics), in contrast to those at mid-latitudes (that typically can be pinned down to within 20 minutes of a MODIS instrument overpass).

In order to help visualize a typical distribution of L2 pixel counts that are used to compute L3 daily statistics, a typical daily pixel count image is shown in Figure 7. This figure shows the number of L2 pixels that typically go into the computation of statistics in each L3 $1^\circ \times 1^\circ$ grid cell for all (daytime only) aerosol-related (10 km resolution at L2) SDSs.

For 5 km (native or sampled) L2 products, the distribution would be similar; however the counts would be roughly 4 times higher. Missing data that include orbital gap regions and the solar zenith angle cut-off region toward high southern latitudes (since this is a daytime only product) is shown by the pure black color.

It is clear from Figure 7 that over 120 L2 pixels were used to compute L3 statistics for this SDS near nadir (the bright red color at mid latitudes) but fewer than 15 L2 pixels were used near the edge of the scan (the light violet color next to the black orbital gap regions at mid-latitudes). Pixel counts for 10 km input products generally range from 0 to 121 (11×11), and pixel counts for 5 km input products generally range from 0 to 484 (22×22) pixels. One-degree L3 grid cell “pixel counts” can exceed these expected maximum numbers where overlapping orbits and near nadir views intersect (typically around 82° latitude). The reduction of counts in the top few rows of latitudes (82° to 90° N) in Figure 7 is due to the reduction of L2 pixel counts due to high view angles (that make the L2 pixels larger) and therefore fewer fall into a given (shrinking, due to Earth geometry) polar L3 grid cell since one L2 pixel can only be assigned to one L3 grid cell.

So for the first and last few latitude bins (88° to 90° latitude), one has the dual drawback of (i) very small (area-wise) L3 grid cells containing small populations of relatively large input L2 pixels that overlap spatially (due to view-angle distortion) in a single orbital pass, and (ii) multiple orbital passes (up to 16 on a given day) that take measurements of the region, which causes the L3 Daily data to be more representative of a daily

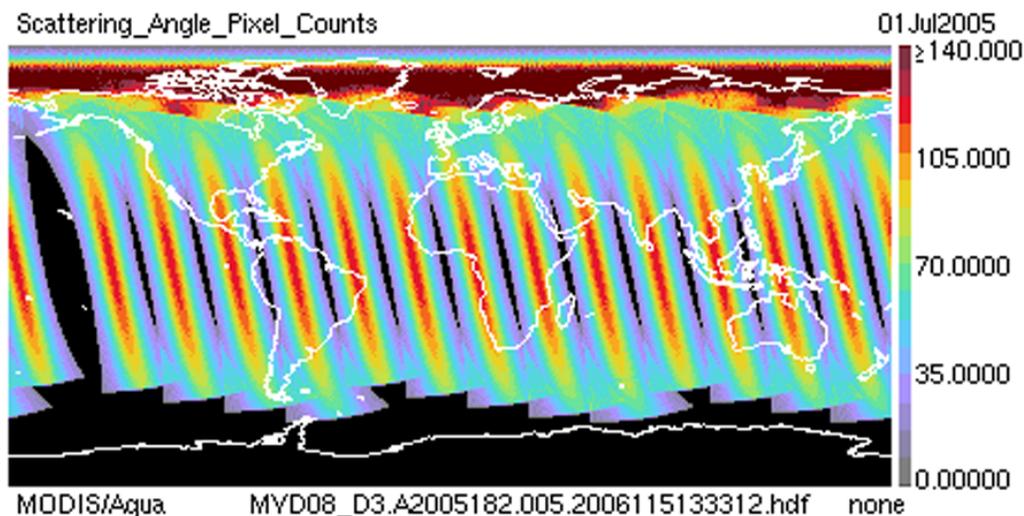


Figure 7. Typical global distribution of daily pixel counts (i.e., the number of L2 data pixels per $1^\circ \times 1^\circ$ L3 grid cell) for 10-km resolution L2 input products (e.g., aerosol derived products). Note this is a daytime (ascending node) only product.

average instead of a single overpass snapshot.

One final property of L3 gridding users should be aware of is, in order to simplify the L3 operational production software, each L2 grid cell is “located” or placed in the L3 grid cell where its center latitude/longitude falls within the L3 grid cell. In cases where a larger (view-angle distorted) L2 pixel actually covers (or intersects) more than one $1^\circ \times 1^\circ$ L3 grid cell, (a situation that typically occurs at high L2 scan angles located over higher latitudes), the L2 pixel is still only assigned (placed) in the single L3 grid cell where the geolocated center point of the L2 pixel is located. Although this shortcoming is relatively minor in the operational L3 software due to the relatively large $1^\circ \times 1^\circ$ L3 grid size and the much smaller L2 pixel size (5 or 10 km at nadir view), when the code is used in research mode to produce products at higher resolution ($0.1^\circ \times 0.1^\circ$ for example), it can produce unwanted effects where unnatural “gaps” are introduced between adjacent L2 pixels, as illustrated in Figure 8.

Finally it should be noted that most of these pixel count variation issues are mitigated in the multiday (eight-day and monthly) products since the orbital gap positions shift from day to day, which allows data from subsequent days to fill in the previous day’s gaps. This will tend to provide complete global coverage over longer time periods (8 days or so).

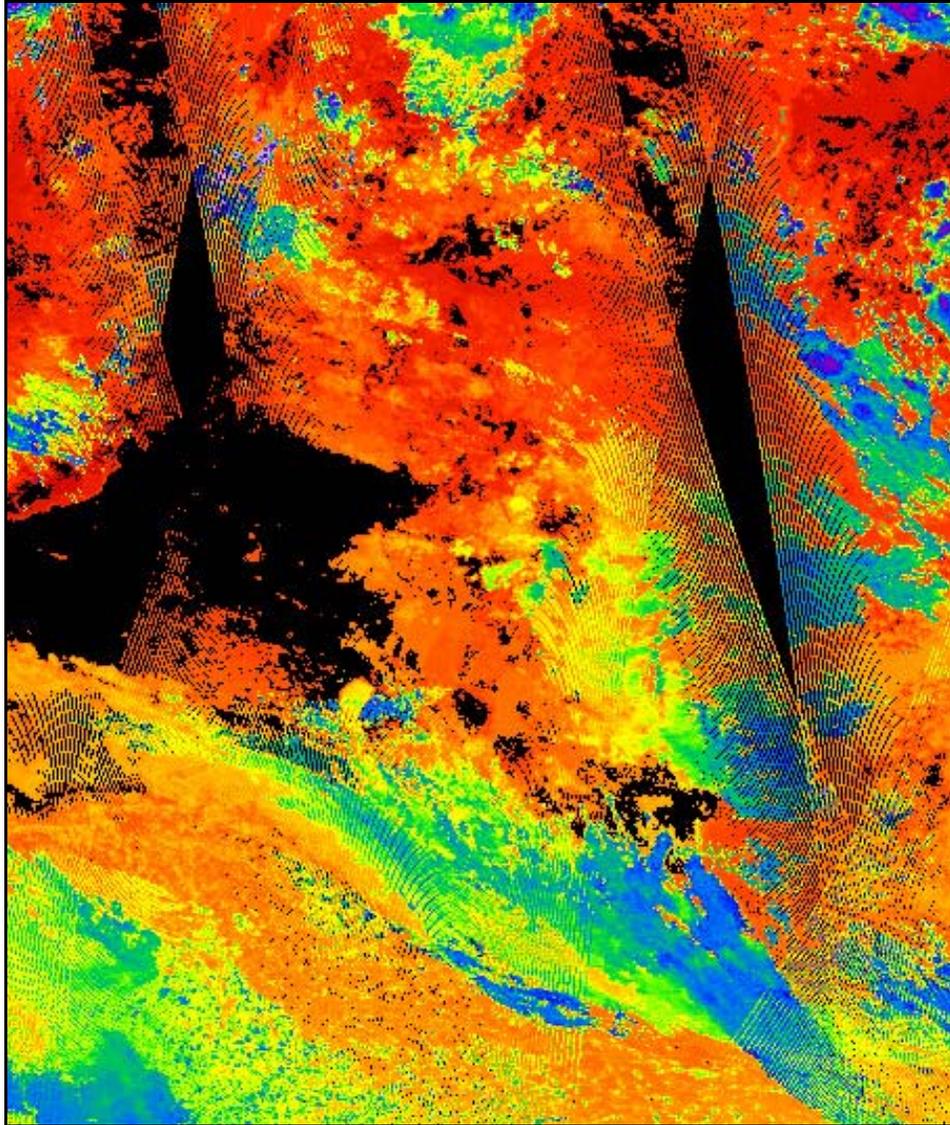


Figure 8. A section of a L3 cloud top temperature image produced for L3 research purposes at $0.1^\circ \times 0.1^\circ$ resolution clearly showing the pixel binning limitation in the operational L3 code for grids finer than 1° (the grid currently implemented in operations).

2.1. Computation of geolocation in L2 atmosphere products

MODIS Atmosphere Level 2 (L2) HDF files always have geolocation arrays stored at either 5-km or 10-km resolution. The Aerosol (04_L2) product contains 10-km resolution geolocation and 10-km retrieval data. All other Atmosphere products used as input into L3 (Water Vapor (05_L2), Cloud (06_L2), Profiles (07_L2)) contain 5-km resolution geolocation and either 1-km or 5-km retrieval data.

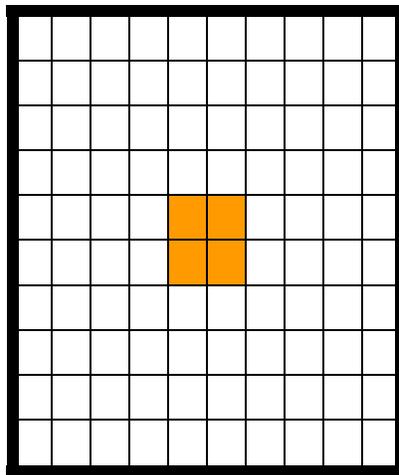


Figure 9. A depiction of what 1 km pixels (orange) are used to compute 10-km resolution L2 geolocation. Geolocation is computed from the 4 central 1-km grid boxes.

Retrieval data for **10-km resolution L2 products** like aerosol properties are generated from 10×10 1-km L1B input pixels. The geolocation for these 10×10 km resolution products is **computed** by averaging the geolocation for the four central (column, row) pairs: (5,5), (5,6), (6,5), (6,6), as shown in Figure 9.

Retrieval data for **5-km resolution L2 data products** is generated from 5×5 1-km L1B input pixels. Retrieval data for **1-km resolution L2 data products** is computed at the native 1 km L1B input pixel resolution. For L2 HDF files containing both of these cases, the geolocation is stored at 5-km resolution. The 5×5 km geolocation is **copied** from center (column, row) pair: (3,3), as shown in Figure 10.

Note that the 2×2 5-km grid area (the outer boundary in Figure 10) matches the 10 km grid area (the outer boundary in Figure 9) exactly.

2.1.1. *Dropped pixels in L2 products*

One additional characteristic should be noted here. A MODIS scan is 1354 1-km pixels in width. For 5 and 10 km resolution L2 products, the last 4 pixels in each MODIS scan are effectively dropped when performing retrievals (a pre-launch decision made by the L2 development teams). For example, for 5 km cloud top property parameters, no retrieval is made for the last 4 (hanging) pixels in each scan line. This also impacts the

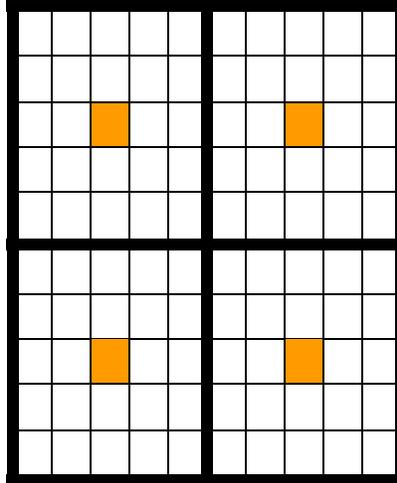


Figure 10. A depiction of what 1 km pixels (orange) are used for 5-km resolution L2 geolocation. Geolocation is copied from the center of the 1-km grid.

geolocation (latitude and longitude) arrays where there is not geolocation computed or stored for the last 4 pixels. To make this clearer, even though there are 1354 1-km pixels in a scanline, there are only 270 5-km cross-scan geolocation grid points and data retrievals in 5-km L2 products (and 135 geolocation grid points and data retrievals in 10-km L2 products). Therefore, in all 5-km and 10-km L2 products (and L3 products derived from those L2 products), a very tiny amount of potential data at the far edge of each MODIS swath is not used, however due to the $1^\circ \times 1^\circ$ averaging in L3, this is more of an issue in 5-km and 10-km L2 products than in $1^\circ \times 1^\circ$ L3.

2.2. Sampling

Sampling of L2 input data for computation of L3 statistics is always performed whenever the input L2 pixel resolution is at a different (finer) resolution than the stored geolocation within the L2 input product file. This occurs in all parameters noted by the red boxes in the Table 1.

So for cases where L2 data are at 1-km resolution and geolocation is stored at 5-km, the L3 software samples the 1-km L2 data every 3rd column and every 3rd row. This ensures the L2 data pixel matches exactly the stored geolocation. (Refer back to Figure 10 and note the orange-colored 1 km pixel.) Note that starting in collection 5, this sampling

Table 1. L2 parameters that are impacted by sampling (note that counts are approximated for nadir view at the Equator).

Derived from Product Family	L2 Data Resolution	L2 Geolocation Resolution	L2 Input Pixel Sampling (Maximum) per 1° L3 Grid Box (Equator)	Impact
Aerosol 04	10 km	10 km	121 out of 121	No
Water Vapor 05	1 km	5 km	484 out of 12,231	Yes
Cirrus Detection 06_CD	1 km	5 km	484 out of 12,231	Yes
Cloud Top Properties 06_CT	5 km	5 km	484 out of 484	No
Cloud Optical Properties 06_OD	1 km	5 km	484 out of 12,321	Yes
Atmosphere Profile 07	5 km	5 km	484 out of 484	No

convention was changed (see section 2.2.1 for details).

There were three reasons why sampling was done in the L3 production: (i) it removed the need to compute geolocation for all (input) 1-km L2 pixels, greatly simplifying the L3 operational software, (ii) fewer L2 pixels needed to be read and operated on, reducing the operational run and CPU time, and (iii) it was found that sampling causes minimal impact of computed statistics at the operational L3 grid resolution of $1^\circ \times 1^\circ$.

2.2.1. Sampling modification to avoid dead detectors

A software modification was introduced in Collection 5 (the latest MODIS Atmosphere data version available as of calendar year 2008), which offset slightly the L2 data point sampled. This patch **only** impacts L3 data that are derived from **sampled** L2 data and includes Water Vapor (05_L2), Cirrus Detection and Cloud Optical Property (06_L2) parameters (see Table 1), all of which are computed and stored at 1 km resolution in the L2 files but are sampled at 5 km resolution for computation of L3 gridded statistics.

In these cases, the geolocation matches the 1-km area corresponding to the orange square in Figure 11. This corresponds to the 1-km pixel located in column 3, row 3. However, the actual L2 data are sampled from the blue square (1 km pixel located in col-

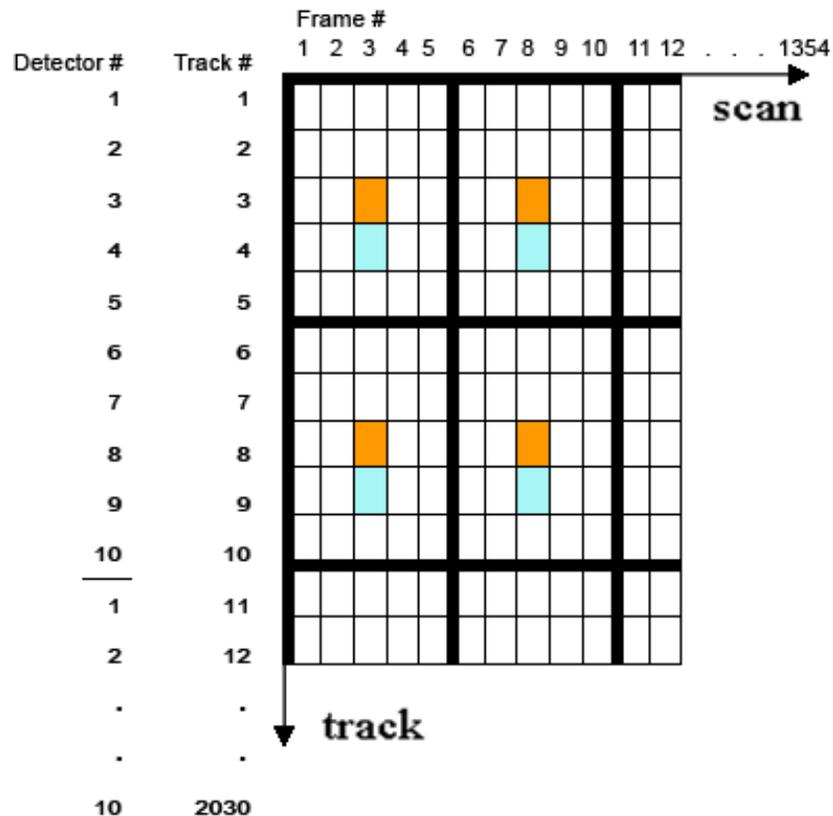


Figure 11. A shift in the L2 “sampling pixel” for 5 km L2 input products was implemented to prevent dead detectors from causing missing L3 data in Collection 5 (Orange = Collection 4 sampling at the center of the 5×5 (column 3, row 3), Blue = Collection 5 Sampling at (column 3, row 4) slightly off center of the 5×5 grid box. In other words, in Collection 4, L2 pixels from detectors 3 & 8 were sampled; in Collection 5, L2 pixels from detectors 4 & 9 were sampled.

umn 3, row 4). In other words, L2 data from MODIS detectors 4 and 9 are sampled (instead of detectors 3 and 8) in the 10-detector set. This patch was necessary because in some critical MODIS spectral bands on the Aqua platform, detectors 3 and 8 both died (that is, always contained missing data); so a shift to adjacent detectors was necessary to prevent many subsampled products from being completely missing (fill values) in L3.

The choice of which alternate pair of detectors to pick [(1 & 6), (2 & 7), (4 & 9), or (5 & 10)] was made to both minimize errors in geolocation as well as inherent detector errors. It turned out that detectors 4 and 9 had the dual benefit of: (i) being immediately adjacent to the geolocation pixel so only a 1 km geolocation error was introduced, and (ii) both detectors 4 and 9 were well behaved and had small residual errors when com-

pared to averages taken over the entire 5×5 km area (Oreopoulos 2005).

Even though the change in start detector (from the 3rd to the 4th) was prompted by the failure of Aqua MODIS detectors 3 and 8 in band 6 (1.64 μm), the change was extended to Terra data as well after a study showed that Terra detector pairs 4 and 9 provided the most representative results over a 5 km grid cell (Oreopoulos 2005) and it was thought that matching the logic between Terra and Aqua versions of the L3 software was prudent.

3.0. Computation of L3 Daily Statistics

The quality of a Level 2 product can be (i) inherited from the L1B radiances, or (ii) associated with the retrieval process. The pixel-based L1B validity flags comprising information on dead and saturated detectors, calibration failure, etc., are examined by L2 algorithms for determination of the radiometric status of each pixel. This information can prevent further calculations from being performed if the “valid L1B input data” criteria are not met by the given algorithm. The granule-level (L1B) QA metadata provide summary information for valid and saturated Earth view observations, and can be useful in screening a granule of data. Details about MODIS L1B QA flags can be found on the MCST web site (www.mcst.ssai.biz/L1B/L1B_docs). The structure and information content of MODIS Atmosphere L2 runtime QA are detailed in the QA Plan available on the MODIS Atmosphere web site (modis-atmos.gsfc.nasa.gov/reference_atbd.php).

