BASIC TERRA FUSION PRODUCT ALGORITHM THEORETICAL BASIS AND DATA SPECIFICATIONS

Guangyu Zhao[1]
Muqun Yang[2]
Landon Clipp[3]
Yizhao Gao[4]
H. Joe Lee[2]
Larry Di Girolamo[1]

[1] Department of Atmospheric Sciences, University of Illinois at Urbana-Champaign
[2] The HDF Group
[3]Department of Electric and Computer Engineering, University of Illinois at Urbana-Champaign
[4]Department of Geography and Geographic Information Science in University of Illinois at Urbana-Champaign

Table of Contents

GLOSSARY OF ACRONYMS

A
ACCESS (Advancing Collaborative Connections for Earth System Science)
ASTER (Advanced Spaceborne Thermal Emission and Reflection Radiometer)

B
BF (Basic Fusion)

C
CERES (Clouds and Earth's Radiant Energy System)
CF (Climate and Forecast)

D
DAAC (Distributed Active Archive Centers)
DOI (Digital Object Identifiers)

E
EOSDIS (Earth Observing System Data and Information System)

H
HDF (Hierarchical Data Format)

I
IFOV (Instantaneous Field of View)

J
JPL (Jet Propulsion Laboratory)

M
MISR (Multi-angle Imaging SpectroRadiometer)
MODIS (Moderate-resolution Imaging Spectroradiometer)
MOPITT (Measurements of Pollution in the Troposphere)

N
NASA (National Aeronautics and Space Administration)
NCSA (National Center for Supercomputing Applications)

S
SDS (Scientific Datasets, multidimensional array of data in HDF)

1. INTRODUCTION

1.1 Purpose

The basic Terra fusion product provides general atmospheric and surface research community a unique temporally-fused set of radiance measurements from all the Terra instruments, namely, the Moderate-resolution Imaging Spectroradiometer (MODIS), the Multi-angle Imaging SpectroRadiometer (MISR), the Advanced Spaceborne Thermal Emission and Reflection Radiometer (ASTER), the Clouds and Earth's Radiant Energy System (CERES), and the Measurements of Pollution in the Troposphere (MOPITT). This product contains (1) radiance values of IOFVs (pixels) for each spectral band at a native resolution for each instrument, (2) their quality flags associated with radiance values, (3) their latitude and longitude information at a native resolution, (4) time of observations, (5) instrument viewing geometry, and (6) solar position.

The intent of this document is to identify and describe sources of the input data, provide the physical theory and mathematical background underlying the derivation of the high-resolution geolocation fields, and describe procedures in data progressing and performance tuning, along with file specifications. To fulfill the requirement of the NASA ACCESS project (**NNH15ZDA001N-ACCESS**), this document is to establish requirements and functionality of the data processing software.

1.2 Scope

This document covers the algorithm theoretical basis and data product specifications for the basic fusion product that is generated at the National Center for Supercomputing Applications (NCSA) at the University of Illinois at Urbana-Champaign. Chapter 1 describes the purpose and scope of the document. Chapter 2 provides a brief overview of this experiment. The processing concept and algorithm description are presented in Chapter 3. Chapter 4 describes the file specifications, and assumptions and limitations are summarized in Chapter 5.

Literature references are indicated by a number in italicized square brackets (e.g., [1]).

[1] MISR Data Products Specifications, JPL D-13963
[2] MODIS Level 1B Product User's Guide, PUB-01-U-0202- REV B
[3] ASTER L1T Product User's Guide, Version 1.0
[4] MOPITT L1B Algorithm Theoretical Basis Document
[5] CERES Single Satellite Footprint TOA/Surface Fluxes and Clouds (SSF) Collection Document

1.3 Revisions

This is original version of the document

2. EXPERIMENT OVERVIEW

2.1 Terra Instruments

Terra is the flagship of NASA's Earth Observing System (EOS). It was launched into orbit on December 18, 1999 and carries five instruments: MODIS, MISR, ASTER, CERES, and MOPITT. The mission remains healthy, continues to receive extremely high ratings from NASA's Senior Review, and carries enough fuel to maintain its current 10:30 am ECT sun- synchronous orbit until 2022. Terra continues to enable scientists to address fundamental questions from NASA's Science Plans, including each of the six Earth Science Research Focus Areas in the latest 2014 Science Plan. Terra is currently one of the longest single-platform satellite record for studying Earth, making it one of our most valuable satellite record for examining Earth's climate and climate change. It is also amongst the most popular NASA EOS datasets. In 2014 alone, more than 230 million files totaling more than 2.2 PB were delivered to more than 100,000 users around the world, resulting in more than 1,600 peer-reviewed publications, and citing other Terra research more than 41,000 times. These metrics have maintained an approximate exponential growth rate since launch. The Terra data serves not just the scientific community, but also government, commercial, and educational communities.

2.2 Objective of Terra Product Generation

The strength of the Terra mission has always been rooted in its five instruments and the ability to fuse the instrument data together for obtaining greater quality of information for Earth Science compared to individual instruments alone. As the data volume grows and the central Earth Science questions shift from process-oriented to climate-oriented questions, the need for data fusion and the ability for scientists to perform large-scale analytics with long records have never been greater. The challenge is particularly acute for Terra, given its growing volume of data (> 1 petabyte), the storage of different instrument data at different archive centers, the different file formats and projection systems employed for different instrument data, and the inadequate cyberinfrastructure for scientists to access and process whole-mission fusion data (including Level 1 data). Sharing newly derived Terra products with the rest of the world also poses challenges.

Our objective is to transfer approximately 1 PB of the mission-wide georectified and radiometric calibrated radiance datasets (L1B) of all the Terra instruments staged across three different DAACs to NCSA and build the necessary tool to create the Basic Fusion (BF) product that merges these L1B granules for all the Terra instruments into one granule.