It should be noted that runtime QA flags are only found in Level 2 (L2) Atmosphere products. Level 3 (L3) Atmosphere products contain no runtime QA flags; however L2 runtime QA flags are used in L3 to compute (aggregate and QA weight) statistics in L3.

3.1. Aggregation and QA weighting of statistics

All L3 products (Daily, Eight-Day, and Monthly) make use of Aggregation and QA weighting capabilities. Aggregation and QA weighting information is read in from and based upon L2 (input) QA bit flag SDSs.

It should also be noted that L3 statistics can be both aggregated and QA weighted; however only a single aggregation and/or QA weighting can be performed for any given output L3 SDS (a limitation in the L3 operational software).

3.1.1. Aggregation of statistics based on physical properties

For some parameters, it is useful to aggregate results into L3 SDSs based on a physical characteristic of the parameter or of the scene. Aggregation refers to the ability to separate L2 input pixel information into various scientifically relevant categories such as liquid water clouds only, ice clouds only, daytime only, nighttime only, clear sky only, dusty scene only, smoky scene only, etc. This aggregation utilizes L2 “Runtime QA Flags” that are designed to convey information on retrieval processing path, input data source, scene characteristics, and the estimated quality of the physical parameters retrieved. In addition, this broad group of flags also includes Cloud Mask flags (initially derived at 1×1 km resolution) that may be recomputed at the spatial resolution of the L2 retrieval for the determination such scene characteristics as cloudy/clear, land surface type, sunglint, day/night, and snow/ice. In L3 these statistics are noted by a suffix to the SDS name (`_Liquid`, `_Dust`, etc.). Additional detail and documentation are always provided in the local attribute “long_name” attached to each L3 SDS within the HDF file. Table 2 lists the various L2 products that have this aggregation information available for use in L3.

3.1.2. QA weighting of statistics based on confidence

QA weighting refers to the ability to weight more heavily what are expected to be more reliable L2 input pixels in the computation of L3 statistics. There are four levels of “reliability” or “confidence” set by the L2 QA Confidence Flags. These four levels are: No Confidence or Fill (QA = 0), Marginal Confidence (QA = 1), Good Confidence (QA = 2), or Very Good Confidence (QA = 3). QA weighted statistics always have the identifying string “QA” somewhere in the Scientific Data Set (SDS) name (for example: “QA_Mean”). Additional detail on the flags used to perform the QA weighting can be found in the local attributes that begin with the string “QA.”

Only six of the derived L3 Daily statistics have the ability to be QA weighted: Mean, StandardDeviation, LogMean, LogStandardDeviation, MeanUncertainty, and LogMeanUncertainty. A seventh related statistic, the Confidence Histogram, does not actually “weight” the statistics by confidence, but does compute (sum) the counts of the various confidence categories so the relative populations of QA confidence categories can be analyzed.

QA Weighted Statistics are computed by weighting all L2 pixels by their QA confidence flag value given in Table 3, using the equation below:

$$\text{L3 QA Weighted Statistic} = \sum d_i w_i / \sum w_i, \quad (1)$$

where,

Table 2. L2 parameters that are aggregated in level-3.

Product Family	Aggregation Property	Aggregation Used ?
Aerosol 04_L2	Dust, Smoke, Sulfate	Yes
Water Vapor 05_L2	Clear, Cloudy	Yes
Cirrus Detection 06_L2 (CD)	n/a	No
Cloud Top Properties 06_L2 (CT)	Day, Night	Yes
Cloud Optical Properties 06_L2 (OD)	Liquid Water, Ice, Undetermined, Combined Phase Clouds Single Layer (1L), Multi Layer (ML) Clouds	Yes
Atmosphere Profile 07_L2	n/a	No

L2 data value = d_i ,

L2 QA value = w_i (weight = 0, 1, 2, 3)

So while all *non-fill* QA = 0 pixels are included in regular statistics; they are screened (removed) from QA weighted statistics. For example: if a L3 1° grid cell had three L2 pixels to average, and one L2 pixel had a QA Confidence value of 3, one had a value of 1, and one had a value of 0; then L2 pixel with QA = 3 will have three times more weighting when computing the L3 gridded mean than the QA = 1 pixel, and the QA = 0 pixel would not be used at all in the QA weighted mean (it would however still be used in the regular mean, where no QA weights are applied). This technique allows for the creation of L3 (QA-weighted) statistics that can selectively exclude *no confidence* (or experimental) L2 results.

Confidence Histograms are computed by totaling the pixel counts (note these may be subsampled, see Table 1) in each L3 1° grid cell of a L2 parameter that fell in one of the four confidence categories described above.

Finally note that for regular (non QA-weighted) statistics, all L2 pixels are given equal weight in the L3 statistical computation; and QA = 0 (no confidence) non-fill L2 input pixels are included in the computation of statistics.

3.2. Types of daily statistics computed

A total of 14 different general types of statistics are computed in the daily product. These general statistical categories are Simple Statistics, QA-weighted Statistics, Fraction Statistics, Log Statistics, Uncertainty Statistics, Regression Statistics, Pixel Counts,

Table 3. The weighting given in L3 QA-weighted statistical computations to various L2 QA Confidence categories.

L2 Confidence QA Value	Weighting in L3 QA-Weighted Statistics
L2 pixels with QA = 3 (very good confidence)	are given a 3x weighting
L2 pixels with QA = 2 (good confidence)	are given a 2x weighting
L2 pixels with QA = 1 (marginal confidence)	are given a 1x weighting
L2 pixels with QA = 0 (no confidence)	are given a 0x weighting (i.e., they're excluded)

Confidence Histograms, Marginal Histograms, and Joint Histograms.

Statistics in the Daily file are always based on the set of L2 input pixels read from the four L2 input product files: Aerosol, Water Vapor, Cloud, and Atmospheric Profile.

Users should note that, in addition to regular simple statistics, L2 QA Confidence flags are also ignored in the L3 computation of pixel count, histogram, joint histogram, and regression statistics.

3.2.1. *Simple statistics*

- **Mean.** Statistics always have the Scientific Data Set (SDS) name suffix “_Mean” and are computed by taking an unweighted average of L2 pixels (sometimes sampled, see Table 1) within a given 1° L3 grid cell.
- **Standard Deviation.** Statistics always have the Scientific Data Set (SDS) name suffix “_Standard_Deviation” and are computed by calculating an unweighted standard deviation of all L2 pixels (sometimes sampled, see Table 1) within a given 1° L3 grid cell.
- **Minimum.** Statistics always have the Scientific Data Set (SDS) name suffix “_Minimum” and are computed by finding the minimum value of L2 pixels (sometimes sampled, see Table 1) within a given 1° L3 grid cell.
- **Maximum.** Statistics always have the Scientific Data Set (SDS) name suffix “_Maximum” and are computed by finding the maximum value of L2 pixels (sometimes sampled, see Table 1) within a given 1° L3 grid cell.

3.2.2. *QA-weighted simple statistics*

- **QA_Mean.** Level 2 Confidence QA Flags that indicate the quality of each L2 pixel are used to weight the pixels when computing the L3 mean (see Section 3.1.2 for details on QA-weighting).
- **QA_Standard_Deviation.** Level 2 Confidence QA Flags that indicate the quality of each L2 pixel are used to weight the pixels when computing the L3

standard deviation (see Section 3.1.2 for details on QA-weighting).

3.2.3. *Fraction statistics.* These statistics are only used for computing cloud fraction.

- **Fraction.** Estimates of cloud fraction based on L2 pixel data. Some fractions are computed from QA Flags that describe the cloudiness of the scene and some directly from the L2 Cloud Fraction SDS data.

3.2.4. *Pixel count statistics.* These statistics are only computed for some parameters. It is similar to a histogram computation except that instead of multiple bins there is only a single bin that covers the full range of all *non-fill* L2 data that are read in for each L3 $1^\circ \times 1^\circ$ grid cell.

- **Pixel Count.** The count of all non-fill L2 pixel data that are read in and used to compute statistics at L3.

3.2.5. *Logarithm statistics.* These statistics are only computed for cloud optical thickness (τ_c) parameters. Because of the curvature of cloud reflectance as a function of optical thickness, the mean optical thickness of an ensemble of pixels does not correspond to the mean reflectance (or albedo) of those pixels. However, the mean of $\log(\tau_c)$ approximates the radiatively-averaged optical thickness because reflectance plotted as a function of $\log(\tau_c)$ is linear over a wide range of optical thickness (excluding small and large values). That is, the mean of $\log(\tau_c)$ gives an optical thickness that approximately corresponds to the average reflectance of the pixels that comprise the mean. The accuracy of this approximation depends on the nature of the optical thickness probability density function (PDF). Studies on the validity of this approximate for MODIS scenes have been reported by Oreopoulos et al. (2007). A similar study on ice clouds by the same authors is ongoing.

- **Log_Mean.** L2 cloud optical thicknesses (τ_c) are converted to base 10 logs. Thus a τ_c of 100 would be converted to a value of 2.0, a τ_c of 10 would be

converted to a value of 1.0, a τ_c of 1.0 would be converted to a value of 0, a τ_c of 0.1 would be converted to a value of -1.0 , and finally a τ_c of 0.01 would be converted to a value of -2.0 . So the valid range of this statistic is -2.0 to 2.0 (corresponding to data values ranging from 0.01 to 100). Once the log values of the L2 input pixel data are calculated, a daily mean value of all the log values is computed.

- **Log_Standard_Deviation.** L2 cloud optical thicknesses (τ_c) are converted to base 10 logs and a standard deviation value of all the L2 log values is computed.
- **QA_Log_Mean.** L2 cloud optical thicknesses (τ_c) are converted to base 10 logs and a QA-weighted mean value of all the L2 log values is computed (see Section 3.1.2 for details on QA-weighting).
- **QA_Log_Standard_Deviation.** L2 cloud optical thicknesses (τ_c) are converted to base 10 logs and a QA-weighted standard deviation value of all the L2 log values is computed (see Section 3.1.2 for details on QA-weighting).

3.2.6. Uncertainty statistics. These statistics are only reported for a few selected Cloud Optical Property parameters. The uncertainty estimate accounts for three error sources only (instrument calibration/modeling error, surface albedo, atmospheric corrections), and as such should be considered an expected minimum uncertainty, i.e., the inclusion of additional (uncorrelated) error sources will increase the uncertainty. Daily uncertainty calculations assume all pixel-level error sources are correlated within a grid box. It should be noted that L2 uncertainties are reported in percentage (valid range 0 to 200%) – these are called relative uncertainties. In L3, uncertainties are reported as absolute uncertainties in the same units as the parameter whose uncertainty is being measured. To convert these back into relative uncertainties (%) one must divide the L3 uncertainty by the L3 mean value of the parameter in question. The conversion of relative to absolute uncertainties in the Daily (D3) file was done to make the computation of the multiday

(E3 and M3) uncertainties easier to add to the multiday production software.

- **Mean_Uncertainty.** An estimate of the absolute uncertainty that is derived from L2 pixel-level relative uncertainties.
- **QA_Mean_Uncertainty.** An estimate of the QA-weighted absolute uncertainty that is derived from L2 pixel-level relative uncertainties (see Section 3.1.2 for details on QA-weighting).

3.2.7. *Logarithm of uncertainty statistics.* An estimate of the log uncertainty that is derived from pixel-level uncertainties.

- **Log_Mean_Uncertainty.** An estimate of the log uncertainty that is derived from pixel-level uncertainties.
- **QA_Log_Mean_Uncertainty.** An estimate of the QA-weighted log uncertainty that is derived from pixel-level uncertainties (see Section 3.1.2 for details on QA-weighting).

3.2.8. *Histogram.* A distribution of L2 pixels.

- **Histogram.** A histogram that contains pixel counts showing the distribution of non-fill L2 pixels that went into the computation of L3 statistics for each L3 grid cell. Histogram bin boundaries are set by a local attribute attached to the histogram SDS. Note that these L2 count values are sampled totals for Water Vapor, Cirrus Detection, and Cloud Optical Property parameters (See Table 1). It should also be noted that the lowest (1st) histogram bin includes L2 data points that fall on either the lowest (1st) bin boundary or the 2nd bin boundary (exactly). All subsequent bins only contain points that fall on the higher bin boundary. Any L2 data point that falls outside the specified range of bin boundaries is not counted.

3.2.9. *Histogram of confidence.* A distribution of L2 pixel-level retrieval confidence.

- **Confidence_Histogram.** A histogram that contains pixel counts showing the number of Questionable (QA = 1), Good (QA = 2), Very Good (QA = 3), and Total Level-2 Input Pixels (# non-fill L2 Pixels) that went into the computation of L3 statistics for each L3 grid cell. Note that these values are sampled totals for Water Vapor, Cirrus Detection, and Cloud Optical Property parameters (See Table 1).

3.2.10. *Joint histogram.* A distribution of L2 pixels comparing one parameter against another.

- **Joint_Histogram.** A 2-dimensional histogram that contains pixel counts showing the distribution of non-fill L2 pixels when comparing one parameter against another. Joint Histogram bin boundaries are set by a local attribute attached to the Joint_Histogram SDS in question. Note that these pixel count values are sampled totals for Cloud Optical Property parameters (See Table 1). It should be noted that only a few Cloud parameters have Joint Histograms defined. The binning logic convention is as follows: the lowest (1st) histogram bin (for both parameters) includes L2 data pixels that fall on either the lowest (1st) bin boundary or the 2nd bin boundary (exactly). All subsequent bins only contain pixels that fall on the higher bin boundary. Any L2 pixels that fall outside the specified bin boundary range are not counted. Pixels for **both** parameters must be defined (non-fill) and within the specified range of bin boundaries for either pixel to be binned.

3.2.11. *Joint regression.* A regression fit of L2 pixels comparing one parameter against another. Note that these are only computed for a few Aerosol parameters.

- **Regression_Slope.** A computation of the regression slope that describes the linear fit distribution of non-fill L2 pixels when comparing one parameter against

another.

- **Regression_Intercept.** A computation of the regression intercept that describes the linear fit distribution of non-fill L2 pixels when comparing one parameter against another.
- **Regression_R_squared.** A computation of the regression r^2 that describes the linear fit distribution of non-fill L2 pixels when comparing one parameter against another.
- **Regression_Mean_Square_Error.** A computation of the regression mean squared error (MSE) that describes the linear fit distribution of non-fill L2 pixels when comparing one parameter against another.

3.3. Special handling logic in daily statistics

“Special handling” logic was introduced into the production software to add functionality or to mitigate problems in one or more parameters.

3.3.1. *Partial retrieval fix for cloud optical thickness*

A change in the L2 Cloud Optical Property retrieval algorithm allows for uncertain quality “partial retrievals” in certain marginal retrieval conditions. In this case a Cloud Optical Thickness retrieval is performed but no Cloud Effective Radius retrieval can be done. In this case, the L2 Cloud Optical Property development team requested that these partial retrievals be screened out from L3 statistical computations.

All Cloud Optical Thickness statistics have a unique local attribute called “Masked_With_QA_Usefulness_Flag,” which can be set to either “True” or “False.” When this attribute is set to true, the L3 algorithm logic prevents partial retrieval data from infiltrating the Cloud Optical Thickness statistics. This logic is applied to all Cloud Optical Thickness statistics.

4.0. Computation of L3 Multiday (8-Day & Monthly) Statistics

There are a number of general L3 Multiday algorithm characteristics:

- a) Only the L3 Daily files are used as input. (Greatly improves algorithm efficiency.)
- b) L3 Daily and Multiday HDF files use an identical grid, SDS dimensions, and histogram bin definitions. (This is an algorithm requirement.)
- c) There is no “valid range” check, only “Fill_Value” grid cells are universally excluded.
- d) One of three weighting schemes are used to compute Multiday statistics from Daily statistics.

4.1. Computational weighting schemes for multiday statistics

The weighting technique used for each SDS is documented in the L3 File Specification (File Spec), which can be found on the MODIS Atmosphere web site (modis-atmos.gsfc.nasa.gov/MOD08_M3/spec.html). This information is also attached to each SDS within the HDF file by a local attribute called “Weighting” – and if this is not set to “None,” there is an additional local attribute called “Weighted_Parameter_Data_Set,” which specifies which SDS in the Daily L3 (input) file is used to weight the daily grid cell statistics in the computation of the multiday statistics.

The various multiday weighting schemes used in L3 are:

- a) **Unweighted** (*a simple time-averaged mean, meaningful for computing temperature averages*).
- b) **Pixel-count weighted** (*a count-averaged mean, used to ensure computed means match means computed from histograms*).
- c) **Pixel-count weighted with pixel-count screen** (*special logic implemented for Aerosol-derived parameters to remove low confidence (low pixel count) daily grids usually occurring at the poleward terminator of the retrieval*

domain).

Table 4 summarizes what multiday weighting scheme is used for each broad set (grouping) of L3 parameters.

4.2. Types of multiday statistics computed

A total of 12 different general types of statistics are computed in the multiday (Eight Day or Monthly) products. They are grouped into broad categories and always computed in pre-determined sets.

Statistics in the Eight Day and Monthly files are always based on the set of L3 Daily pixels read from the L3 Daily input product files that cover the time period being summarized (8 days or 1 month).

4.2.1. Simple statistics based on the daily mean

Mean_Mean and Std_Deviation_Mean statistics can either be unweighted or weighted depending on a local attribute setting (set per instructions from the L2 algorithm team) in the HDF structure file. An unweighted statistic is computed by taking an average of all L3 daily values for a given $1^\circ \times 1^\circ$ L3 grid cell for the time period in question. Weighted statistics are computed by using information stored in an SDS local

Table 4. L3 Multiday weighting scheme used (generally) for categories of parameters.

Product Family	Multiday Weighting Scheme *	Detail
Aerosol 04_L2	Pixel Count Weighted with Pixel Count Screen	Each daily grid weighted by the pixel count & grids with counts < 6 are excluded
Water Vapor 05_L2	Pixel Count Weighted	Each daily grid weighted by the pixel count
Cirrus Detection 06_L2 (CD)	Unweighted	Each daily grid given the same weight
Cloud Top Properties 06_L2 (CT)	Unweighted	Each daily grid given the same weight
Cloud Optical Properties 06_L2 (OD)	Pixel Count Weighted	Each daily grid weighted by the pixel count (categorical, by phase)
Atmosphere Profile 07_L2	Pixel Count Weighted	Each daily grid weighted by the pixel count

attribute called “Weighted_Parameter_Data_Set,” typically a Daily Pixel Count or Fraction SDS. To determine the technique used to weight the daily grids, refer to the local attribute “Weighting” found in the file specification or a local attribute attached to each SDS in the HDF file.

- **Mean_Mean.** The Scientific Data Set (SDS) name suffix “_Mean_Mean” stands for “Mean of the Daily Mean.” This statistic is computed by averaging non-fill Daily Means from the L3 Daily files within the (eight day or monthly) time period being summarized. The averaging computation uses the weighting scheme specified in the local attribute “Weighting.” This can be set to “Unweighted,” “Pixel-Weighted,” or “Pixel_Weighted_Screen” (See 4.1 for details).
- **Mean_Std.** The Scientific Data Set (SDS) name suffix “_Mean_Std” stands for “Standard Deviation of the Daily Mean.” This statistic is computed by reading all the non-fill Daily Means from the L3 Daily files within the (eight day or monthly) time period being summarized and then computing their standard deviation.
- **Mean_Min.** The Scientific Data Set (SDS) name suffix “_Mean_Min” stands for “Minimum of the Daily Mean.” This statistic is computed by reading all the non-fill Daily Means from the L3 Daily files within the (eight day or monthly) time period being summarized and reporting the minimum value. There is never any weighting performed for this statistic.
- **Mean_Max.** The Scientific Data Set (SDS) name suffix “_Mean_Max” stands for “Maximum of the Daily Mean.” This statistic is computed by reading all the non-fill Daily Means from the L3 Daily files within the (eight day or monthly) time period being summarized and reporting the maximum value. There is never any weighting performed for this statistic.