2.3 Basic Fusion Strategy

We intend to reserve the contents and structures of the datasets in their original product granules as much as possible in the BF product. The contents of a single fusion granule will include: (1) radiance values of IOFVs (pixels) for each spectral band at a native resolution for each instrument, (2) their quality flags associated with radiance values, (3) their latitude and longitude information at a native resolution, (4) time of observations, (5) instrument viewing geometry, and (6) solar position. As for content (1), except for MOPITT and CERES, the radiance values need to be converted from digital numbers stored as integers in the original product granules by using the scale and offset values as well as gain setting imbedded in metadata/attributes. For content (3), the geolocation information (latitude and longitude) is not provided at a pixel level for all of the native resolutions for ASTER, MISR, and MODIS. This information is given at a coarse resolution either in the L1B granules as separate fields or in a separate product from the L1B granules. For example, latitude and longitude at 250m and 500m resolutions for MODIS, 275m resolution for MISR, and all the resolution levels for ASTER need to be interpolated from coarse resolution latitude and longitude information provided in the original products.

The reprocessed L1B granules for each instrument will be merged and packed into one fusion granule. After evaluating the storage settings of Blue Waters, processing approach, application programs and distribution strategies, we choose Terra orbit as the granularity of the BF product. The BF granules are stored in the HDF5 format, which supports high performance parallel I/O with no limitation of file size and the dataset size or the number of the objects.

3. ALGORITHM DESCRPTION

3.1 Processing Outline

Processing flow concepts are shown diagrammatically throughout the document.

The convention for the various elements displayed in these diagrams is shown in Figure 1.

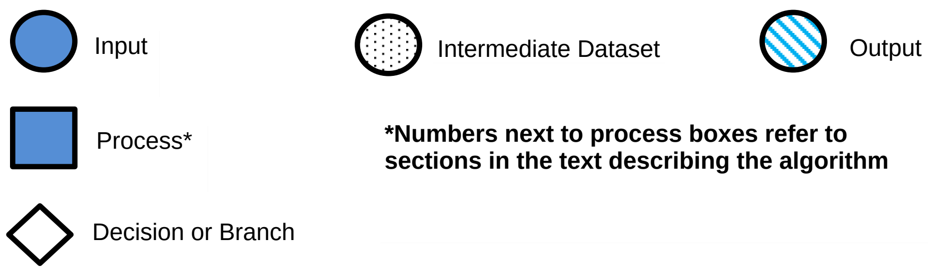

Figure 1. Conventions used in processing flow diagrams

Overviews of the processing flow concept are shown in Figures 3.1

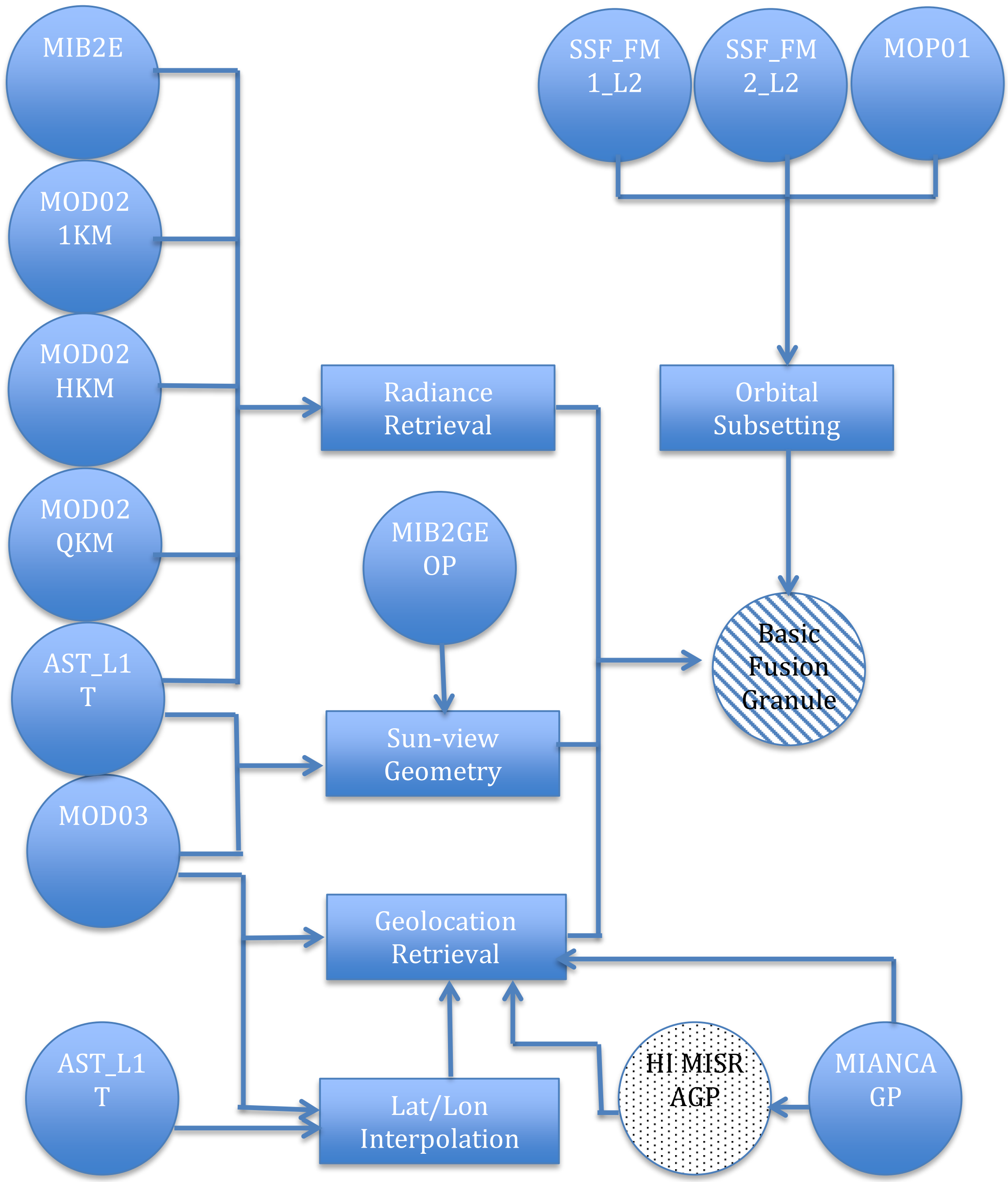

Figure 3.1. Processing flow chart, The DOI and version number for each of the Terra product IDs listed in the input diagram are given in Table 3.1. "HI MISR AGP" derived from the MISR AGP product contains latitude and longitude information for the MISR pixels at a 275m resolution (see section 3.3.4 for details).