4.2.2. *Simple statistics based on the daily standard deviation*

- **Std_Deviation_Mean.** The Scientific Data Set (SDS) name suffix “Std_Deviation_Mean” is a proxy for “Mean of the Daily Standard Deviation.” This statistic is computed by reading all the non-fill Daily Standard Deviations from the L3 Daily files within the (eight day or monthly) time period being summarized, then averaging the standard deviations using the same weighting as used for the Mean_Mean.

4.2.3. *QA-weighted statistics based on the daily mean*

All Mean and Standard Deviation statistics computed can either be unweighted or weighted depending on a local attribute setting in the HDF structure file. An unweighted statistic is computed by taking an average of all L3 daily values for a given 1° L3 grid cell for the time period in question. A weighted statistic is computed by using a “Weighted_Parameter_Data_Set,” typically a Pixel Count or Fraction SDS, to weight each of the Daily grid cell values. To determine the technique used to weight the daily grids, refer to the local attribute “Weighting” found in the file specification or a local attribute attached to each SDS in the HDF file.

- **QA_Mean_Mean.** The Scientific Data Set (SDS) name suffix “QA_Mean_Mean” stands for “Mean of the Daily QA-weighted Mean.” This statistic is computed by averaging all the non-fill Daily QA-weighted Mean grid cells from the L3 Daily files within the (eight day or monthly) time period being summarized using the same weighting used for the Mean_Mean.
- **QA_Mean_Std.** The Scientific Data Set (SDS) name suffix “QA_Mean_Std” stands for “Standard Deviation of the Daily QA-weighted Mean.” This statistic is computed by reading all the non-fill Daily QA-weighted Means from the L3 Daily files within the (eight day or monthly) time period being summarized, then computing their standard deviation.

- **QA_Mean_Min.** The Scientific Data Set (SDS) name suffix “QA_Mean_Min” stands for “Minimum of the Daily QA-weighted Mean.” This statistic is computed by reading all the non-fill Daily QA-weighted Means from the L3 Daily files within the (eight day or monthly) time period being summarized and reporting the minimum value. There is no additional weighting performed for this statistic.
- **QA_Mean_Max.** The Scientific Data Set (SDS) name suffix “QA_Mean_Max” stands for “Maximum of the Daily QA-weighted Mean.” This statistic is computed by reading all the non-fill Daily QA-weighted Means from the L3 Daily files within the (eight day or monthly) time period being summarized and reporting the maximum value. There is no additional weighting performed for this statistic.
- **QA_Std_Deviation_Mean.** The Scientific Data Set (SDS) name suffix “QA_Std_Deviation_Mean” is a proxy for “Mean of the Daily QA-weighted Standard Deviation.” This statistic is computed by reading all the non-fill Daily QA-weighted Standard Deviations from the L3 Daily files within the (eight day or monthly) time period being summarized, then averaging the standard deviations using the same weighting as used for the Mean_Mean.

4.2.4. *Fraction statistics.* These statistics are only used for computing cloud fraction.

- **FMean.** Mean of the Daily Cloud Fraction.
- **FStd.** Standard Deviation of the Daily Cloud Fraction.

4.2.5. *Pixel count statistics.* These statistics are only computed for some parameters. It is similar to a histogram computation except that instead of multiple bins there is only a single bin that covers the full range of all non-fill L2 data that are read in for each L3 $1^\circ \times 1^\circ$ grid cell.

- **Pixel Count.** The count of all non-fill L2 pixel data that are read in and used to

compute statistics at L3. This is computed by simply summing the Daily Pixel Count SDS.

4.2.6 *Logarithm statistics.* These statistics are only computed for cloud optical thickness parameters. The mean statistic is typically weighted using the “Pixel_Weighted” scheme, which weights each daily mean by a pixel count SDS (this allows the multiday means to match those computed from histograms).

- **Log_Mean_Mean.** Mean of the Daily Log Mean.
- **Log_Mean_Std.** Standard Deviation of the Daily Log Mean.
- **Log_Mean_Min.** Minimum of the Daily Log Mean.
- **Log_Mean_Max.** Maximum of the Daily Log Mean.
- **Log_Std_Deviation_Mean.** Mean of the Daily Log Standard Deviation.
- **QA_Log_Mean_Mean.** Mean of the Daily QA Log Mean.
- **QA_Log_Mean_Std.** Standard Deviation of the Daily QA Log Mean.
- **QA_Log_Mean_Min.** Minimum of the Daily QA Log Mean.
- **QA_Log_Mean_Max.** Maximum of the Daily QA Log Mean.
- **QA_Log_Std_Deviation_Mean.** Mean of the Daily QA Log Standard Deviation.

4.2.7. *Uncertainty statistics.* These statistics are only reported for a few selected Cloud Optical Property parameters. Monthly uncertainty calculations are derived from daily uncertainties; calculations assume that individual daily uncertainties are uncorrelated with respect to each other. The equation to compute multiday uncertainties is as follows.

$$\text{Uncertainty} = \sqrt{\sum_i w_i^2 \sigma_i^2} / \sqrt{\sum_i w_i^2}, \quad (2)$$

where,

σ_i = DailyUncertainties

w_i = DailyPixelCountWeights

It should be noted that L2 uncertainties are reported in percentage (valid range 0 to 200%) – these are called relative uncertainties. In L3, uncertainties are reported as absolute uncertainties in the same units as the parameter whose uncertainty is being measured. To convert these back into relative uncertainties (%) one must divide the L3 uncertainty by the L3 mean value of the parameter in question. The conversion of relative to absolute uncertainties in the Daily (D3) file was done to make the computation of the multiday (E3 and M3) uncertainties easier to add to the multiday production software.

- **Mean_Uncertainty.** A multiday absolute uncertainty estimate derived from the daily absolute uncertainty estimate.
- **QA_Mean_Uncertainty.** A multiday QA-weighted absolute uncertainty estimate derived from the daily QA-weighted absolute uncertainty estimate.

4.2.8. *Logarithm of uncertainty statistics.* An estimate of the log uncertainty that is derived from pixel-level uncertainties.

- **Log_Mean_Uncertainty.** A multiday absolute log uncertainty estimate derived from the daily absolute log uncertainty estimate.
- **QA_Log_Mean_Uncertainty.** A multiday QA-weighted absolute log uncertainty estimate derived from the daily QA-weighted absolute log uncertainty estimate.

4.2.9. *Histogram.* A distribution of L2 pixels.

- **Histogram.** A histogram that contains pixel counts showing the distribution of non-fill L2 Pixels that went into the computation of L3 statistics for each L3 grid cell. In the eight day and monthly products, this is computed by simply summing the daily counts in each histogram bin. Histogram bin boundaries are set by a local attribute attached to the histogram SDS. Note that these L2 count values are sampled totals for Water Vapor, Cirrus Detection, and Cloud Optical Property parameters (See Table 1). It should also be noted that the lowest (1st) histogram

bin includes L2 data points that fall on either the lowest (1st) bin boundary or the 2nd bin boundary (exactly). All subsequent bins only contain points that fall on the higher bin boundary. Any L2 data point that falls outside the specified range of bin boundaries is not counted.

4.2.10. *Histogram of confidence.* A distribution of L2 pixel-level retrieval confidence.

- **Confidence_Histogram.** A histogram that contains pixel counts showing the number of Questionable (QA = 1), Good (QA = 2), Very Good (QA = 3), and Total Level-2 Input Pixels (# non-fill L2 Pixels) that went into the computation of L3 statistics for each L3 grid cell. Note that these values are sampled totals for Water Vapor, Cirrus Detection, and Cloud Optical Property parameters (See Table 1). In the eight day and monthly products, this is computed by simply summing the daily counts in each bin.

4.2.11. *Joint histogram.* A distribution of L2 pixels comparing one parameter against another.

- **Joint_Histogram.** A 2-dimensional histogram that contains pixel counts showing the distribution of non-fill L2 Pixels comparing one parameter against another. It should be noted that only a few Cloud parameters have Joint Histograms defined. In the eight day and monthly products, this is computed by simply summing the daily counts in each joint histogram bin. Histogram bin boundaries are set by a local attribute attached to the histogram SDS. Note that these L2 count values are sampled totals for Cloud Optical Property parameters (See Table 1). It should also be noted that the lowest (1st) histogram bin includes L2 data points that fall on either the lowest (1st) bin boundary or the 2nd bin boundary (exactly). All subsequent bins only contain points that fall on the higher bin boundary. Any L2 data point that falls outside the specified range of bin boundaries are not counted.

4.2.12. *Joint regression.* A regression fit of L2 pixels comparing one parameter against another. Note that these are only computed for a few Aerosol parameters.

- **Regression_Slope.** An unweighted mean of the Daily regression slope that describes the linear fit distribution of L2 pixels when comparing one parameter against another.
- **Regression_Intercept.** An unweighted mean of the Daily regression intercept that describes the linear fit distribution of non-fill L2 pixels when comparing one parameter against another.
- **Regression_R_squared.** An unweighted mean of the Daily regression r^2 that describes the linear fit distribution of non-fill L2 pixels when comparing one parameter against another.
- **Regression_Mean_Square_Error.** An unweighted mean of the Daily regression mean squared error (MSE) that describes the linear fit distribution of non-fill L2 pixels when comparing one parameter against another.

5.0. Reading and Unpacking HDF File Data

Hierarchical Data Format (HDF) is the standard data format for all NASA Earth Observing System (EOS) data products. HDF is a multi-object file format developed by the HDF Group at the National Center for Supercomputing Applications (NCSA) at the University of Illinois (www.ncsa.uiuc.edu).

The HDF Group developed HDF to assist users in the transfer and manipulation of scientific data across diverse operating systems and computer platforms, using Fortran and C calling interfaces and utilities. HDF supports a variety of data types: n-Dimensional scientific data arrays, tables, text annotations, several types of raster images and their associated color palettes, and metadata. The HDF library contains interfaces for storing and retrieving these data types in either compressed or uncompressed formats.

For each data object in an HDF file, predefined tags identify the type, amount, and

dimensions of the data, and the file location of various objects. The self-describing capability of HDF files helps users to fully understand the file’s structure and contents from the information stored in the file itself. A program interprets and identifies tag types in an HDF file and processes the corresponding data.

Because many Earth science data structures need to be geolocated, the HDF Group developed the HDF-EOS format with additional conventions and data types for HDF files. HDF-EOS supports three geospatial data types: grid, point, and swath, providing uniform access to diverse data types in a geospatial context. All MODIS Atmosphere Level 3 HDF files use the grid geospatial data type. The HDF-EOS software library allows a user to query or subset of the contents of a file by Earth coordinates and time if there is a spatial dimension in the data. Tools that process standard HDF files also read HDF-EOS files; however, standard HDF library calls cannot access geolocation data, time data, and product metadata as easily as with HDF-EOS library calls.

Finally, it’s important to note that all MODIS data are archived using HDF-EOS4, which is based on HDF4. This file format is not compatible with the newer HDF5 format; and there are no plans to ever switch to HDF5 for MODIS data.

5.1. Descaling the SDS data

The local attributes “scale_factor” and “add_offset,” attached to each and every SDS, are used for the conversion of stored (packed) SDS integer data to geophysical floating point numbers. The formula to descale the data follows conventional HDF usage (see HDF User’s Guide):

$$\text{Geophysical Float Value} = \text{scale_factor} * (\text{Stored Integer} - \text{add_offset}). \quad (3)$$

It is probably confusing to most users that the HDF convention calls the offset “add_offset” even though it’s subtracted from the stored integer when unpacking the data. It seems likely that this terminology originated from the programmer’s, and not the end user’s, perspective, since to pack the data the offset is added.

The units of the unpacked geophysical floating point value are indicated by the “units” local attribute that is also provided with each SDS.

The “valid_range” local attribute applies to the packed data (before de-scaling). The two valid range values given are the expected low and high values of valid (non-fill) packed data. Note that **no** valid range screening on the input L2 data or the output L3 data is performed. The reason for this is sometimes absolute valid ranges are difficult to determine in advance and the algorithm developers wanted to avoid the potential loss of good data. Therefore users should not be surprised to find non-fill data points that fall outside the documented valid range; however it should raise a flag for the user to make sure they are reading the data correctly.

5.2. Definitions of local attributes

Attached to each Scientific Data Set (SDS) within an HDF file are a number of local attributes. These local attributes serve as a key to unpack the data, drive the logic in the L3 software, and finally provide documentation for end users.

5.2.1. Local attributes used in all L3 files. There is a set of local attributes that can be used in all L3 files (Daily, Eight-Day, and Monthly). Note that not all of these are attached to every SDS (only the local attributes necessary are attached).

- **long_name.** A character string description of the SDS.
- **units.** Units of the unpacked data (if there are no units, “none” is specified).
- **valid_range.** An array of two numbers that describe the expected low and high values (valid range) of the packed data (before applying the scale and offset); ignoring the Fill Value.
- **FillValue.** The value of missing or fill data in the packed SDS.
- **scale_factor.** The scaling factor used to unpack the data (see Eq. (2)).
- **add_offset.** The offset used to unpack the data (see Eq. (2)).
- **Statistic_Type.** A fixed string that specifies the statistical group. (This attribute

drives the logic in the operational production program).

- **Quality_Assurance_Data_Set.** This specifies which L2 SDS to read to obtain the L2 QA. (In E3 and M3 HDF files this is for documentation only.)
- **QA_Byte.** Specifies which L2 SDS byte to read to obtain the L2 QA (in the E3 and M3 files this is for documentation only).
- **QA_Useful_Flag_Bit.** Specifies which L2 SDS bit to read to obtain the L2 QA Useful Flag bit. (In E3 and M3 HDF files this is for documentation only.)
- **QA_Value_Start_Bit.** Specifies which L2 SDS bit to begin reading to obtain the L2 QA Confidence Flag. (In E3 and M3 HDF files this is for documentation only.)
- **QA_Value_Num_Bits.** Specifies the number of L2 SDS bits to read to obtain the L2 QA Confidence Flag (In E3 and M3 HDF files this is for documentation only.)
- **Aggregation_Data_Set.** Specifies which L2 SDS to read to obtain the L2 Aggregation information. (In the E3 and M3 HDF files this is for documentation only.)
- **Aggregation_Byte.** Specifies which L2 SDS byte to read to obtain the L2 Aggregation information (in the E3 and M3 files this is for documentation only).
- **Aggregation_Value_Start_Bit.** Specifies which L2 SDS bit to begin reading to obtain the L2 Aggregation information. (In E3 and M3 HDF files this is for documentation only.)
- **Aggregation_Value_Num_Bits.** Specifies the number of L2 SDS bits to read to obtain the L2 Aggregation information. (In E3 and M3 HDF files this is for documentation only.)
- **Aggregation_Valid_Category_Values.** Specifies all the Aggregation values that are valid (not fill). (In E3 and M3 HDF files this is for documentation only.)
- **Aggregation_Category_Values.** Specifies all the Aggregation values that are to be used to aggregate the statistic. (In E3 and M3 HDF files this is for

documentation only.)

- **Level_2_Pixel_Values_Read_As.** A fixed string that tells the operational production program what the format is of the L2 input data (byte, integer, real).
- **Included_Level_2_Nighttime_Data.** A documentation string telling the user if the Statistic contains any nighttime data. (Set to True or False.)
- **Histogram_Bin_Boundaries.** An array of numbers describing the histogram bin boundaries of the primary parameter.
- **Joint_Parameter_Histogram_Bin_Boundaries.** An array of numbers describing the histogram bin boundaries of any secondary parameter. (Used for Joint Histogram statistics.)

5.2.2. *Local attributes used in the L3 daily file only.* There is a set of local attributes that are used only in the L3 Daily file.

- **Derived_From_Level_2_Data_Set.** L2 SDS read to obtain the input data.
- **Masked_With_QA_Usefulness_Flag.** Set to “True” or “False.” If True, the QA Usefulness Flag is used to screen input data. (This attribute is used to screen out partial cloud optical property retrievals from L3).
- **Wave_Band.** Used to specify what band to read when a L2 input parameter has more than 1 solution (wave_band).
- **Joint_Parameter_Wave_Band.** Used to specify what band to read when a L2 input parameter that is being used as a joint parameter in L3 has more than 1 solution (wave_band).

5.2.3. *Local attributes used in the L3 eight day and monthly file only.* There are a set of local attributes that are used only in the L3 Eight Day and Monthly file. These are:

- **Derived_From_Level_3_Data_Set.** Daily SDS read to obtain the input data.
- **Weighting.** Specifies what weighting scheme is used to average the Daily L3 pixels (see Section 4.1).

- **Weighted_Parameter_Data_Set.** Specifies what Daily SDS is used for weighting the Daily L3 pixels. This local attribute is used (specified) only when Weighting = “True.”

6.0. Using the MODIS-Atmosphere Web Site

A user-friendly web-based interface on the MODIS-Atmosphere web site can be used to view primary statistics in all the MODIS-Atmosphere L3 data products. Users may select any daily, 8-day, or monthly HDF files and any group of parameters to view (see Figure 12). Once you select the product group, date, and projection; select ‘Go’ to see the resulting images (see Figure 13).

For Daily and Eight day products, images can only be viewed in the native latitude-longitude (rectangular) grid. For Monthly browse images both the native equal-angle (latitude-longitude) grid, as well as an equal-area (Hammer-Aitoff) grid are available.

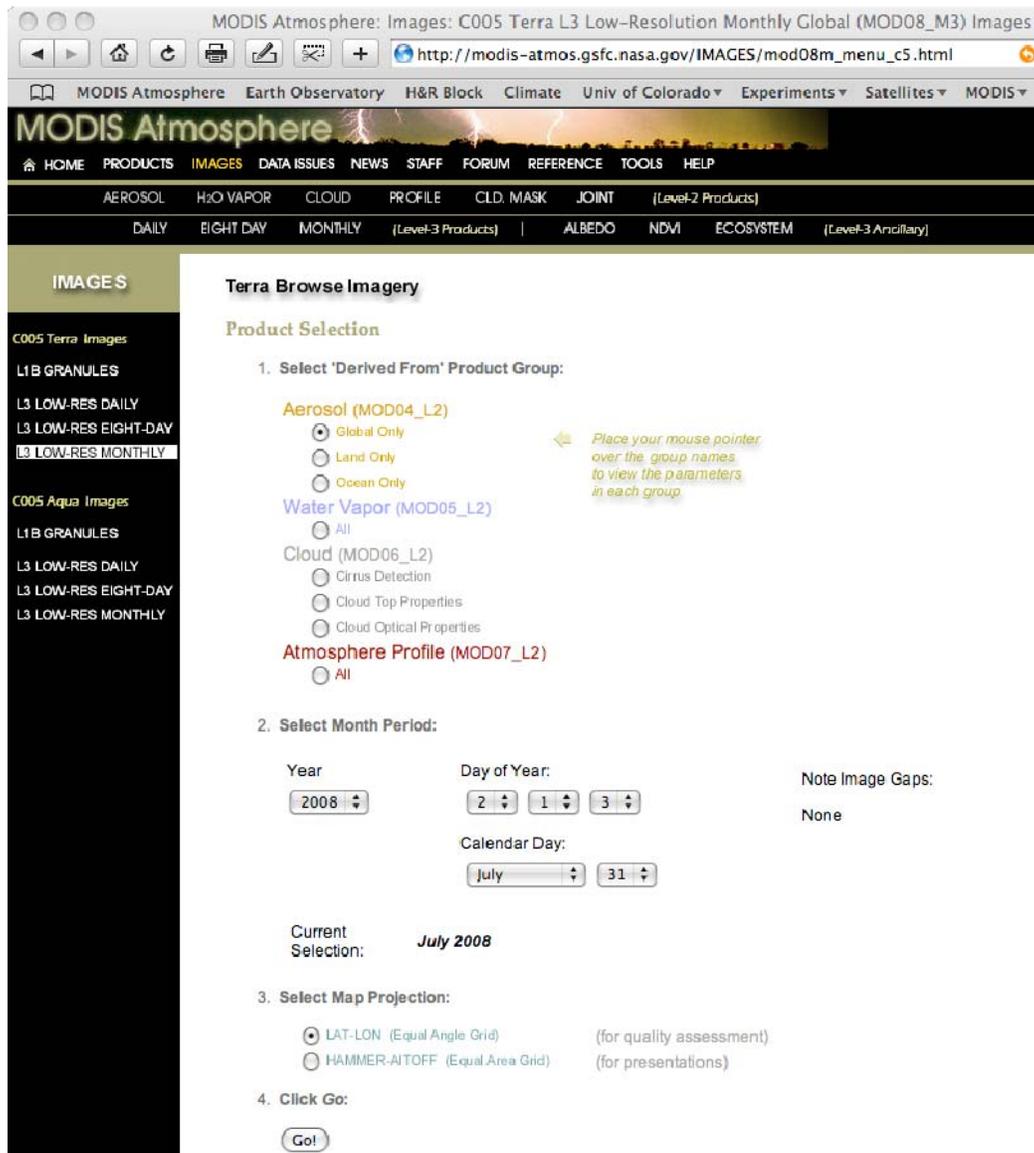


Figure 12. The web-based interface that allows viewing of primary statistical images in all L3 HDF files. A user selects a “derived-from” product group, the time period, and map projection.

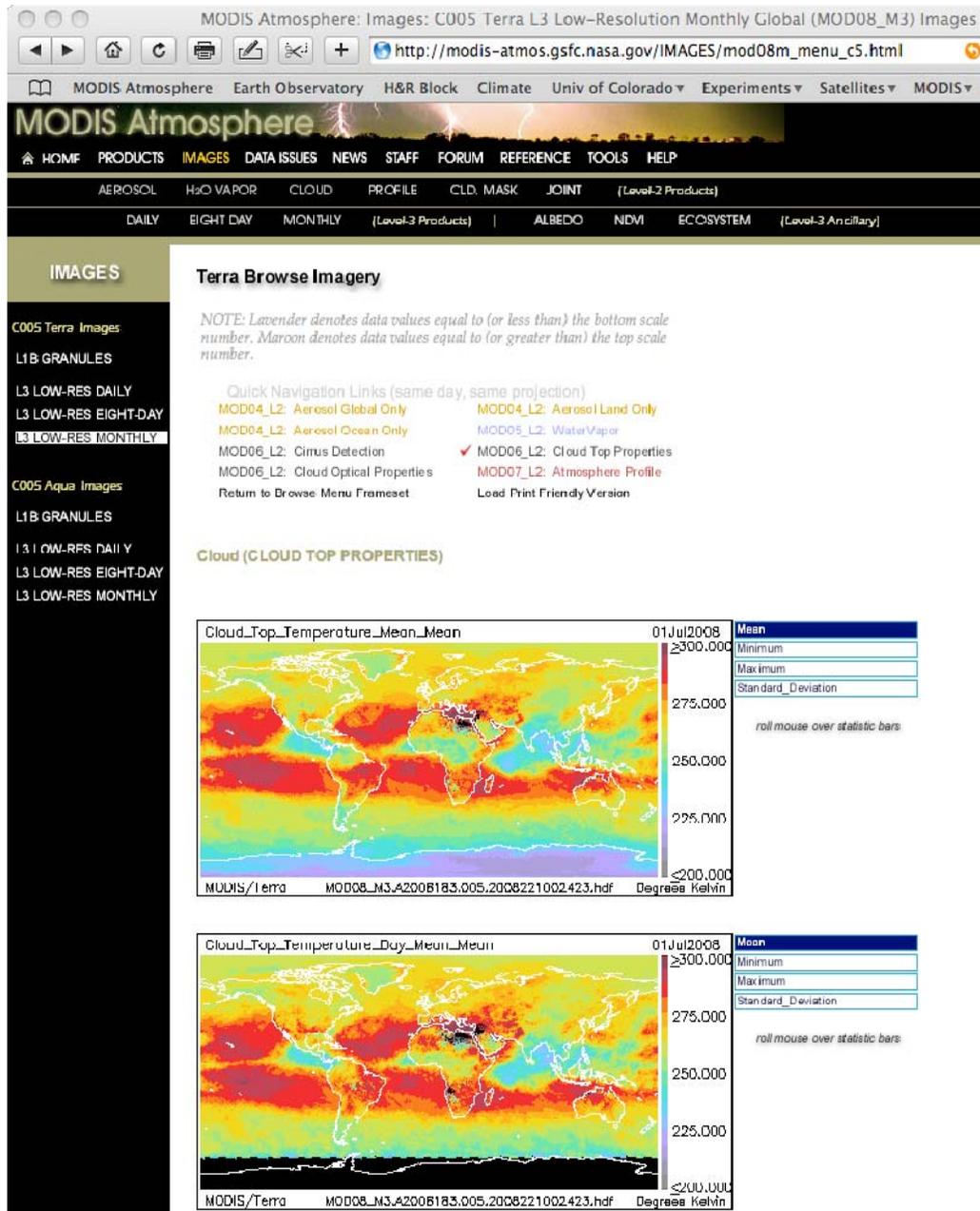


Figure 13. The L3 browse images showing the Mean image by default. Other statistical images can be viewed by using a mouse-over function implemented on the statistic bars to the right of each image. The color bar is a discrete modified rainbow and the data scale has been pre-determined to maximize detail.

Users should be aware that the Hammer-Aitoff (equal-area) version of the L3 browse image is derived directly from the Latitude-Longitude (equal-angle) image using a packaged Interactive Data Language (IDL) function. This function can cause some mild distortion and occasionally even loss of data while converting from one map projection to

another inregions (especially high latitudes) where there are sparse data (data gaps); so if one is performing any detailed quantitative analysis on SDSs that contain data gaps, the native Latitude-Longitude version is the preferred choice for viewing (cf. Figure 14).

A discretized modified-rainbow color bar is applied to each image and the data scale is optimized to maximize image detail. Fill (missing) data are always colored black. Multiple statistics (SDSs) for a single parameter can be viewed by using a built-in “mouse-over” functionality on the web page by rolling your mouse over the statistic “bars” to the right of each image (see Figure 13).

Finally there is a myriad of additional information that can be gleaned from the

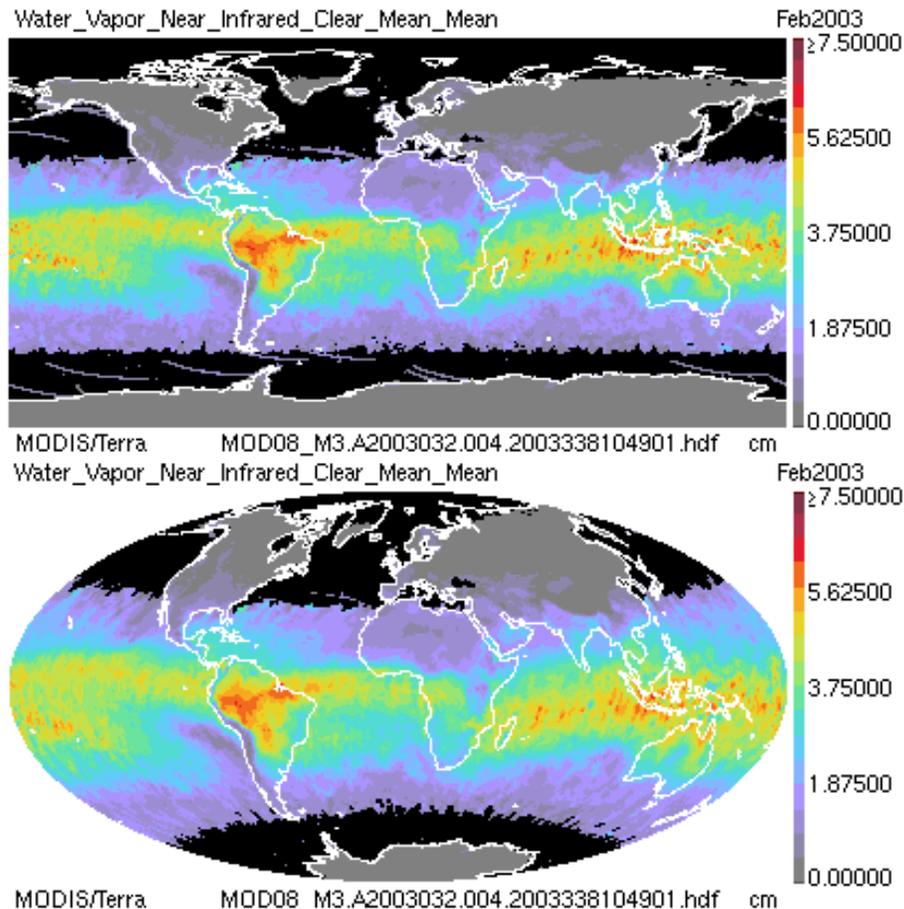


Figure 14. Identical Collection 004 data mapped on the native Latitude-Longitude grid (top) and remapped on the Hammer-Aitoff grid bottom. Note the distortion in sparse data regions especially at high latitudes on the Hammer grid (e.g., the disappearance of the bad data scans (purple streaks) in the southern Ocean); however in solid data regions (30°N to 30°S) the Hammer remapping shows little distortion.

MODIS-Atmosphere web site including a complete file specification, details on the format and content, modification history, and known problems. Also available are programs and tools that can read and image L3 HDF file SDSs.

7.0. Interpretation of Data: Frequently Asked Questions

Since launch, MODIS data users have asked the L3 development team and associated MODIS Atmosphere staff a number of questions. This section summarizes the most common questions and their answers.

7.1. How do the cloud mask and optical properties ‘cloud fractions’ differ?

In order for users to understand the meaning of cloud fraction in MODIS L3 products, an understanding of how cloud fraction is defined and computed in each L3 cloud fraction parameter is needed. From the simplest perspective, cloud fraction at L3 is the number of cloudy L2 pixels (that fit pre-determined criteria, such as daytime only or liquid water clouds only) divided by the total number of non-fill L2 pixels within a $1^\circ \times 1^\circ$ L3 grid box. The key detail is how each parameter defines a pixel as “cloudy”. There are two primary cloud fraction parameters in L3 MODIS Atmosphere files.

7.1.1. Cloud mask cloud fraction

The L3 cloud fraction that garners the most interest from MODIS data users is cloud fraction derived directly from the 35_L2 Cloud Mask (note the 35_L2 Cloud Mask flags are duplicated in the 06_L2 Cloud product). These L3 cloud mask cloud fraction SDSs have the prefix

- [Cloud_Fraction](#)
- [Cloud_Fraction_Day](#)
- [Cloud_Fraction_Night](#)

The first parameter listed above contains both day and night retrievals, the second parameter contains daytime only retrievals (solar zenith angle less than 81.4°), and the

third parameter contains nighttime only retrievals.

The L2 cloud mask cloud fraction, which L3 uses to compute statistics, is calculated in the cloud top properties algorithm and stored in the 06_L2 HDF file. The L2 cloud mask cloud fraction is derived in L2 (the L2 SDS names are identical to those shown above) by reading the 1×1 km Cloud Mask “Cloudiness Probability Flags” or “Cloudiness Flags” (see Table 5).

The Cloud Mask Cloudiness Flag can have the following settings: 0 = confident cloudy, 1 = probably cloudy, 2 = probably clear, and 3 = confident clear. In the computation of the L2 cloud mask cloud fraction, the first two flags are assigned 100% cloudy and the last two flags 100% clear. Then in each 5×5 km L2 retrieval box, the average L2 cloudiness is computed by computing the fraction of the 1-km pixels that are cloudy.

In the L3 Daily product, the cloud mask cloud fraction is computed by taking an unweighted average of these 5×5 km L2 cloud fractions in each $1^\circ \times 1^\circ$ grid box.

7.1.2. *Cloud optical properties cloud fraction*

The second most utilized L3 cloud fraction is that derived from the Cloud Optical Properties retrieval. All optical property cloud fractions are computed for **daytime scenes only** (solar zenith angle less than 81.4°). These L3 cloud optical properties cloud fraction SDSs have the prefix

- [Cloud_Fraction_Combined](#)
- [Cloud_Fraction_Liquid](#)
- [Cloud_Fraction_Ice](#)
- [Cloud_Fraction_Undetermined](#)

Table 5. Two key L2 Cloud Mask Flags used to compute the L2 Cloud Mask Cloud Fraction.

L2 QA Flag	Flag Value	Meaning
Cloud Mask Status Flag	0	Undetermined
	1	Determined
Cloud Mask Cloudiness Flag	0	Confident Cloudy (or Fill if Status Flag = 0)
	1	Probably Cloudy
	2	Probably Clear
	3	Confident Clear

The first parameter represents the cloud fraction for all cloud phases; the second, liquid water clouds only; the third, ice clouds only; and the fourth, undetermined cloud phase clouds only.

Unlike the cloud mask cloud fraction (detailed in section 7.1.1.), which is initially computed in L2, the optical property cloud fraction is not available in L2. The optical property cloud fraction is computed in L3 using the Primary Cloud Retrieval Phase Flag and the Primary Cloud Retrieval Outcome Flag (both a part of the Quality_Assurance_1km SDS in 06_L2 product files) for a sampled set of L2 grid points (every 5th row, every 5th column (see Section 2.2)). See Table 6.

Every sampled L2 grid point that has a Primary Cloud Retrieval Outcome Flag = 1 (Retrieval Successful) and a Primary Cloud Retrieval Phase Flag of 2 (Liquid Water Cloud), 3 (Ice Cloud), or 4 (Undetermined Phase Cloud) are taken as 100% cloudy for the cloud phase category in question.

However if any of these aforementioned L2 grid points have a Primary Cloud Retrieval Outcome Flag = 0 (Retrieval Not Attempted or Unsuccessful) then they are thrown out of the computation completely (even if the Cloud Retrieval Phase Flag was set to a valid phase). This logic was implemented because the L2 Cloud Optical Properties retrieval algorithm sets the Phase Flag before successful completion of the retrieval. In the unsuccessful case, the retrieval algorithm might have become unsure of the cloud phase, or even if the scene actually contains a cloud. Therefore it was decided to use the Retrieval Outcome Flag as a screen when calculating the L3 cloud fraction. (Another reason for this logic was to maintain correct “successful retrieval” pixel count statistics in

Table 6. Two key L2 QA Flags used to compute the Cloud Optical Properties Cloud Fraction.

L2 QA Flag	Flag Value	Meaning
Primary Cloud Retrieval Phase Flag	0	Cloud Mask Undetermined (Missing or Fill)
	1	Not Processed (Typically Clear)
	2	Liquid Water Cloud
	3	Ice Cloud
	4	Undetermined Phase Cloud
Primary Cloud Retrieval Outcome Flag	0	Retrieval Not Attempted or Unsuccessful
	1	Retrieval Successful

the Daily file, which are later used to weight individual Daily statistics in the computation of L3 Eight Day and Monthly statistics. See section 4.1).

Finally, sampled L2 grid points that have a Primary Cloud Retrieval Phase Flag of 1 (Not Processed) are counted as 100% clear. These retrievals will always have a Retrieval Outcome Flag of 0 (Unsuccessful).

The optical property cloud fractions are then computed from the following equations:

$$\text{Cloud Fraction Liquid} = \text{SCRW} / \text{Total}$$

$$\text{Cloud Fraction Ice} = \text{SCRI} / \text{Total}$$

$$\text{Cloud Fraction Undetermined} = \text{SCRU} / \text{Total}$$

$$\text{Cloud Fraction Combined} = (\text{SCRW} + \text{SCRI} + \text{SCRU}) / \text{Total}$$

where,

SCRW = Number of **Successful** Liquid Water Cloud Retrievals

SCRI = Number of **Successful** Ice Cloud Retrievals

SCRU = Number of **Successful** Undetermined Phase Cloud Retrievals

Total = Number of Clear + **Successful** Cloudy Retrievals for all Phases

So in summary, the Cloud Optical Properties Cloud Fraction is really a “Successful Cloud Optical Properties Retrieval Fraction.” However this does approximate fairly effectively real world cloudiness, but will tend to throw out some clouds where retrievals become problematic.

7.1.3. Comparing the cloud mask with the optical properties cloud fraction

Since the cloud optical property cloud fraction is computed for daytime scenes only and the cloud mask cloud fraction is computed for all cloud phases only, users should only attempt to make direct comparisons of the cloud mask cloud fraction for **daytime** (Cloud_Fraction_Day, highlighted in section 7.1.1) with the cloud optical property cloud fraction for **combined phase** (Cloud_Fraction_Combined, highlighted in section 7.1.2).

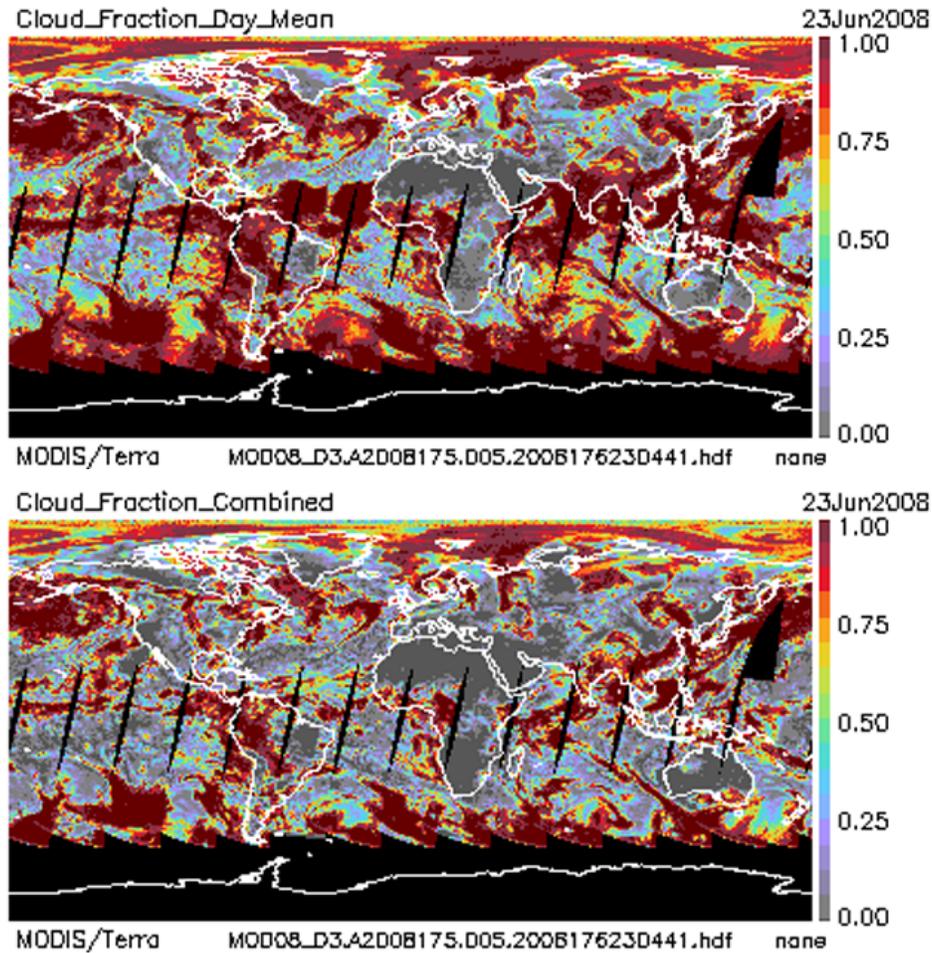


Figure 15. A comparison of a Daily Cloud Mask Cloud Fraction for daytime (top) and Optical Properties Cloud Fraction for clouds of all phases (also computed for daytime only) (bottom). It is normal to find the optical property cloud fractions smaller than the cloud mask cloud fraction, since failed optical property retrievals are thrown out and a “clear sky restoration” algorithm in the optical properties code might set some cloudy cloud mask pixels back to clear.