3.2 Input Files

A complete list of the EOSDIS DOIs of all of the input products, which include the radiance datasets and ancillary files for all of the Terra instruments that are fed into the basic fusion software, is given in the Table 3.1. The DOI system provides a persistent link to a detailed description of each input product located at the NASA EOSDIS' websites.

Table 3.1. A list of DOIs of all the input products

Instrument	Product DOIs
ASTER	10.5067/ASTER/AST_L1T.003
CERES	10.5067/TERRA/CERES/SSF_Terra-FM1_L2.004A 10.5067/TERRA/CERES/SSF_Terra-FM2_L2.004A
MISR	10.5067/Terra/MISR/MI1B2E_L1.003 10.5067/TERRA/MISR/MIANCAGP_Ancillary.001 10.5067/Terra/MISR/MIB2GEOP_L1.002
MODIS	10.5067/MODIS/MOD02QKM.006 10.5067/MODIS/MOD02HKM.006 10.5067/MODIS/MOD021KM.006 10.5067/MODIS/MOD03.006
MOPITT	10.5067/TERRA/MOPITT/MOP01_L1.007

3.3 Theoretical Descriptions

3.3.1 Subsetting by Terra orbits

The granularity of the BF product is chosen to be one Terra orbit in accordance with the granularity of the MISR radiance product. Factors also taken into account for this choice include the I/O performance, processing speed, memory usage and transfer rate based on the cyberinfrastructure and specifications of computational facilities at NCSA, where the BF product is produced, processed, and staged. The size of one orbital BF file typically ranges between 20 GigaBytes (GB) and 50 GB with the in-memory compression scheme applied to most fields.

The starting and ending time of Terra orbits were generated using the MISR toolkit developed by JPL (version 1.4.1 available for download from The Open Channel Foundation http://www.openchannelsoftware.org/projects/MISR_Toolkit). One granule of the BF product contains 1, ~20, 2-3, and 1-1 granules of the MISR, MODIS, CERES and MOPITT radiance products. The number of the ASTER granules stored in the BF product vary from one granule to another, depending on the collection mode of the ASTER instrument, who cameras primarily open over land and remain closed over ocean.

The temporal information stored in the original Terra instrument granules is used to calculate the associated orbit number that each of the granules is ascribed to. For ASTER and MODIS, the data fields for their entire granules will be incorporated into a BF granule without any sub-setting if and only if the starting time of their granules falls within the starting and ending time of the orbit of the BF granule.

Only CERES and MOPITT products provide the time stamps for all of the pixels at their native resolutions. After converting their time format into Coordinated Universal Time (UTC) format, only pixels whose time stamp are within the starting and ending time of an orbit are included into the granule for the orbit. Subletting CERES data fields, however, turns out not always following our original assumption that the observed time is stored in a monotonically temporal order in a dataset. This assumption does not hold true for data which were collected when the CERES instruments are in the biaxial mode. Therefore, some CERES radiance data fused in one orbit may not be necessarily belong to that orbit. Nevertheless, the current algorithm still ensures the monotonic order of the first and the last time stamp in one orbit and the time stamps prior and next to them. In addition, there are no missing valid CERES radiance data although some data may be misplaced to an orbit neighboring to the orbit they should belong to.

The orbit starting time and ending time were generated using the MISR toolkit as mentioned in section 3.3.1. The orbit for a BF granule may or may not match the orbit provided in the metadata for some of the ASTER and MODIS granules, as long as the starting time of their granules falls within the starting and ending time of the orbit of the BF granule. This does not affect the subsetting accuracy since the starting and ending time of a ASTER or MODIS granule contained its filename is used to determine whether the granule is ascribed to an orbit.

3.3.2 Radiance Conversion

Except for CERES and MOPITT, the Level-1B radiance granules for the Terra instruments contain 8-bit or 16-bit scaled integer representation of the calibrated digital signals instead of physical radiance values in a floating-point format. In the BF product, these digital signals have been converted to radiance using scale factors and offsets written as attributes in the original granules, and they have been stored as a single-precision floating-point format.

The conversion formulas and procedures used for MISR, MODIS and ASTER are documented in details in the MISR Level-1 Radiance Scaling and Conditioning Algorithm Theoretical Basis [1] (available for download at https://eospso.nasa.gov/sites/default/files/atbd/atbd-misr-01.pdf), the MODIS Level 1B Product User's Guide [2](https://mcst.gsfc.nasa.gov/sites/mcst.gsfc/files/file_attachments/M1054.pdf), and the ASTER L1T Product User's guide [3] (https://lpdaac.usgs.gov/sites/default/files/public/product_documentation/aster_l1t_users_guide.pdf), respectively. In brief, the MISR radiance was obtained from the 16-bit integer Radiance/RDQI field by right-shifting 2 bits, then multiplying the results by the scale factor contained in the grid metadata. For MODIS, the radiance was calculated by multiplying the difference between the 16-bit integer Digital Numbers (DN) and offset value by a scale factor. Both the scale factor and offset values are provided as SDS attributes in the MODIS L1B product. The ASTER radiance was converted from the 8-bit integer DN by subtracting it by 1 than multiplying the results by unit conversion coefficient specified for each spectral bands and gain setting.

3.3.3 Quality Flags

The data fields that contain quality flags for radiance values in the original MODIS, ASTER, CERES and MOPITT granules are directly copied into the BF product.

For MISR, the quality flags, which are called Radiometric Data Quality Indicator (RDQI), are encoded in 16-bit integers along with scaled radiance values. These quality flags were decoded first following the steps described in in the MISR Level-1 Radiance Scaling and Conditioning Algorithm Theoretical Basis [1]. However, the RDQI is not directly stored as an individual data fields in the BF product. Instead, only the spatial-index location of the pixels with the RDQI equal to1(reduced accuracy measurement) are stored as a separate data field. The purpose of doing this is to save storage space given that the majority of the MISR radiance pixels are high quality and having a RDQI value of zero. The radiance values for the pixels with RDQI larger than one are considered either "Not usable for science" or "Unusable for any propose" [1]. The radiance values for such pixels are set to -999.0. The radiance values for the pixels whose 16-integer scaled radiance values equal to 16378 (out of bound) or 16380 (high RDQI) are also set to -999.0.