Because the unsuccessful retrievals are thrown out in the computation of the Optical Properties Cloud Fraction and additional tests in the cloud optical properties retrieval can set some pixels identified as cloudy by the cloud mask back to clear (clear sky restoration), the Optical Property Cloud Fraction for Combined Phase Clouds (Cloud_Fraction_Combined) can show some marked divergence from the Cloud Mask Cloud Fraction for Daytime (Cloud_Fraction_Day), even though the Cloud Optical Properties algorithm is based on (starts with) results from the Cloud Mask. This can be seen on examination of Figure 15.

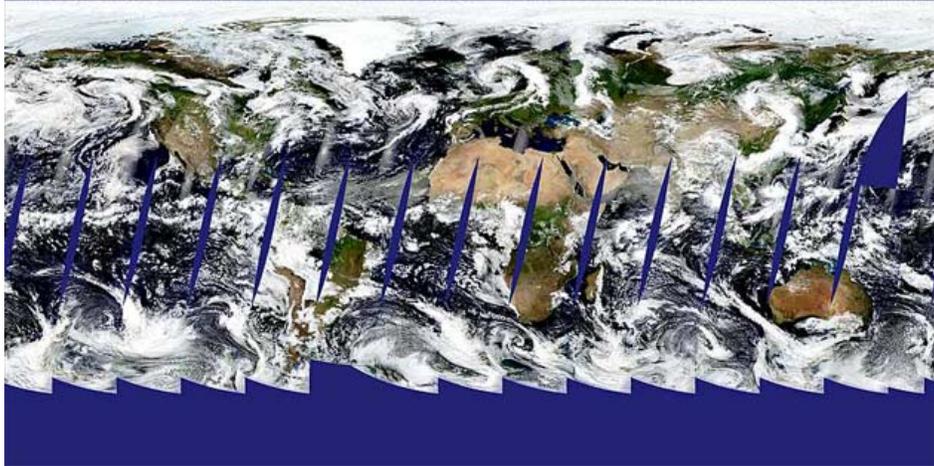


Figure 16. RGB true-color composite global image for 23 June 2008 (daytime only). Clouds and aerosol appear white.

Finally, one should remember that the L3 Cloud Optical Property Cloud Fraction is computed from sampled L2 cloudiness at 5-km resolution in the $1^\circ \times 1^\circ$ L3 grid box; while the Cloud Mask Cloud Fraction is computed in the $1^\circ \times 1^\circ$ L3 grid box from average cloud fractions in the 5×5 km region (not sampled at L3). Although it is not believed this sampling vs. averaging difference has a serious impact on the final results, users should be aware that the two cloud fractions are computed differently.

The primary benefits to the cloud optical properties cloud fraction is it allows an analysis of the cloud fractions separated by cloud phase (liquid water clouds vs. ice clouds). The primary benefits to the cloud mask cloud fraction is it's the most direct way to view cloud mask results; and has cloud fraction results at night.

As to which of these two cloud fraction SDSs more accurately represents “real world” cloudiness, Figure 16 shows a global RGB composite (tuned to show clouds and aerosol as white) for the same time period as the two cloud fraction images in Figure 15. Although there are large areas of broken clouds in the image, it's not clear using a L1B image at this scale, which cloud fraction did a better job.

To further this investigation, consider the region of the Arabian Sea to the west of India and to the southeast of Oman and Yemen in Figure 15, where the Cloud Mask Cloud Fraction shows over 90% cloudy and the Cloud Optical Properties Cloud Fraction

shows less than 25% cloudy. Which is closer to the truth?

Figure 17 shows a zoom of this region for all three previous images mapped on a latitude-longitude grid to minimize mapping distortion. The L1B image in Figure 17 shows a distinct signature of blowing dust and sand from the African continent towards and over India. Added to this set of images is an Aerosol Optical Depth over ocean image that confirms high concentrations of blowing dust over portions of the Arabian Sea (dark red color). It seems clear that for this case and this region, the Cloud Mask interpreted the blowing dust and sand over the Arabian Sea as cloud, whereas the Cloud Optical Properties algorithm was able to correctly reduce the measured cloudiness over this region.

This same effect can be seen over the Atlantic Ocean west of Africa's Sahara Desert where more blowing dust (sand) was interpreted by the Cloud Mask as cloud (refer back to Figure 15).

However, it should be noted that there are other areas where the Cloud Mask cloud fraction will do a better job than the Cloud Optical Properties cloud fraction, namely regions where there are large quantities of broken (scattered) cloudiness, especially found over ocean. This is primarily due to a "clear sky restoration" algorithm in the Cloud Optical Properties code, which restores pixels identified as cloud by cloud mask to clear if they touch another pixel that was identified as clear. This clear sky restoration "edge detection" algorithm was implemented because these cloud edge pixels are not retrieved (for all cloud optical property parameters) with enough accuracy, in general, to warrant their inclusion. In this case, the Cloud Optical Properties cloud fraction will show numbers that are most likely too low when compared to the Cloud Mask cloud fraction.

7.2. What do the 'undetermined' and 'combined' cloud phases mean?

The undetermined cloud phase means the cloud optical properties retrieval algorithm could not make a determination of the cloud phase (liquid water or ice). This may have

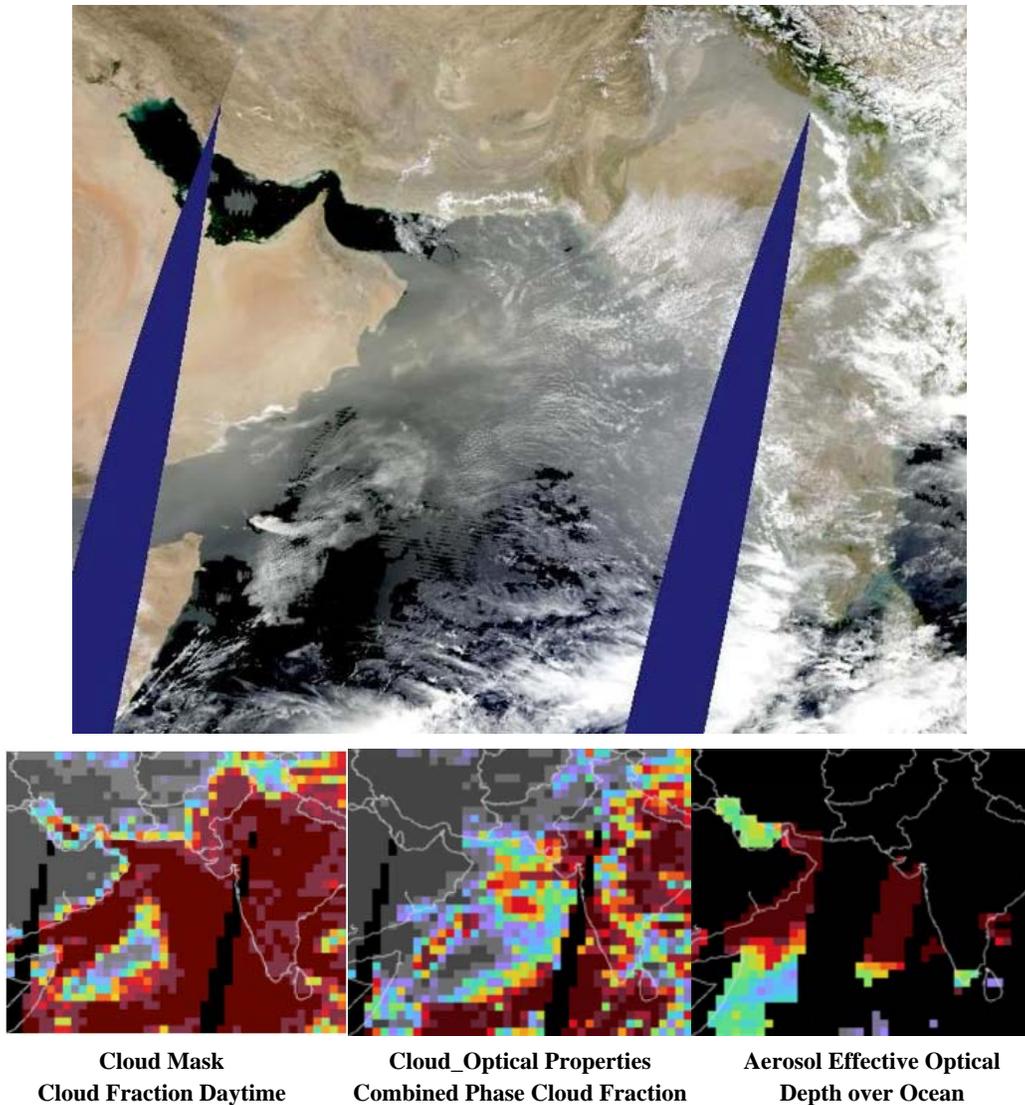


Figure 17. Top Image: RGB true-color composite regional image for 23 June 2008 (daytime only) over the Arabian Sea west of India. Clouds and aerosol appear white. Second Row Image Left: Cloud Mask Cloud Fraction. Center: Cloud Optical Properties Cloud Fraction. Right: Aerosol Optical Depth over Ocean.

been caused by viewing anomalies in the retrieval (sunglint), contamination of the scene by aerosol, or a multi-layer cloud with mixed phases (e.g., thin cirrus overlying liquid water clouds). For these undetermined retrievals the liquid water libraries are used in the cloud optical properties retrievals, but the retrievals are considered to be of lower confidence (and quality) than those that are placed in one of the other primary phase categories.

The combined phase is simply a combination of all cloud phase categories: liquid

water, ice, and undetermined.

7.3. What is the minimum illumination requirement for a MODIS retrieval?

Many MODIS parameters produce results for both daytime and nighttime retrievals; so for those there is no minimum illumination requirement. MODIS parameters that are earmarked “Daytime Only” are (i) Aerosol, (ii) Near-Infrared Water Vapor, (iii) Cirrus Detection, and (iv) Cloud Optical Properties. For these daytime only parameters, the Aerosol retrieval requires the solar zenith angle to be less than 72° , all others require the solar zenith angle to be less than 81.4° (which matches the cloud mask classification of “daytime”).

7.4. What is the best way to display (marginal) histogram data?

Most parameters in the L3 product files have marginal (1D) histogram data. (For a list of parameters with marginal histogram data, see Appendix B. For a summary of the histogram boundaries for all parameters for Collection 005 and/or Collection 006, see Appendix D.)

Marginal histogram data are typically displayed using a series of rectangular bars whose two dimensions represent (i) the bin size and (ii) the histogram value. The histogram value plotted can either be pure counts or a normalized probability density function (PDF).

Figure 18 shows a marginal histogram of counts for liquid water cloud effective particle radius pixels over ocean only for June 2005 as observed by the Aqua MODIS instrument. The bin boundaries are on the x-axis and the height of each bar represents the total L2 pixel counts (sampled) in each bin. The visual discontinuity in counts (vertical bar heights) in Figure 18, starting at the bin boundary of $20\ \mu\text{m}$, is due to the change of bin widths (going from $1\ \mu\text{m}$ to $2\ \mu\text{m}$ bins).

Also it should be noted that the first two bins (2-3 and 3-4) contain zero counts due to a Collection 005 change in the L2 cloud optical properties code **not** allowing any effec-

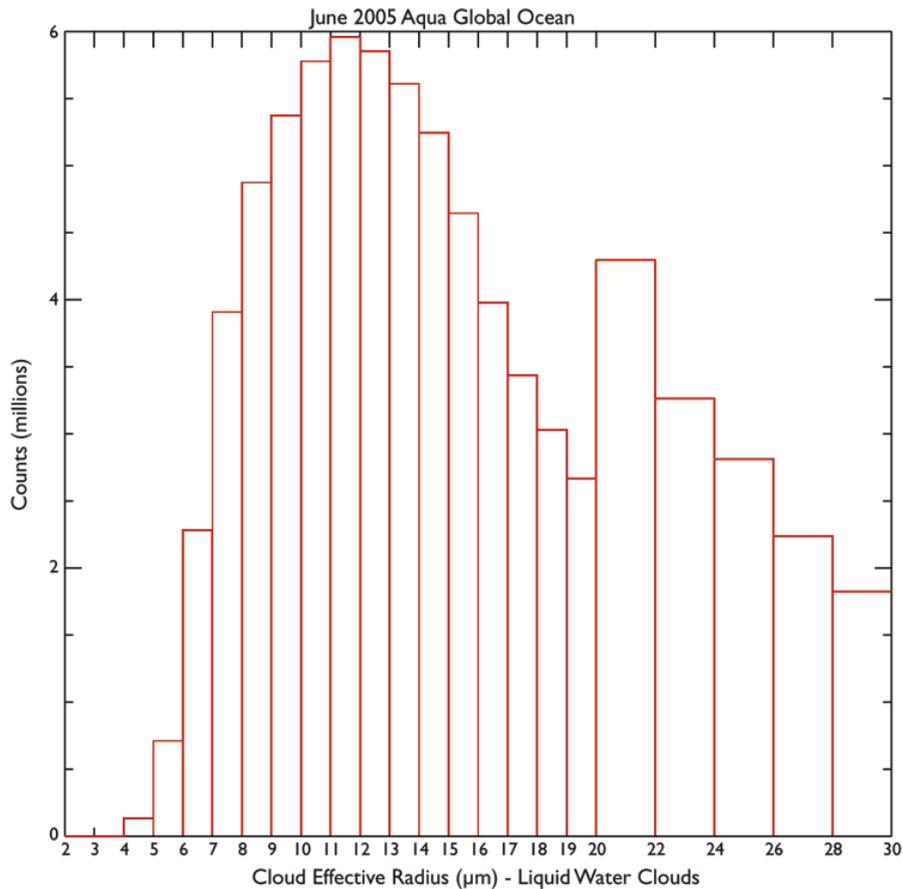


Figure 18. Marginal Histogram of L2 Pixel Counts (sampled) for Liquid Water Cloud Effective Radius during June 2005 (daytime only) for Global Ocean data from Aqua/MODIS.

tive radii $\leq 4 \mu\text{m}$ to be computed or stored. In Collection 006, the bottom bin boundary for cloud effective radius in L3 will be adjusted to start at $4.0 \mu\text{m}$ instead of $2.0 \mu\text{m}$.

Users should keep in mind that one can plot “counts” (which is how the data are stored in the L3 HDF file) in each bin; or normalize the data by taking into account the bin sizes and plot “normalized probability” in each bin. This normalized probability calculation is performed as follows:

$$\text{Normalized Probability} = (\text{bin_counts}) / (\text{bin_width} * \text{total_counts}) \quad (4)$$

Figure 19 shows the same marginal histogram except plotted as a normalized probability density function (PDF). In this plot the double wide bins are taken into account, so the heights of the vertical bars starting at $20 \mu\text{m}$ are half as high as those in Figure 18.

For a normalized one-dimensional histogram plot, if one calculates the (size of bin) x

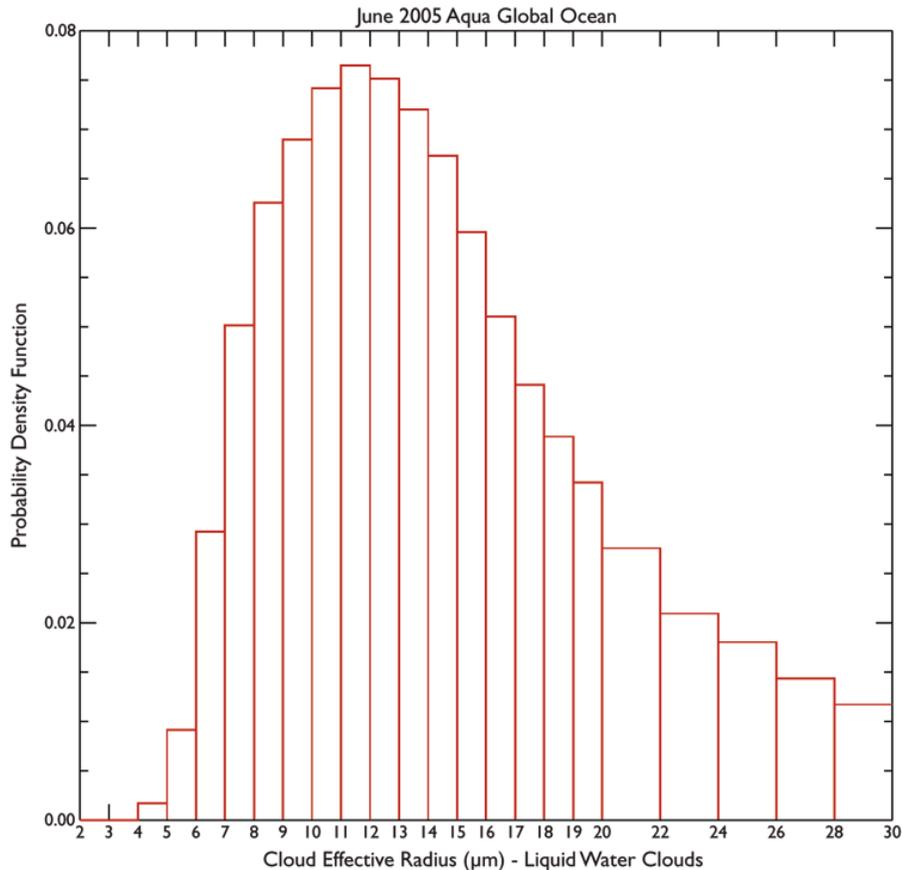


Figure 19. Marginal Histogram of Normalized Probability Density Function (PDF) (sampled) for Liquid Water Cloud Effective Radius during June 2005 (daytime only) for Global Ocean data from Aqua/MODIS.

(PDF value of bin) then an “area of the rectangle” calculation is actually being performed. When one sums all the rectangle areas, one ends up with the area (or integral) under the ‘curve’ of 1.0 or 100%.

For example, the peak PDF histogram bin value shown in Figure 19 is 0.076. (This occurs in the tenth bin, which corresponds to $11 \mu\text{m} < r_e \leq 12 \mu\text{m}$.) This means that 7.6% of the total area of the entire plot fell in this bin. This can also be interpreted as the probability of any one pixel falling in this peak histogram bin is 7.6%. If one performs this calculation for every bin (rectangle), and sums them, one ends up with 1.0 or 100%.

7.5. What is the best way to display joint histogram data?

Common ways to display Joint Histogram data are (i) 3D lego plots, (ii) smoothed

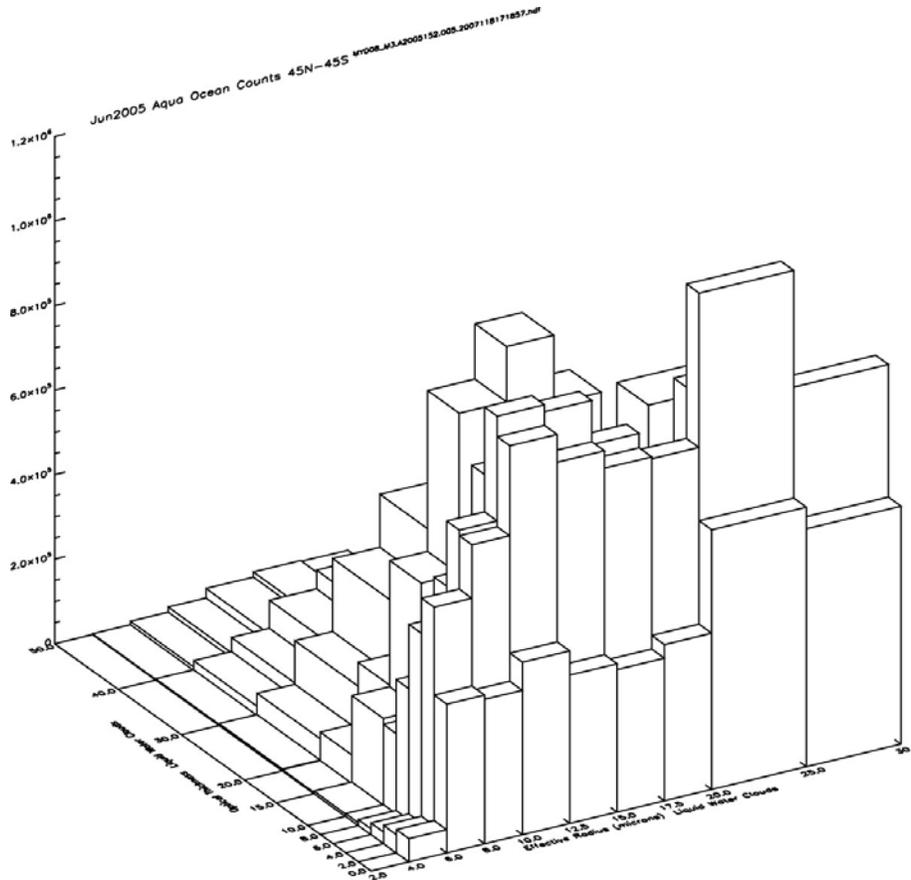


Figure 20. A June 2005 MYD08_M3 Counts Joint Histogram of cloud optical thickness (y-axis) vs. cloud effective radius (x-axis) for liquid water clouds displayed as a “3D lego plot” with post-processing to limit the data to ocean-only L3 grid cells that range from 45°N to 45°S. The top bin of cloud optical thickness from 50 to 100 was chopped off. The height of each Lego bar represents the number of counts in each bin. It is often difficult to impossible to orient lego plots to make all bins visible.

contour plots, and (iii) color-coded histogram bin plots.

Lego plots (Figure 20) are easily made in many graphics packages (such as Interactive Data Language, IDL); however it’s often difficult to read the exact “height” of each lego block, and some (or most) viewing orientations may cause some blocks to be hidden by larger blocks in front of them.

Smoothed contour plots (Figure 21) are useful as they allow quick visual interpretation of data (without an intimate knowledge of the color bar scale) and therefore lend themselves well to time-series animations, where the time to view each joint histogram movie frame is short. However smoothing options can be hard to fine tune and the result

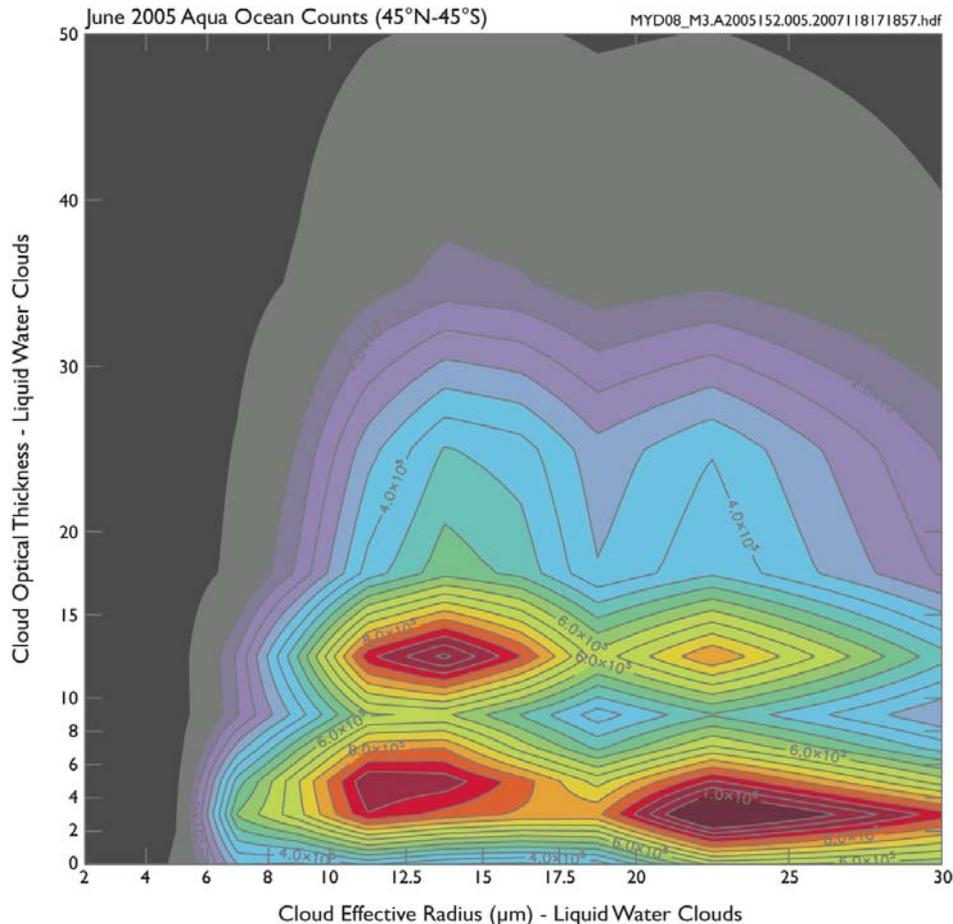


Figure 21. A June 2005 MYD08_M3 Counts Joint Histogram of cloud optical thickness (y-axis) vs. cloud effective radius (x-axis) for liquid water clouds displayed as a “colored contour plot” with post-processing to limit the data to ocean-only L3 grid cells that range from 45°N to 45°S. The top bin of cloud optical thickness from 50 to 100 was chopped off. The contour colors represent the number of counts in each bin (low is grey, high is red). Contour plots show a distorted (smoothed) view of the data stored in the joint histogram.

is always a plot that shows a distorted (smoothed) view of the actual data.

The third option, color-coded histogram bin plots (Figure 22), work the best in most cases and have few drawbacks; however some graphics packages cannot produce these plots without expert knowledge of the language and some user programming. The pros of this final method are there is no distortion (smoothing) of data in the x-y plane (as is the case with smoothed contour plots) and all cells (bins) can be viewed without any visual obstruction (as is the case with lego plots). Figures 20, 21, and 22 show identical MODIS L3 joint histogram counts data plotted using these three described techniques.

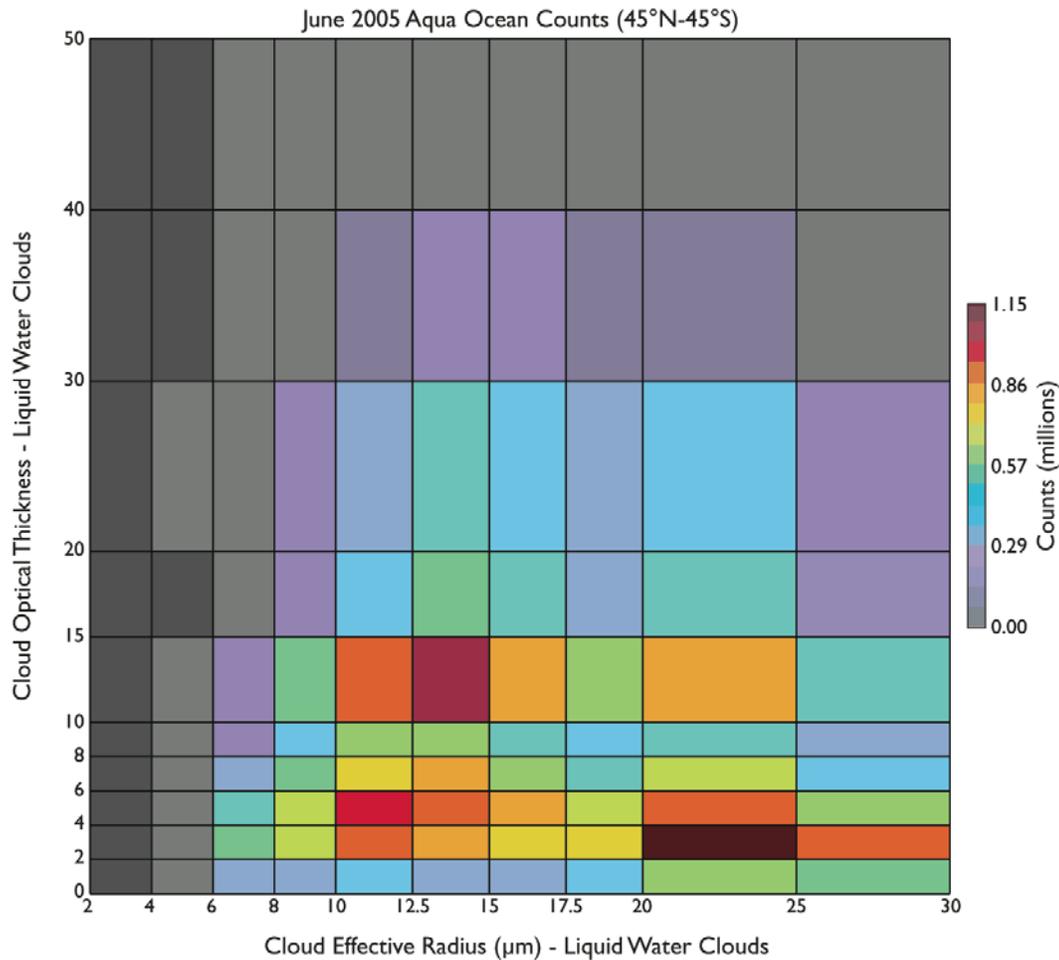


Figure 22. A June 2005 MYD08_M3 Counts Joint Histogram of cloud optical thickness (y-axis) vs. cloud effective radius (x-axis) for liquid water clouds displayed as a “colored histogram bin plot” with post-processing to limit the data to ocean-only L3 grid cells that range from 45°N to 45°S. The top bin of cloud optical thickness from 50 to 100 was chopped off. The bin colors represent the number of counts in each bin (low is grey, high is red).

For optimal display, it’s best to have the graphics package print tick marks and labels at the exact histogram bin boundaries on both axes, so that the bin boundaries and sizes are clear to the viewer. These bin boundaries can be obtained from local attributes attached to each Joint Histogram SDS or from the CDL file specification available on the MODIS Atmosphere web site (modis-atmos.gsfc.nasa.gov). They are also provided in Appendix E for convenience.

Since all L3 joint histograms are stored as 4D arrays (Lat, Lon, Parameter1Bin, Parameter2Bin), MODIS data users can read (and view) a unique joint histogram for every

L3 $1^\circ \times 1^\circ$ grid cell. This allows users to easily perform post-processing to narrow the application of the joint histogram, which is then optimized to show specific scientific results.

Users can apply a latitude/longitude screen to the $1^\circ \times 1^\circ$ gridded histograms so a newly computed (summed) joint histogram applies to a specified range of latitudes and longitudes only.

Users can also apply land only or ocean only masks when summing individual $1^\circ \times 1^\circ$ gridded histograms so the newly summed joint histogram applies to land-only or ocean-only regions. (Note that a land/ocean mask for L3 data can be obtained from the MODIS-Atmosphere web site (modis-atmos.gsfc.nasa.gov/MOD08_M3/grids.html)).

Users might also sometimes find it advantageous to delete (remove) the highest bin along one or both data axes (which typically have a large data range and few counts) in order to show more detail in other more scientifically relevant bins.

Another item to keep in mind is one can plot “counts” (which is how the data are stored in the L3 HDF file) in each bin; or sometimes it is more useful to normalize the data by taking into account the bin sizes and then plot “normalized probability” in each bin. This normalized probability calculation is performed as follows:

$$\text{Normalized Probability} = (\text{bin_box_counts}) / (\text{bin_box_area} * \text{total_counts}) \quad (5)$$

Converting the counts to normalized probabilities removes the visual anomalies that occur when comparing bins of different sizes in joint histogram plots. Figure 23 shows how the data in Figure 22 change when going from raw counts to normalized probability, where an adjustment is made for the bin sizes. It’s clear that normalizing the data can really change one’s perspective of the results.

Finally, some ask, what do the Probability Density Function (PDF) numbers computed for each histogram bin in Figure 23 actually (physically) represent? If a normalized one-dimensional histogram plot means the area under the curve equals one; then a normalized joint histogram PDF plot means the volume under the surface equals one.

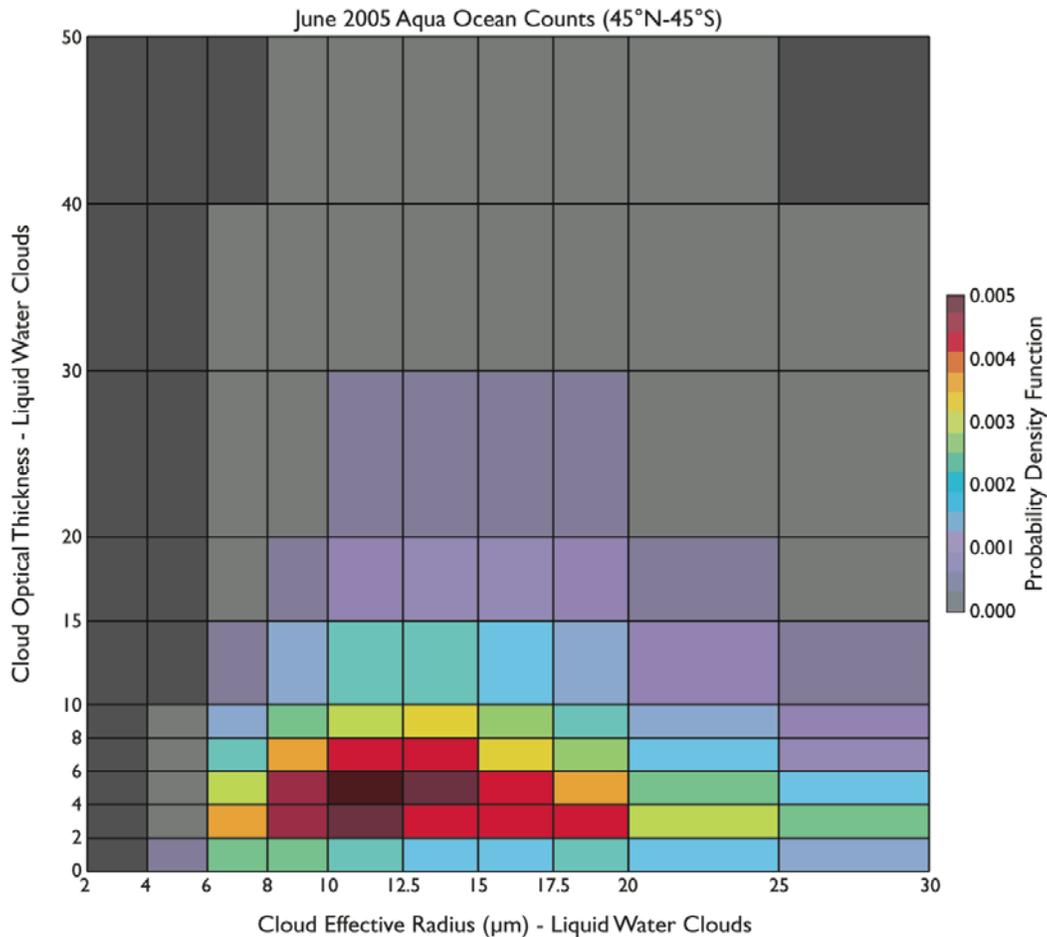


Figure 23. A June 2005 MYD08_M3 Normalized Probability Density Function (PDF) Joint Histogram of Cloud Optical Thickness (y-axis) vs. Cloud Effective Radius (x-axis) for Liquid Water Clouds displayed as a “colored histogram bin plot” with post-processing to limit the data to Ocean-only L3 grid cells that range from 45N to 45S. The top bin of Cloud Optical Thickness from 50 to 100 was removed. The bin colors represent the PDF in each bin (low probability is grey, high probability is red).

(The surface being defined as z-axis heights or PDF values in each bin.) If one calculates the (area of bin) \times (height of bin) or, alternatively (area of bin) \times (PDF value of bin) then a “volume of the block” calculation is actually being performed. When one sums all the block volumes, one ends up with the volume under the surface, which for the normalized JPDF surface is 1.0 or 100%.

For example, the peak PDF histogram bin value shown in Figure 23 is 0.005. If you take this number and multiply by the area of the bin ($2.5 \times 2.0 = 5.0$), one gets a bin (block) volume of 0.025. This means that 2.5% of the total volume under the surface of

the entire plot fell in this bin (block). This can also be interpreted as the probability of any one pixel falling in this peak joint histogram bin box is 2.5%. If one performs this calculation for every bin (block), and sums them, one ends up with 1.0 or 100%.

It should be noted that for the joint histogram in Figure 23, the top bin ($50 < \text{Optical Thickness} \leq 100$) was cut off in order to show (allow) more visual detail in the remaining bins. In order to get the proper summed value of 1.0 (100%), one would need to include this deleted bin in the calculation.

7.6. How is the aerosol cloud mask computed and how should it be interpreted?

According to the L2 Aerosol development team, the L2 SDSs “Cloud_Fraction_Ocean” and “Cloud_Fraction_Land,” used to compute the C05 L3 SDS “Aerosol_Cloud_Mask_Cloud_Fraction,” were intended to be a measure of the cloud fraction in 10×10 km scenes, regardless of whether an aerosol retrieval was attempted or successful.

However, there were bugs in the L2 aerosol cloud mask algorithm that caused the following problems for some retrievals: (i) where no aerosol retrievals were attempted, the L2 SDS was incorrectly assigned a fill value, (ii) in sunglint regions over ocean, the aerosol cloud fraction algorithm assigned either fill or 100% cloudy values even for regions where the cloud mask detected clear, and (iii) in polar regions over ocean, the aerosol cloud mask algorithm assigned either fill or very high (approaching 100%) cloudy values, even for regions where the cloud mask detected less cloudiness.

In L3, since fill values are thrown out, the $1^\circ \times 1^\circ$ average aerosol cloud fraction incorrectly tended towards 100% over ocean sunglint and in polar regions over ocean.

Figure 24 shows a comparison of cloud fraction from cloud mask vs. the aerosol cloud mask over land and ocean. With the bug in the aerosol cloud mask, the cloud fraction is clearly too high in some regions and too low in other regions when compared to the cloud mask cloud fraction, which tends to track real world cloudiness.