3.3.4 Derivation of Latitude and Longitude at Native Resolution

The latitude and longitude for each pixel at its native resolution for all of the radiance fields is provided in the Basic Fusion (BF) product, following the same conventions where latitude ranges between -90 and 90 degrees and longitude ranges between -180 and 180 degrees. For MOPITT, this information is given in their radiance products, from which their geolocation fields are directly copied into the BF product without any modifications. For CERES, colatitude instead of latitude is given in the original radiance dataset and longitude ranges between 0 and 360 degrees. The CERE latitude and longitude are converted to conform the same conventions as the other instruments before being packed in the BF product.

MISR geolocation information is only provided at a resolution of 1.1km in the MISR Ancillary Geographic Product (AGP). There is no publicly available MISR product that provides geolocation information at a resolution of 275m, at which the radiance data for all of the bands for the MISR nadir camera and the red band for all of the off-nadir cameras are collected. Because the MISR data are stored in the Space Oblique Mercator (SOM) grids, the geolocation of a 275m pixel can be mathematically calculated given its orbit number, line, sample and block number. The MISR toolkit is used to calculate latitude and longitude at a resolution of 275m resolution for each of the 233 MISR paths. The results are stored as the MISR HI AGP files in an HDF4 format in the same way as how the geolocation fields are stored in the MISR AGP product.

The MODIS MOD03 product contains geolocation fields at a 1km resolution, but not at 250 and 500m resolutions, which have to be derived mathematically. Based on the co-registration arrangement of MODIS cells (Figure X1, Gumley *et al*. 2003), a bilinear interpolation is used to calculate the coordinates of 500m-resolution pixels from the 1000m resolution geolocation fields. The same procedure was repeated to achieve the 250m-resolution geolocations from 500m-resolution ones. Bilinear interpolation is a method to interpolated the value at a specific location based on the values of its four neighboring points from a rectilinear 2D grid. Counterintuitively, in this application, the latitudes and the longitudes are the values to be interpolated, while the input locations in the interpolation are the relative pixel counts (e.g, 0.25 pixels along line direction and 0.5 pixel along sample direction). In a bilinear interpolation, as shown in Figure3.2, the value at a new location P is estimated based on values of four neighboring points (A_{11}, A_{12}, A_{21}, A_{22}) using a two-phase linear interpolation. First, the value at B_1 is linearly interpolated using values at A_{11} and A_{21} based on the length of $A_{11}B_1$ and B_1A_{21}, and the value at B_2 is linearly interpolated using values at A_{12} and A_{22}. Then the value at P is linearly interpolated using values at B_1 and B_2. Suppose $f = \frac{|A_{11}B_1|}{|A_{11}A_{21}|} = \frac{|A_{12}B_2|}{|A_{12}A_{22}|}$ and $f = \frac{|B_1P|}{|B_1B_2|}$. The value at P (V_p) can be estimated from V_{11}, V_{21}, V_{12} and V_{22}, as

$$V_P = [1-f \quad f]\begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix}\begin{bmatrix} 1-g \\ g \end{bmatrix} \tag{3.1}$$

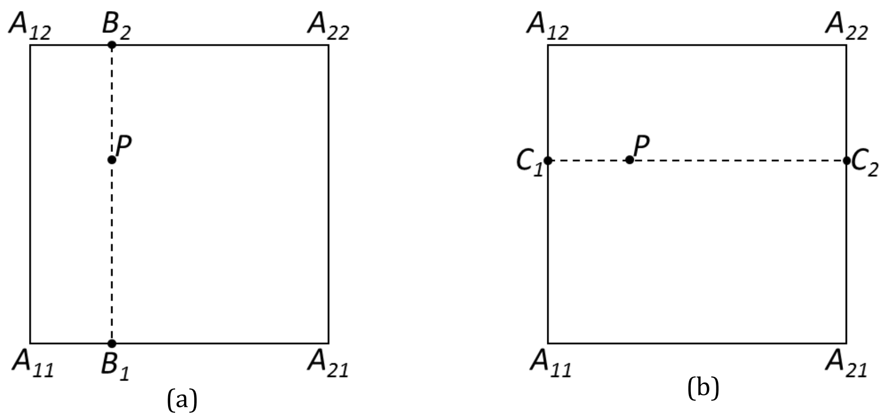

Figure 3.2 An illustration of bilinear interpolation to calculate the value at P using all the values from neighboring four points A_{11},A_{12}, A_{22}, and A_{21} with a two-step approach shown as (a) and (b).

Using the latitudes and longitudes as values in conventional bilinear interpolation is problematic on a sphere. The average of latitudes and longitudes of two points is different from the midpoint of these two locations. As a result, a pseudo bilinear interpolation based on spherical surface is used as an alternative. Rather than using a linear interpolation to calculate the latitudes and longitudes of B_1 (B_2 and P), the new latitudes and longitudes are calculated as the interpolation points along the great circle arc $A_{11}A_{21}$ ($A_{21}A_{22}$ and B_1B_2). The procedure to calculate the spherical interpolation point is shown below.

If the two end points of an spherical arc can be expressed as P_1(latitude φ_1, longitude λ_1) and P_2(latitude φ_2, longitude λ_2), we can then calculate the location of a new point P_{New}(latitude φ_{New}, longitude λ_{New}) at fraction f along the great circle arc (e.g., f=0 when P_{New} is at P_1, f=1 when P_{New} is at P_2). First, the angular distance θ between P_1 and P_2 are calculated using the haversine formula:

$$\theta = 2\arcsin\sqrt{\sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos\varphi_1 * \cos\varphi_2 * \sin^2\left(\frac{\Delta\lambda}{2}\right)} \tag{3.2}$$

where $\Delta\varphi = \varphi_1 - \varphi_2$, and $\Delta\lambda = \lambda_1 - \lambda_2$. Then the new coordinates φ_{New} and λ_{New} can be calculated:

$$a = \frac{\sin((1-f)*\theta)}{\sin\theta} \tag{3.3}$$

$$b = \frac{\sin(f\theta)}{\sin\theta} \tag{3.4}$$

$$x = a * \cos\varphi_1 * \cos\lambda_1 + b * \cos\varphi_2 * \cos\lambda_2 \tag{3.5}$$

$$y = a * \cos\varphi_1 * \sin\lambda_1 + b * \cos\varphi_2 * \sin\lambda_2 \quad (3.6)$$
$$z = a * \sin\varphi_1 + b * \sin\varphi_2 \quad (3.7)$$
$$\varphi_{New} = \text{atan}\,2\left(z, \sqrt{x^2 + y^2}\right) \quad (3.8)$$
$$\lambda_{New} = \text{atan}\,2(y, x) \quad (3.9)$$

This method can also be used for extrapolation, when $f < 0$ or $f > 1$. The extrapolation is used to estimate the first and last row, and the last column of each scan.