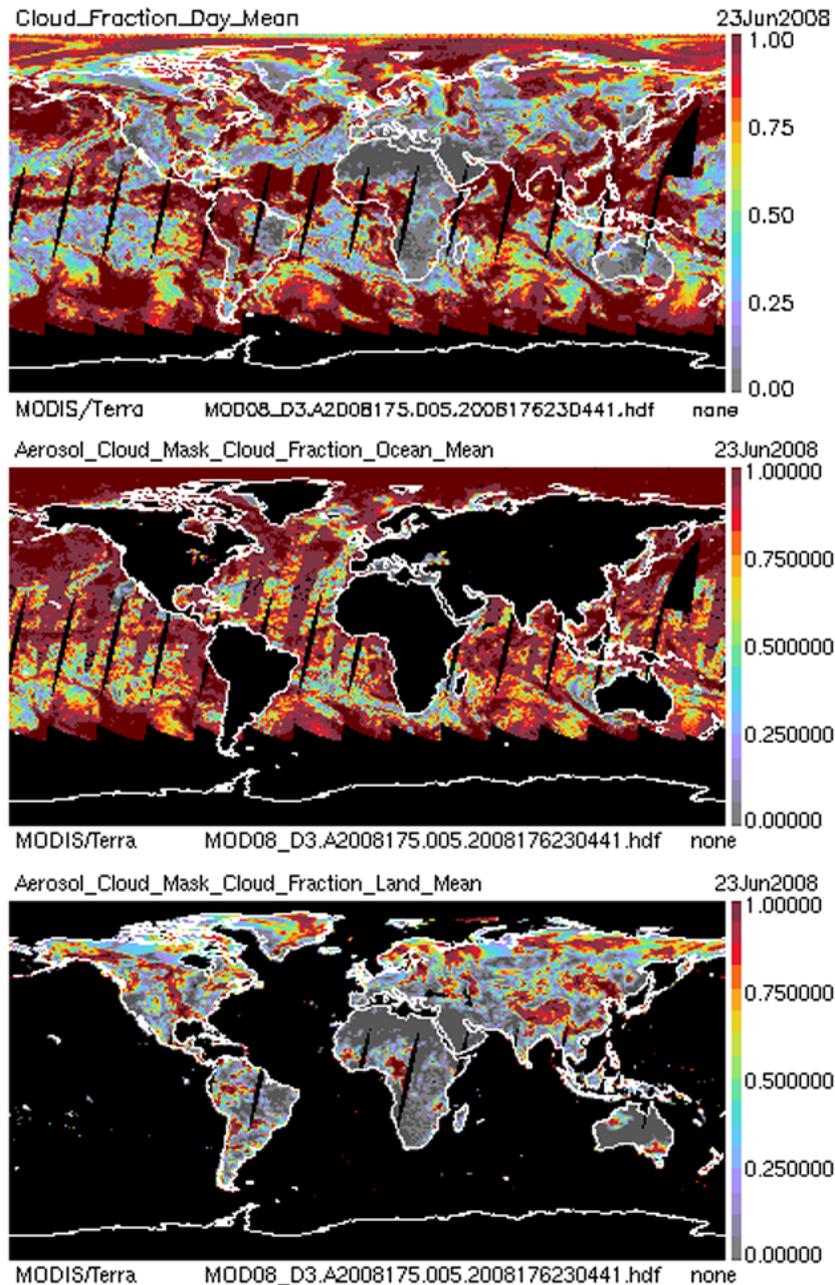


Figure 24. A comparison of a Daily Cloud Mask Cloud Fraction for daytime and Aerosol Cloud Mask Cloud Fraction (Land and Ocean), also computed for daytime only. The bug in the “Aerosol Cloud Mask” makes a substantial impact over ocean, and less so over land.

At the time of publication of this ATBD, it was expected that these software bugs would be fixed at the start of Collection 051 (an interim collection expected to begin in the Fall of 2008), so there could be a marked improvement in these L3 SDSs in Collection 051.

It should also be noted that the future of this parameter in L3 for Collection 006 is under discussion.

7.7. How should the various multilayer cloud fractions be interpreted?

There are a number of multilayer cloud fractions that use the “Multilayer Cloud and Phase Flag” to compute various cloud fractions.

The `Cloud_Fraction_1L_(phase)` parameter is computed the same way as the `Cloud_Fraction_(phase)` parameter, outlined in section 7.1.2, except that it uses the “Primary Cloud Retrieval Multilayer Cloud & Phase Flag” (See MODIS Atmosphere QA Plan) where the flag is set to “single layer cloud.” This gives the cloud fraction separated by phase for single layer clouds only.

The `Cloud_Fraction_ML_(phase)` parameter is computed the same way, except that it uses the “Primary Cloud Retrieval Multilayer Cloud & Phase Flag,” where the flag is set to “multi-layer cloud.” This gives the cloud fraction separated by phase for multi-layer clouds only.

Both of the cloud fraction parameters outlined above are true cloud fractions whose computation includes clear-sky pixel counts in the denominator. Figure 25 (top) shows a sample Cloud Fraction image for single-layer liquid water clouds, and Figure 25 (bottom) shows a sample Cloud Fraction image for multi-layer liquid water clouds. It is typical for the multi-layer liquid water cloud fraction to be less than the single-layer liquid water cloud fraction. When these two fractions are added, it will equal the `Cloud_Fraction_Liquid SDS`, which depicts the liquid water cloud fraction for all layer clouds.

Finally there are a set of cloud “fractions” that do not include clear-sky pixel counts in the denominator; that is they are not standard cloud fractions but instead are cloud ratios. SDSs that have the prefix “`ML_Fraction_(phase)`” show a ratio of multi-layer clouds to all-layer clouds for various cloud phases. In other words, it shows the percen-

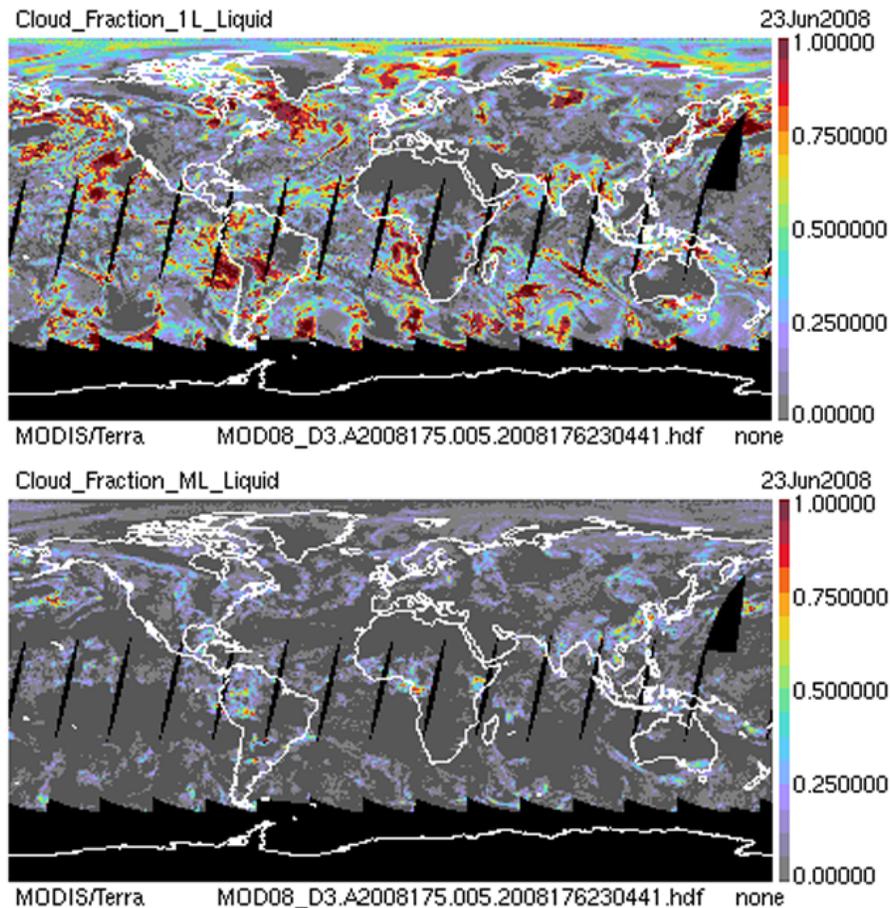


Figure 25. A comparison of a single-layer liquid water cloud fraction with multi-layer liquid water cloud fraction.

tage of clouds that were detected as multi-layer for various cloud phases (cf. Figure 26).

Users should be aware that there are SDS name length limits; sometimes an SDS name might not be descriptive enough to determine its content. If at any time MODIS data users need more clarity on what any SDS actually contains, additional documentation as well as computational details can often be obtained from the local attribute “long_name” attached to each SDS. These attributes can also be read from the online version of the file specifications, available from the MODIS-Atmosphere web site (modis-atmos.gsfc.nasa.gov/MOD08_M3/spec_c5.html).

7.8. Is there a minimum L2 pixel count requirement to compute L3 statistics?

Within the L3 Daily product file, for any given $1^\circ \times 1^\circ$ grid cell, only a single L2

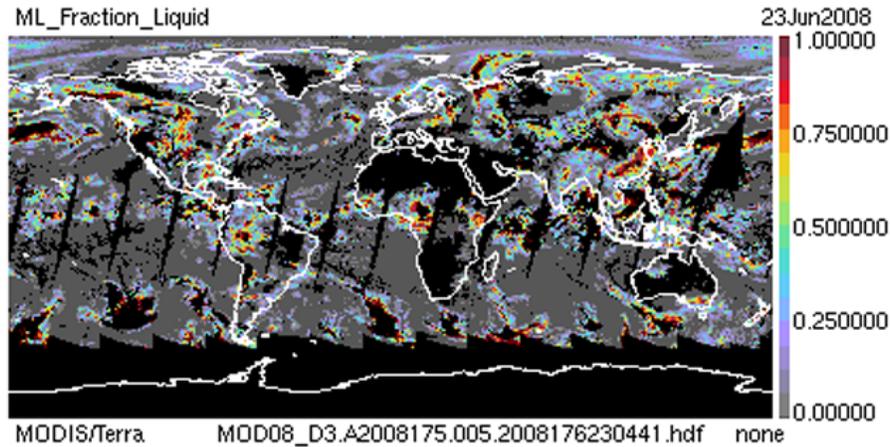


Figure 26. A depiction of the ratio of multilayer liquid water clouds to all liquid water clouds. In other words, what fraction of liquid water clouds were detected as multi-layer.

non-fill pixel is needed to create L3 daily statistics. This tends to cause an artificial spatial expansion of sparse L2 data when going to L3. For example, a sizable region might have only a few scattered L2 retrievals, but L3 might show solid coverage of $1^\circ \times 1^\circ$ statistics.

MODIS L3 data users have a relatively easy way to monitor the number of L2 pixels that went into L3 statistics for each $1^\circ \times 1^\circ$ grid cell. Pixel count data are stored within the L3 Daily product file in either the “Pixel_Counts” or the 4th dimension of the “Confidence_Histogram” SDS.

Likewise for the L3 Multiday product files, for any given $1^\circ \times 1^\circ$ grid cell, only a non-fill daily grid cell for a single day is needed to create L3 multiday statistics.

The only exception to this multiday statistic computation rule is in L3 products derived from L2 Aerosol parameters. The Aerosol group requested a special weighting scheme for L3 Multiday products, where a daily grid cell must have a pixel count of at least 6 to be included in the multiday statistic computation (See Table 4 in Section 4.2.) This was done because the Aerosol group felt that low pixel count daily grid cells were not very reliable and should be thrown out. This screening procedure also helped to reduce the number of low confidence multiday statistic grid cells near the poleward boundary (terminator) of their retrieval algorithm.

7.9. Are statistics within a L3 Daily file a single orbit snapshot or a multiple orbit average?

Two EOS satellites, Terra and Aqua, both carrying the MODIS sensor, are in sun-synchronous orbits. The Terra overpass time is around 1030 local solar time at the equator in its descending (daytime) mode and 2230 local solar time in its ascending (nighttime) mode. The Aqua overpass time is around 1330 local solar time at the equator in ascending (daytime) mode and 0130 local solar time in descending (nighttime) mode.

L3 SDSs within either a Terra or Aqua Daily file are made up of L2 data that are collected for daytime only, nighttime only, or combined daytime and nighttime scenes. Only L3 Daily SDSs that are based on L2 data collected for daytime only or nighttime only scenes, have the chance (depending on global location) to be pinned down to an approximate local solar time. MODIS L3 SDSs that are based on L2 data collected for combined daytime and nighttime scenes are, for most regions of the globe, a mixture of at least two MODIS overpasses approximately 12 hours apart.

To determine if a L3 SDS contains daytime data only, nighttime data only, or combined daytime and nighttime data, one needs to query the local attribute “Included_Nighttime_Data.” If this attribute is set to “False,” then it’s a daytime only pa-

Table 7. Availability of daytime only, nighttime only, or combined daytime and nighttime parameters in L3 organized by the “derived from” product group.

Derived from Product	Daytime Only SDSs?  DAYTIME	Combined Day & Night SDSs?  DAYTIME  NIGHTTIME	Nighttime Only SDSs?  NIGHTTIME
Aerosol 04_L2	Yes	No	No
Water Vapor 05_L2	Yes	Yes*	No
Cirrus Detection 06_L2 (CD)	Yes	No	No
Cloud Top Properties 06_L2 (CT)	Yes	Yes	Yes
Cloud Optical Prop. 06_L2 (OD)	Yes	No	No
Atmosphere Profile 07_L2	No	Yes	No

* All Combined Day & Night Water Vapor (05_L2) SDSs are copied from Atmosphere Profile (07_L2)

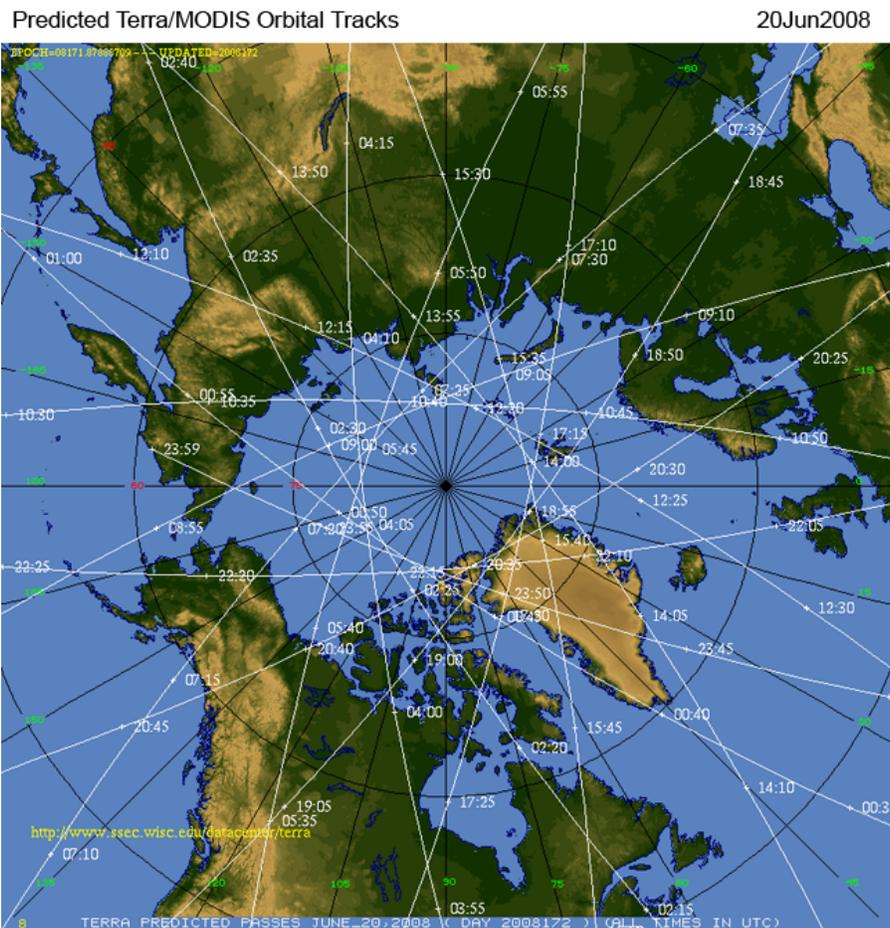


Figure 27. A polar view of predicted Terra MODIS orbit tracks for June 20, 2008. The overlapping pattern of orbits in the polar region is clear.

parameter (SDS). If this attribute is set to “True,” then it’s combined daytime and nighttime; unless the string “_Night_” appears somewhere in the SDS name, in which case it’s a nighttime only parameter. It should be noted that this last nighttime-only case is of low incidence, only occurring in a few cloud top property derived parameters (see Table 7).

Due to overlapping orbits toward the poles, **only** the daytime only or nighttime only Daily SDSs from approximately 23°N to 23°S can be pinned down to an approximate local solar time (since they are made up of observations from a single MODIS overpass only). Poleward of 23°, L3 Daily data become an average of several overlapping orbits approximately 100 minutes apart (see Figure 27).

Figure 28 shows the predicted Terra orbit tracks and a Terra L3 image of input (L2)

pixel counts for a daytime only parameter for June 20, 2008. The Terra orbit track image shows both the descending (day) node (white lines that slope from the upper right to lower left portion of the global image) and the ascending (night) node (white lines that slope from the lower right to upper left). The pixel count image shows the typical pattern of single orbit sampling between about 23°S and 23°N. Note that the black color in the pixel count image in Figure 28 shows regions of 0 pixel counts (orbital gap regions between approximately 23°N and 23°S, and polar darkness south of approximately 50°S for this date).

So in summary, to pin down a L3 Daily SDS to an approximate local solar time, it

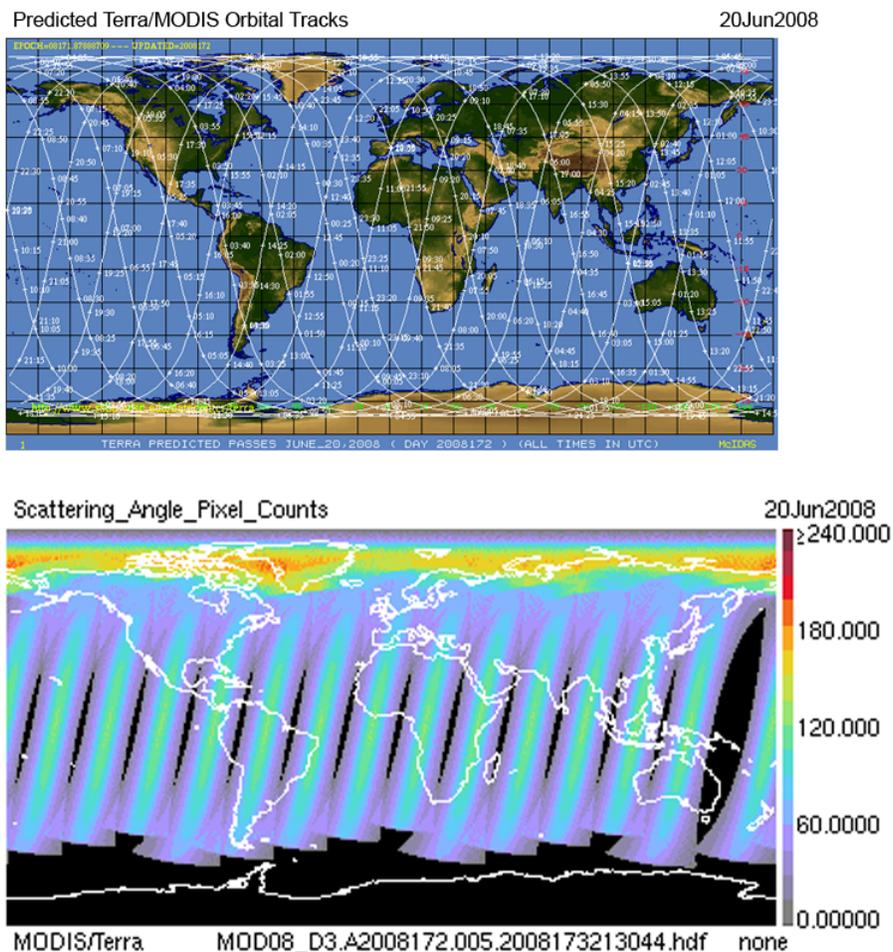


Figure 28. Predicted Terra MODIS orbit tracks (top) and a typical pattern of input pixel counts for a daytime only parameter (bottom) for June 20, 2008. The marked increase in pixel counts towards the pole is due to overlapping orbits causing the L3 data to become a multiple orbit average, over mid-latitudes the pattern of pixel counts is typical for a single orbit snapshot that can be assigned a single (approximate) local solar time.

must be (i) either daytime only or nighttime only and (ii) constrained to L3 grid cells between approximately 23°N and 23°S latitude. Only in both these situations does any Daily SDS contain data from a single MODIS overpass, whose approximate local solar time is given above.