For a bilinear interpolation, it does not matter whether the value at P is estimated from B_1 and B_2, or C_1 and C_2. For the pseudo bilinear interpolation based on spherical surface, the two results may differ very slightly. The difference, however, is extraordinarily small, since for both MODIS, the four sides of the four cornering points are almost identical in length.

There also does not exist any ASTER products that provide geolocation information for each of the ASTER radiance pixels at their native resolutions of 15, 30, 90m. For each ASTER granule, only a 11×11 grid of latitudes and longitudes are given for uniformly-spaced line and sample locations covering the entire ASTER image. The $(1{,}1), (1{,}11), (11{,}1)$ and $(11{,}11)$ points in the 11×11 grids correspond to the pixel centers four cornering pixels of the image. The same bilinear interpolation methods used to calculate the MODIS geolocation fields at 500m and 250m resolutions as descripted in Equations 3.1-3.9 is used to compute the ASTER geolocations for pixels at resolutions of 15, 30, and 90m, respectively.

3.3.5 Sun-View Geometry Fields

All of data fields containing sun-view geometry information either from the original L1B products or ancillary products are directly copied into the BF product without any modification. The sun-view geometry information includes solar zenith angle, solar azimuth angle, viewing zenith angle and viewing azimuth angle.

3.3.6 Data Storage Format and compression scheme

The storage format of the BF product is chosen to be HDF5. HDF5 employs in-memory compression, multidimensional extensible datasets, and chunking technologies to improve access, management, and storage efficiency. The HDF5 format and library doesn't set restriction to the file size and the number of objects in an HDF5 file. This enables the HDF5 store variables with much bigger size and many objects in one file, which is exactly the case for the BF product. Because of the support of the group hierarchy, the HDF5 library makes it straightforward group the non-trivial number of physical and geolocation fields of the five instruments to one HDF5 file. Furthermore, MPI-IO, other rich optimization features and the potential support for the cloud environment inside the HDF5 library make the implementation of the IO module of the BF analysis programs less difficult.

To cater for broad user communities, the file structure of the BF product was constructed to mostly comply with Climate and Forecast(CF) conventions, which follow the netCDF-4 data model enabling NetCDF4 tools to access and explore the contents of the BF product. A detailed description of the CF conventions is available at http://cfconventions.org. The CF conventions have been widely used both in atmospheric modelling and remote sensing communities, mainly because the CF conventions make the data interoperability easily achieved. Detailed information on the CF metadata in a BF file can be found in section 4.3.

The total size of 16 years of the BF granules generated without using any compression scheme is close to 9 Petabytes. To reduce the BF file size, we apply the deflate lossless compression scheme on most of the radiance and geo-location fields for MISR, ASTER and MODIS. To use the compression feature in HDF5, data arrays must be split into chunks first. The data in each chunk is then compressed and stored separately in the file. To optimize the IO performance, we choose the chunk shape to be the same as the shape and size of the radiance and geo-location arrays except the MODIS radiance fields. Each chunk for MOIDS radiance array is a subset of the original array size. It stores the MODIS radiance data per band. For CERES and MOPITT, we don't apply any compression scheme, since their data storage spaces are already small. With compression, the size of a BF granule has reduced by ~two thirds at the expenses of I/O performance, which decreases by nearly half accordingly.

3.4 Metadata production

NASA maintains metadata repository system called "Common Metadata Repository (CMR)" to allow users to search the data products distributed by NASA DAACs easily. There are two kinds of metadata that NASA CMR maintains - collection and granule. Collection metadata covers the shared information among granules for the same product. Granule metadata contains specific information for an individual data file. For the basic fusion product, collection metadata holds information such as who produced data, the contact information for data producers, and temporal/spatial coverage of the entire granules under the whole collection. Granule metadata describes the file contents of a granule. Therefore, granule metadata may vary significantly from one orbit to another depending on the orbit information, what products are fused, which datasets are available, and the quality of data inside the file.

The BF collection metadata was generated manually since only one collection metadata is necessary for the same product. The collection metadata information is stored in a single XML file. The storage structure and format in the XML file follows the ECHO10 schema that NASA CMR team provides. The BF collection metadata includes the existing collection level CMR record of the original Terra data products that have been fused into the BF product.

The BF granule metadata contains the basic fusion file size, file creation time and a list of all of the original input granule file names along with their NASA CMR information retrieved from the NASA CMR search engine. The content inside each input granule includes data quality information, temporal and spatial information, and sensor information etc. Since the granularity of the basic fusion product is the same as the MISR Level 1 products, the BF granule metadata has the similar layout to MISR. In total, 84303 granule metadata files in the XML format were generated. The granule metadata is still provided for the BF granules that have no valid radiance values even for all of the five Terra instruments.

3.5. Large Scale processing

The Basic Fusion program itself is entirely serial in that it takes advantage of no parallel libraries. One instance of the program is designed to generate a single granule of data, i.e. one Terra orbit. Because of the large number of orbits that must be processed (85,430 orbits in total), the program is executed in an embarrassingly (or pleasingly) parallel fashion to vastly decrease the time required to process the whole mission. The fact that there are no interdependencies between the jobs greatly increases the ease of processing. The entire BF file set is processed on the Blue Waters supercomputer housed at the University of Illinois at Urbana-Champaign. Blue Waters provides a total of 362,240 AMD Bulldozer compute cores, more than 250 petabytes of Nearline archive tape storage and 26.4 petabytes of Online high-performance disk storage.

The processing of all the data heavily relies on three main components: the input data, the SQlite dataset, and the repackaging program itself. A detailed description of each component is given below:

The Basic Fusion program takes as one of its arguments a list of input files spanning one Terra orbit. The task of querying the list of files available for processing is delegated to an SQLite database. This database can be queried in a various number of ways, however for the purposes of this project queries are only performed using the Terra orbit number.

To generate the database itself, a Python script was written that parses a raw, unordered text file containing all of the existing MOPITT, MISR, ASTER, CERES and MODIS files, as well as a text file containing the start and ending times of all Terra orbits. The data products used for each of the instruments all have different file naming conventions as well as different file granularities. The Python script must parse each filename and determine that file's start time, end time and absolute directory, storing that information into one master table. The start times for some of the instruments are explicitly given, making it very easy to fill the start time record. However, some of the instruments only give orbit number or perhaps a simple date (as is the case for MOPITT). None of the instruments provide information on the file's end time in the file name itself, so the only way to determine the end time of a file, short of using HDF API calls to go inside the file itself, is to infer end time by using the published documentation on granularity for each instrument.

By storing the start and end time of each Terra orbit, the path number of each orbit, and the start and end time of each file, a series of useful SQLite calls can be constructed that take advantage of this information. As stated before, queries based on orbit number are the only type that are used for BF generation, but this does not limit future users to use their own queries if needed.

One of the requirements of the BF program is that the input text file has its HDF files listed in a specific and predictable way. Querying the database will not return the requested files in the correctly ordered way, so a script has been written that orders all of the files properly, also checking for all possible errors within the final input text file that might cause either the generation of an erroneous fusion file or an unrecoverable program crash downstream. The details of how the input file must be ordered can be found on the Basic Fusion GitHub page.

4. OUTPUT FILE SPECIFICATIONS

4.1 File naming conventions

The BF product is composed of the file granules with names constructed as "Terra BF L1B short name"_"Orbit Number"_"Start date and time of an orbit in UTC"_ "Software update version number"_"Collection version number". Table 4.1 provides example values of these fields.

Table 4.1. File naming convention

File Name Field	Format	Example Value
L1B Short Name	TERRA_BF_L1B	TERRA_BF_L1B
Oribt number	Oxxxxx	O68138
Start Date-Time-Group	YYYYMMDDhhmmss	20121009081300
Software update version	Ffff	F000
Collection version number	Vnnn	V001

4.2 Data variable descriptions

The majority of the data variable names and contents in a BF are directly copied from the original L1B products or associated ancillary products for all of the Terra instrument. Users are encouraged to refer to the references [1][2][3][4][5] for a detailed description of each data variable. The original data variables that have been modified in the process of the BF production and new data variables are described in the following tables 4.2-4.2.6.

4.2.1 ASTER

All the data fields for ASTER are stored under the root group name of "/ASTER" in a BF granule. One BF granule contains a variable number of the original ASTER L1T granules, each of which is stored as a separate and individual HDF5 subgroup, whose name is partially copied from the associated ASTER L1T file name, includes the starting time of the granule. For example, the subgroup name of granule_05032000141102 contains the data fields for the ASTER L1T granule having a starting time of 14:11:02 (UTC) on May 3, 2000.

Table 4.2 HDF data variables for each ASTER under the subgroup of /ASTER/granule_mmddyyyyhhxxss, where mmddyyyyhhmmss, stands for month (mm), date(dd), year(yyyy), hour(hh), minute(xx), and second(ss) of the starting time of data acquisition. The group path of "/ASTER/granule_mmddyyyyhhxxss" is abbreviated to ".../" in the table.

Path	Name	Dimension	Unit	Type	Description
.../VNIR	ImageData1	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
.../VNIR	ImageData2	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
.../VNIR	ImageData3N	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]

…/VNIR/Geolocation/	Latitude	Varies by scene	degrees_north	Float64	The same dimension as the radiance fields under …/VNIR at a resolution of 15m
…/VNIR/Geolocation	Longitude	Varies by scene	degrees_east	Float64	The same dimension as the radiance fields under …/VNIR at a resolution of 15m
…/SWIR	ImageData4	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
…/SWIR	ImageData5	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
…/SWIR	ImageData6	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
…/SWIR	ImageData7	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
…/SWIR	ImageData8	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
…/SWIR	ImageData9	Varies by scene	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[3]
…/SWIR/Geolocation/	Latitude	Varies by scene	degrees_north	Float64	The same dimension as the radiance fields under …/SWIR at a resolution of 30m
…/SWIR/Geolocation	Longitude	Varies by scene	degrees_east	Float64	The same dimension as the radiance fields under …/SWR at a resolution of 30m
…/TIR	ImageData10	Varies by scene	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[3]
…/TIR	ImageData11	Varies by scene	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[3]
…/TIR	ImageData12	Varies by scene	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[3]
…/TIR	ImageData13	Varies by scene	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[3]
…/TIR	ImageData14	Varies by scene	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[3]
…/TIR/Geolocation/	Latitude	Varies by scene	Degrees_north	Float64	The same dimension as the radiance fields under …/TIR at a resolution of 90m

…/TIR/Geolocation	Longitude	Varies by scene	Degrees_east[1]	Float64	The same dimension as the radiance fields under …/TIR at a resolution of 90m
.../Geolocation	Latitude	11 x 11	Degrees_north	Float64	Coarse resolution of latitude uniformly spaced to cover the entire scene
.../Geolocation	Longitude	11 x 11	Degrees_east	Float64	Coarse resolution of longitude uniformly spaced to cover the entire scene
…/Solar_Geometry	SolarAzimuth	1	Degree[1]	Float32	[3]
…/Solar_Geometry	SolarElevation	1	Degree	Float32	[3]
…/PointAngle	SWIR	1	Degree	Float32	[3]
…/PointAngle	SWIR	1	Degree	Float32	[3]
…/PointAngle	SWIR	1	Degree	Float32	[3]

4.2.2 CERES

All the data fields for CERES FM1 and FM2 are stored under the root group name of "/CERES/FM1" and "/CERES/FM2", respectively, in a BF granule. One BF graule contains two or three hourly CERES SSF granule files, each of which is stored as a separate and individual HDF5 subgroup, whose names are partially copied from their associated SSF file names, includes the starting time of the SSF file. For example, the CERES subgroup name of granule_200092705 contains the data fields for the CERES SSF granule having a starting time of 05:00 (UTC) on September 27, 2009. All of the data fields were directly copied from the CERES SSF product without any modifications.