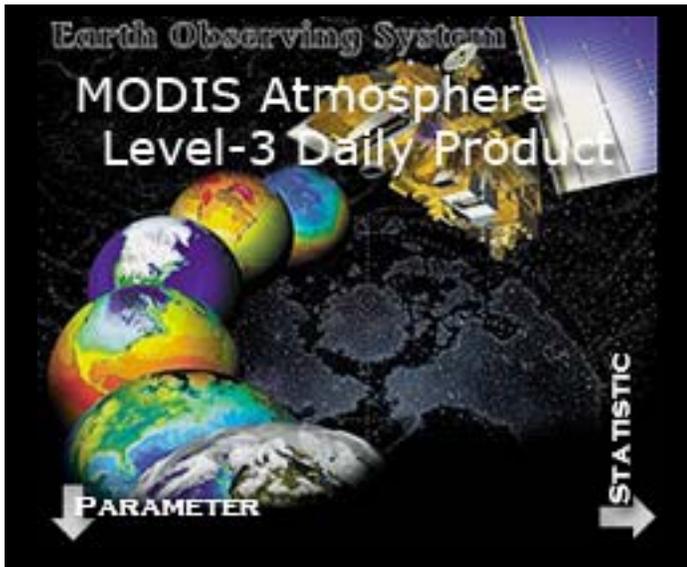
8.0. References

- Oreopoulos, L., 2005: The impact of subsampling on MODIS Level-3 statistics of cloud optical thickness and effective radius. *IEEE Trans. Geosci. Remote Sens.*, **43**, 366–373.
- Oreopoulos, L., R. Cahalan, and S. Platnick, 2007: The plane-parallel albedo bias of liquid clouds from MODIS observations. *J. Climate*, **20**, 5114–5125.

Appendix A:

**Scientific Data Sets
found in the
L3 Daily File
only**

*Statistics only in the L3 Daily (08_D3)
Collection 005 and 051*

	Mean	Standard_Deviation	Minimum	Maximum	OA_Mean	OA_Standard_Deviation	Fraction	Log_Mean	Log_Standard_Deviation	OA_Log_Mean	OA_Log_Standard_Deviation

Derived from L2 Aerosol (04_L2)

Combined Land & Ocean

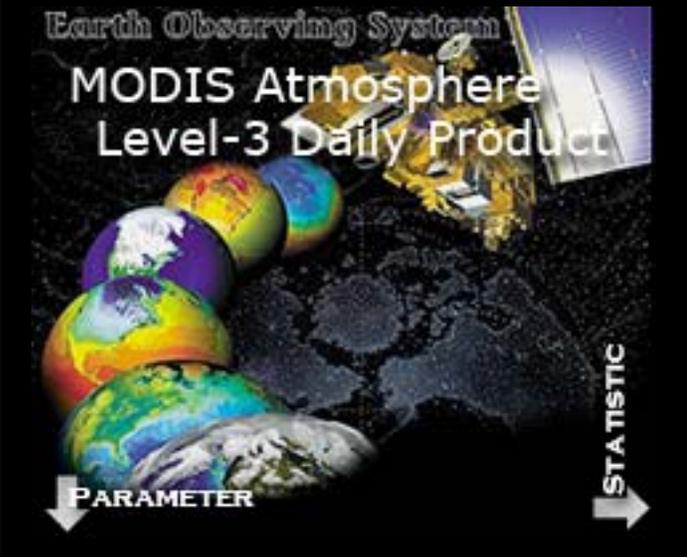
01. Scattering_Angle	•	•	•	•							
02. Solar_Zenith	•	•	•	•							
03. Solar_Azimuth	•	•	•	•							
04. Sensor_Zenith	•	•	•	•							
05. Sensor_Azimuth	•	•	•	•							
06. Optical_Depth_Land_And_Ocean	•	•	•	•							
07. Optical_Depth_Ratio_Small_Land_And_Ocean	•	•	•	•							

Land Only

08. Corrected_Optical_Depth_Land (3)	•	•	•	•	•	•					
09. Optical_Depth_Ratio_Small_Land	•	•	•	•	•	•					
10. Mass_Concentration_Land	•	•	•	•	•	•					
11. Angstrom_Exponent_Land ¹	•	•	•	•	•	•					
12. Mean_Reflectance_Land_All_QA47 (3)	•	•	•	•	•	•					
13. Mean_Reflectance_Land_All_QA66 (3)	•	•	•	•	•	•					
14. Path_Radiance_Land_QA47 (2)	•	•	•	•	•	•					
15. Path_Radiance_Land_QA66 (2)	•	•	•	•	•	•					
16. Critical_Reflectance_Land_QA47 (2)	•	•	•	•	•	•					
17. Critical_Reflectance_Land_QA66 (2)	•	•	•	•	•	•					
18. Aerosol_Cloud_Mask_Cloud_Fraction_Land	•	•	•	•							
19. Number_Pixels_Used_Land (2)	•	•	•	•							
20. Deep_Blue_Aerosol_Optical_Depth_Land (3)	•	•	•	•	•	•					
21. Deep_Blue_Aerosol_Optical_Depth_550_Land	•	•	•	•	•	•					
22. Deep_Blue_Angstrom_Exponent_Land	•	•	•	•	•	•					
23. Deep_Blue_Single_Scattering_Albedo_Land (3)	•	•	•	•	•	•					

Ocean Only

24. Effective_Optical_Depth_Average_Ocean (7)	•	•	•	•	•	•					
25. Optical_Depth_Ratio_Small_Ocean	•	•	•	•	•	•					

105. Atmospheric_Water_Vapor_High	•	•	•	•	•	•														
106. Retrieved_Temperature_Profile	•	•	•	•																
 <p>Earth Observing System MODIS Atmosphere Level-3 Daily Product</p> <p>PARAMETER</p> <p>STATISTIC</p>																				
	Mean	Standard_Deviation	Minimum	Maximum	OA_Mean	OA_Standard_Deviation	Fraction	Log_Mean	Log_Standard_Deviation	OA_Log_Mean	OA_Log_Standard_Deviation									

- ¹ Regressions are against Corrected Optical Depth Land (1st or 0.47 μm solution)
- ² Regressions are against Effective Optical Depth Average Ocean (2nd or 0.55 μm solution)
- ³ Regressions are against Effective Optical Depth Average Ocean (4th or 0.87 μm solution)
- ⁴ Joint Histogram with Cloud_Effective_Radius_1L_Liquid
- ⁵ Joint Histogram with Cloud_Effective_Radius_1L_Ice
- ⁶ Joint Histogram with Cloud_Effective_Radius_1621_Liquid
- ⁷ Joint Histogram with Cloud_Effective_Radius_1621_Ice
- ⁸ Joint Histogram uses ISCCP-like bin boundaries

Appendix B:

**Scientific Data Sets
common to both the
L3 Daily
and
L3 Multiday Files**

Appendix C:

**Scientific Data Sets
found in the
L3 Multiday File
only**



Mean_Mean
Mean_Std
Mean_Min
Mean_Max
Std_Deviation_Mean
OA_Mean_Mean
OA_Mean_Std
OA_Mean_Min
OA_Mean_Max
OA_Std_Deviation_Mean
FMean
FStd
Log_Mean_Mean
Log_Mean_Std
Log_Mean_Min
Log_Mean_Max
Log_Std_Deviation_Mean
OA_Log_Mean_Mean
OA_Log_Mean_Std
OA_Log_Mean_Min
OA_Log_Mean_Max
OA_Log_Std_Deviation_Mean

- 1 Regressions are against Corrected Optical Depth Land (1st or 0.47 μm solution)
- 2 Regressions are against Effective Optical Depth Average Ocean (2nd or 0.55 μm solution)
- 3 Regressions are against Effective Optical Depth Average Ocean (4th or 0.87 μm solution)
- 4 Joint Histogram with Cloud_Effective_Radius_1L_Liquid
- 5 Joint Histogram with Cloud_Effective_Radius_1L_Ice
- 6 Joint Histogram with Cloud_Effective_Radius_1621_Liquid
- 7 Joint Histogram with Cloud_Effective_Radius_1621_Ice
- 8 Joint Histogram uses ISCCP-like bin boundaries

Appendix D:
Histogram Parameters
&
Bin Boundaries

Histogram Parameters and Bin Boundaries

Identical in Daily, Eight Day & Monthly Statistics

Black = Collection 005 and 051

Aerosol Product (04_L2)

Land Only

Parameter	Bin Boundaries *	Bins	Units
1. Corrected_Optical_Depth_Land	-0.1, 0.1, 0.3, 0.6, 1.5, 5.0	5	none
2. Optical_Depth_Ratio_Small_Land	0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 3.0	6	none
3. Mass_Concentration_Land	1.0, 10.0, 100.0, 1000.0, 10000.0	4	10 ⁻⁶ g cm ⁻²
4. Angstrom_Exponent_Land	-0.5, 0.0, 1.0, 2.0, 3.0, 5.0	5	none
5. Mean_Reflectance_Land_All	-0.1, 0.1, 0.3, 0.6, 1.5, 5.0	5	none
6. Deep_Blue_Aerosol_Optical_Depth_550_Land	0.0, 0.1, 0.3, 0.6, 1.5, 5.0	5	none
7. Deep_Blue_Angstrom_Exponent_Land	-0.5, 0.0, 1.0, 2.0, 3.0, 5.0	5	none
8. Deep_Blue_Single_Scattering_Albedo_Land	0.7, 0.86, 0.88, 0.90, 0.92, 0.94, 0.96, 1.00	7	none

Ocean Only

Parameter	Bin Boundaries *	Bins	Units
1. Effective_Optical_Depth_Average_Ocean	-0.1, 0.1, 0.3, 0.6, 1.5, 5.0	5	none
2. Optical_Depth_Ratio_Small_Ocean	0.0, 0.2, 0.4, 0.6, 0.8, 1.0, 3.0	6	none
3. Effective_Radius_Ocean	0.0, 0.1, 0.3, 0.5, 1.0, 2.0, 3.0, 4.0, 5.0	8	µm
4. Mass_Concentration_Ocean	1.0, 10.0, 100.0, 1000.0, 10000.0	4	10 ⁻⁶ g cm ⁻²
5. Cloud_Condensation_Nuclei_Ocean	1.0, 10.0, 10E2, 10E3, 10E4, 10E5, 10E6, 10E7, 10E8, 10E9, 10E10	10	cm ⁻²
6. Angstrom_Exponent_1_Ocean	-0.5, 0.0, 1.0, 2.0, 3.0, 5.0	5	none
7. Angstrom_Exponent_2_Ocean	-0.5, 0.0, 1.0, 2.0, 3.0, 5.0	5	none
8. Asymmetry_Factor_Average_Ocean	0.0, 0.3, 0.5, 0.7, 0.8, 0.9, 1.0	6	none
9. Backscattering_Ratio_Average_Ocean	0.0, 0.10, 0.15, 0.20, 0.30, 0.40, 0.50	6	none

Water Vapor Product (05_L2)

Parameter	Bin Boundaries *	Bins	Units
1. Water_Vapor_Near_Infrared_Clear	0.0, 0.05, 0.10, 0.15, 0.20, 0.30, 0.35, 0.40, 0.50, .0.60, 0.70, 0.80, 0.90, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 5.5, 6.0, 7.0, 8.0, 9.0, 10.0, 11.0, 12.0, 13.0, 15.0, 20.0	44	cm

2. Water_Vapor_Near_Infrared_Cloud	0.0, 0.05, 0.10, 0.15, 0.20, 0.30, 0.35, 0.40, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0, 1.2, 1.4, 1.6, 1.8, 2.0, 2.2, 2.4, 2.6, 2.8, 3.0, 3.2, 3.4, 3.6, 3.8, 4.0, 4.2, 4.4, 4.6, 4.8, 5.0, 10.0	34	cm
------------------------------------	--	----	----

Cloud Product (06_L2)

Cirrus Detection

Parameter	Bin Boundaries *	Bins	Units
1. Cirrus_Reflectance	0.0, 0.005, 0.01, 0.015, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.12, 0.14, 0.16, 0.18, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.60, 0.70, 0.80, 0.90, 1.0, 1.2, 1.6	30	<i>none</i>

Cloud Top Properties

Parameter	Bin Boundaries *	Bins	Units
1. Cloud_Top_Pressure 2. Cloud_Top_Pressure_Day 3. Cloud_Top_Pressure_Night	1, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100	11	hPa
4. Cloud_Top_Temperature 5. Cloud_Top_Temperature_Day 6. Cloud_Top_Temperature_Night	190, 200, 210, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290	17	K
7. Cloud_Effective_Emissivity 8. Cloud_Effective_Emissivity_Day 9. Cloud_Effective_Emissivity_Night	0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0	10	<i>none</i>
10. Cloud_Fraction 11. Cloud_Fraction_Day 12. Cloud_Fraction_Night	0.0, 0.2, 0.4, 0.6, 0.8, 1.0	5	<i>none</i>
15. Cloud_Phase_Infrared 16. Cloud_Phase_Infrared_Day 17. Cloud_Phase_Infrared_Night	0, 1, 2, 3, 4, 5, 6 <i>(note: 0=clear, 1=water, 2=ice, 3=mixed, 6=undetermined phase)</i>	7	<i>none</i>

Cloud Optical Properties

Parameter	Bin Boundaries *	Bins	Units
18. Cloud_Optical_Thickness_Liquid 35. Cloud_Optical_Thickness_1L_Liquid 49. Cloud_Optical_Thickness_1621_Liquid	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 60, 70, 80, 90, 100	45	<i>none</i>
19. Cloud_Optical_Thickness_Ice 36. Cloud_Optical_Thickness_1L_Ice 50. Cloud_Optical_Thickness_1621_Ice	0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, 50, 60, 70, 80, 90.0, 100.0	30	<i>none</i>
22. Cloud_Effective_Radius_Liquid 37. Cloud_Effective_Radius_1L_Liquid	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 22.0, 24,	23	μm

51. Cloud_Effective_Radius_1621_Liquid	26, 28, 30		
23. Cloud_Effective_Radius_Ice 38. Cloud_Effective_Radius_1L_Ice 52. Cloud_Effective_Radius_1621_Ice	5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 90	13	µm
31. Cloud_Water_Path_Liquid 39. Cloud_Water_Path_1L_Liquid 53. Cloud_Water_Path_1621_Liquid	0, 10, 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 2000	14	g m ⁻²
32. Cloud_Water_Path_Ice 40. Cloud_Water_Path_1L_Ice 54. Cloud_Water_Path_1621_Ice	0, 10, 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 1000, 2000, 4000	15	g m ⁻²

Atmosphere Profile Product (07_L2)

Parameter	Bin Boundaries *	Bins	Units
1. Total_Ozone	0, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500	10	Dobson Units
2. Total_Totals	0, 10, 20, 30, 40, 50, 60, 70, 80	8	K
3. Lifted_Index	-20, -15, -10, -5, 0, 5, 10, 15, 20, 25, 30, 35, 40	12	K
3. Atmosphere_Water_Vapor 4. Atmosphere_Water_Vapor_Low 5. Atmosphere_Water_Vapor_High	0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10	10	cm

** Note: 1st bin includes values equal to both the low and high bin boundary,
2nd through the nth bin includes values equal to the high bin boundary only*

Appendix E:

**Joint Histogram Parameters
&
Bin Boundaries**

Joint Histogram Parameters and Bin Boundaries

Identical in Daily, Eight Day & Monthly Statistics

Black = Collection 005 and 051

Parameter 1	(Bins): Bin Boundaries *	Parameter 2	(Bins): Bin Boundaries *	D/N
<i>Liquid Water Clouds</i>				
Optical thickness	(11): 0, 2, 4, 6, 8, 10, 15, 20, 30, 40, 50, 100	Effective radius (μm)	(10): 2, 4, 6, 8, 10, 12.5, 15, 17.5, 20, 25, 30	D
Optical thickness <u>Single Layer</u>	"	Effective radius (μm) <u>Single Layer</u>	"	D
Optical thickness <u>1621</u>	"	Effective radius (μm) <u>1621</u>	"	D
Optical thickness	"	Cloud top temperature (K)	(12): 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290	D
Optical thickness	"	Cloud top pressure (hPa)	(13): 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1100	D
Effective radius (μm)	(10): 2, 4, 6, 8, 10, 12.5, 15, 17.5, 20, 25, 30	Cloud top temperature (K)	(12): 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290	D
Effective radius (μm)	"	Cloud top pressure (hPa)	(13): 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1100	D
<i>Ice Clouds</i>				
Optical thickness	(11): 0, 0.5, 1, 2.5, 5, 7.5, 10, 15, 20, 30, 50, 100	Effective radius (μm)	(13): 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 90	D
Optical thickness <u>Single Layer</u>	"	Effective radius (μm) <u>Single Layer</u>	"	D
Optical thickness <u>1621</u>	"	Effective radius (μm) <u>1621</u>	"	D
Optical thickness	"	Cloud top temperature (K)	(13): 190, 200, 210, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270	D
Optical thickness	"	Cloud top pressure (hPa)	(16): 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1100	D
Optical thickness	"	Effective emissivity	(10): 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0	D
Optical thickness <u>Single Layer</u>	"	Effective emissivity	(10): 0.0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0	D
Effective radius (μm)	(13): 5, 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 90	Cloud top temperature (K)	(13): 190, 200, 210, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270	D

Effective radius (μm)	"	Cloud top pressure (hPa)	(16): 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 700, 800, 900, 1000, 1100	D
------------------------------------	---	--------------------------	--	---

All Clouds

Parameter 1	(Bins): Bin Boundaries *	Parameter 2	(Bins): Bin Boundaries *	D/N
Cloud top pressure (hPa)	(3): 1, 400, 700, 1000	Effective emissivity	(3): 0.0, 0.5, 0.95, 1.0	D
Cloud top pressure (hPa)	"	Effective emissivity	"	N
Cloud top pressure (hPa)	"	Effective emissivity	"	D+N
Cloud top temperature (K)	(17): 190, 200, 210, 220, 225, 230, 235, 240, 245, 250, 255, 260, 265, 270, 275, 280, 285, 290	Cloud Phase Optical Properties	(3): Water, Ice, Undetermined	D
Cloud top temperature (K)	"	Cloud Phase IR	(4): Water, Ice, Mixed, Uncertain	D
Cloud top temperature (K)	"	Cloud Phase IR	"	N
Optical thickness (ISCCP boundaries)	(7): 0, 0.3, 1.3, 3.6, 9.4, 23, 60, 100	Cloud top pressure (hPa) (ISCCP boundaries)	(7): 0, 180, 310, 440, 560, 680, 800, 1100	D

** Note: 1st bin includes values equal to both the low and high bin boundary, 2nd through the nth bin includes values equal to the high bin boundary only*