Table 4.3 HDF data variables for CERES under the subgroup of /CERES/FM1/granule_yyyymmddhh or /CERES/FM2/granule_yyyymmddhh, where yyyymmddhh, stands for year(yyyy), ,month (mm), and hour(hh) of the starting time of data acquisition. The group path of "/CERES/{FM1,FM2}/granule_mmddyyyyhh" is abbreviated to "…/" in the table.

Path	Name	Dimension	Unit	Type	Description
…/Radiances	LW_Radiance	Varies by scene	$Wm^{-2} sr^{-1}$	Float32	[5]
…/Radiances	Radiance_Mode_Flags	Varies by scene	$Wm^{-2} str^{-1}$	Float32	[5]
…/Radiances	SW_Filtered_Radiance	Varies by scene	$Wm^{-2} sr^{-1}$	Float32	[5]
…/Radiances	SW_Radiance	Varies by scene	$Wm^{-2} sr^{-1}$	Float32	[5]
…/Radiances	TOT_Filtered_Radiance	Varies by scene	$Wm^{-2} sr^{-1}$	Float32	[5]
…/Radiances	WN_Filtered_Radiance	Varies by	$Wm^{-2} sr^{-1}$	Float32	[5]

		scene			
…/Radiances	WN_Radiance	Varies by scene	$Wm^{-2} sr^{-1}$	Float32	[5]
…/Time_and_Position	Latitude	Varies by scene	Degrees_north	Float32	[5]
…/Time_and_Position	Longitude	Varies by scene	Degrees_east	Float32	[5]
…/Time_and_Postion	Time_of_observation	Varies by scene	Day	Float64	[5]
…/Viewing_Angles	Relative_Azimuth	Varies by scene	Degree[1]	Float32	[5]
…/Viewing_Angles	Solar_Zenith	Varies by scene	Degree	Float32	[5]
…/Viewing_Angles	Viewing_Azimuth	Varies by scene	Degree[1]	Float32	[5]
…/Viewing_Angles	Viewing_Zenith	Varies by scene	Degree	Float32	[5]

4.2.3 MISR

All the data fields for MISR are stored under the root group name of "/MISR" in a BF granule. One BF granule contains one orbital MISR data for all of the MISR cameras. The designated MISR cameras name (DF, CF, BF, AF, AN, AA, BA, CA, DA) are used to name subgroups, where radiance fields are stored.

Table 4.4 HDF data variables for MISR. The root group name of "/MISR/" is abbreviated to "…/" in the table. In the table, {cam} following "…/" represents the subgroups named by one of the nine MISR cameras designated as (DF, CF, BF, AF, AN, AA, BA, CA, and DA).

Path	Name	Dimension	Unit	Type	Description
…/{cam}/BRF_Conversion_Factors	BlueConversionFactor	180×3×32	N/A	Float32	[1]
…/{cam}/BRF_Conversion_Factors	GreenConversionFactor	180×3×32	N/A	Float32	[1]
…/{cam}//BRF_Conversion_Factors	RedConversionFactor	180×3×32	N/A	Float32	[1]
…/{cam}//BRF_Conversion_Factors	NIRConversionFactor	180×3×32	N/A	Float32	[1]
…/{cam}//	BlockCenterTime	180	UTC[1]	Float32	[1]
…/{cam}//Data_Fields	Blue_Radiance	180×128×512 for off-nadir cameras 180×512×2048 for AN	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[1]
…/{cam}//Data_Fields	Green_Radiance	180×128×512 for off-nadir cameras 180×512×2048 for AN	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[1]
…/{cam}//Data_Fields	Red_Radiance	180×512×2048	$Wm^{-2}\mu m^{-1}sr^{-1}$	Float32	[1]
…/{cam}//Data_Fields	NIR_Radiance	180×128×512 for off-nadir cameras 180×512×2048 for AN	$Wm^{-2}\mu m^{-1st}sr^{-1}$	Float32	[1]
…/{cam}//Data_Fields	Blue_Radiance_low_accuracy_index	n×3; n is the number of pixels with reduced	N/A	Unsigned short	Only appear if pixels with RDQI=1

		accuracy, 3 records coordinates (block, sample, line) of these pixels			exist
…/{cam}//Data_Fields	Green_Radiance_low_accuracy_index	n×3; n is the number of pixels with low RDQI, 3 records coordinates (block, sample, line) of these pixels	N/A	Un-Int16	Only appear if pixels with RDQI ≥1 exist
…/{cam}//Data_Fields	Red_Radiance_low_accuracy_index	n×3; n is the number of pixels with low RDQI, 3 records coordinates (block, sample, line) of these pixels	N/A	Un-Int16	Only appear if pixels with RDQI ≥1 exist
…/{cam}//Data_Fields	NIR_Radiance_low_accuracy_index	n×3; n is the number of pixels with low RDQI, 3 records coordinates (block, sample, line) of these pixels	N/A	Un-Int16	Only appear if pixels with RDQI ≥1 exist
…/{cam}//Sensor_Geometry	{cam}Azimuth	180×3×32	Degree	double	[1]
…/{cam}//Sensor_Geometry	{cam}Glitter	180×3×32	Degree	double	[1]
…/{cam}//Sensor_Geometry	{cam}Scatter	180×3×32	Degree	double	[1]
…/{cam}//Sensor_Geometry	{cam}Zenith	180×3×32	Degree	double	[1]
	More fields				
…/Geolocation	GeoLatitude	180×128×512	Degrees_north	Float32	[1]
…/Geolocation	GeoLongitude	180×128×512	Degrees_east	Float32	[1]
…/HRGeolocation	GeoLatitude	180×512×2048	Degrees_north	Float32	[1]
…/HRGeolocation	GeoLatitude	180×512×2048	Degrees_north	Float32	[1]
…/Solar_Geometry	SolarAzimuth	180×3×32	Degree	double	[1]
…/Solar_Geometry	SolarZenith	180×3×32	Degree	double	[1]

4.2.4 MODIS

All the data fields for MODIS are stored under the root group name of "/MODIS" in a BF granule. One BF granule contains 18-20 the original MODIS 5-minute granules, each of which is stored as a separate and individual HDF5 subgroup, whose name is partially copied from the associated original file name, includes the starting time of the granule in the original time format. For example, the subgroup name of granule_2009270_0610 contains the data fields for the original MODIS granule having a starting time of 06:10 (UTC) on the 270th day of year 2000.

Table 4.5 HDF data variables for MODIS under the group of /MODIS/granule_yyyyddd_hhmm, where yyyyddd stands for year and julian date (ddd), and hhmm gives hour and minute(xx) of the starting time of data acquisition. The group path of /MODIS/granule_yyyyddd_hhmm" is abbreviated to "…/" in the table.

Path	Name	Dimension	Unit	Type	Description
…/_1KM/Data_Fields	EV_1KM_Emissive	16×[1950-2100]×1354	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[2]
…/_1KM/Data_Fields	EV_1KM_Emissive_Uncert_Indexes	16×[1950-2100]×1354	N/A	Float32	[2]
…/_1KM/Data_Fields	EV_1KM_RefSB	16×[1950-2100]×1354	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[2]
…/_1KM/Data_Fields	EV_1KM_RefSB_Uncert_Indexes	16×[1950-2100]×1354	N/A	Float32	[2]
…/_1KM/Data_Fields	EV_250_Aggr1km_RefSB	16×[1950-2100]×1354	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[2]
…/_1KM/Data_Fields	EV_250_Aggr1km_Uncert_Indexes	16×[1950-2100]×1354	N/A	Float32	[2]
…/_1KM/Data_Fields	EV_500_Aggr1km_RefSB	16×[1950-2100]×1354	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[2]
…/_1KM/Data_Fields	EV_500_Aggr1km_Uncert_Indexes	16×[1950-2100]×1354	N/A	Float32	[2]
…/_1KM/Geolocation	Latitude	16×[1950-2100]×1354	Degrees_north	Float32	[2]
…/_1KM/Geolocation	Longitude	16×[1950-2100]×1354	Degrees_east	Float32	[2]
…/_250m/Data_Fields	EV_250 _RefSB	2×[7800-8400]×5416	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[2]
…/_250m/Data_Fields	EV_250_RefSB_ Uncert_Indexes	2×[7800-8400]×5416	N/A	Float32	[2]
…/_250m/Geolocation	Latitude	2×[7800-8400]×5416	Degrees_north	Float32	[2]
…/_250m/Geolocation	Longitude	2×[7800-8400]×5416	Degrees_east	Float32	[2]
…/_500m/Data_Fields	EV_500 _RefSB	5×[3900-4200]×2708	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[2]
…/_500m/Data_Fields	EV_500_RefSB_ Uncert_Indexes	5×[3900-4200]×2708	N/A	Float32	[2]
…/_500m/Data_Fields	EV_250_Aggr500_RefSB	5×[3900-4200]×2708	$Wm^{-2}\mu m^{-1}str^{-1}$	Float32	[2]
…/_500m/Data_Fields	EV_250_Aggr500_Uncert_Indexes	5×[3900-4200]×2708	N/A	Float32	[2]
…/_500m/Geolocation	Latitude	5×[3900-4200]×2708	Degrees_north	Float32	[2]
…/_500m/Geolocation	Longitude	5×[3900-4200]×2708	Degrees_east	Float32	[2]
…/	SensorAzimuth	[1950-2100]×1354	Degree	Float32	[2]
…/	SensorZenith	[1950-2100]×1354	Degree	Float32	[2]
…/	SolarAzimuth	[1950-2100]×1354	Degree	Float32	[2]
…/	SolarZenith	[1950-2100]×1354	Degree	Float32	[2]

4.2.5 MOPITT

All the data fields for MOPITT are stored under the root group name of "/MOPITT" in a BF granule. One BF granule contains 1-2 the original MOPITT daily granules, each of which is stored as a separate and individual HDF5 subgroup, whose name is partially copied from the associated original file name, includes the day of the granule in the original time format. For example, the subgroup name of granule_20130213 contains the data fields for the original MOPITT granule on February 13 in year 2000. The entire data fields in the original MOPITT L1B products are completely copied and repacked in the BF product, given that the total size of these data fields is small and some data fields other than the radiance and geolocation fields may be useful for MOPITT users.

Table 4.6 HDF data variables for MOPITT under the group of /MOPITT/granule_yyyyddd, where yyyyddd stands for year and calendar date of data acquisition. The group path of /MOPITT/granule_yyyyddd" is abbreviated to "…/" in the table.

Path	Name	Dimension	Unit	Type	Description
…/Data_Fields	CalibrationData	n×4×8×2×8, n is the number of cross-tracks	N/A	Float32	[4]
…/Data_Fields	DailyGainDev	4×8×2	N/A	Float32	[4]
…/Data_Fields	DailyMeanNoise	4×8×2	N/A $^{-1}$	Float32	[4]
…/Data_Fields	DailyMeanPositionNoise	4×8×2×5	N/A	Float32	[4]
…/Data_Fields	EngineeringData	n×34×2	N/A	Float32	[4]
…/Data_Fields	Level0StdDev	n×29×4×8×2	N/A	Float32	[4]
…/Data_Fields	MOPITTRadiances	n×29×4×8×2	$Wm^{-2}sr^{-1}$	Float32	[4]
…/Data_Fields	PacketPositions	n×29	N/A	Float32	[4]
…/Data_Fields	SatelliteAzimuth	n×29×4	Degree	Float32	[4]
…/Data_Fields	SatelliteZenith	n×29×4	Degree	Float32	[4]
…/Data_Fields	SectorCalibrationData	n×4×8×4×8	N/A	Float32	[4]
…/Data_Fields	SolarAzimuth	n×29×4	Degree	Float32	[4]
…/Data_Fields	SolarZenith	n×29×4	Degree	Float32	[4]
…/Data_Fields	SwathQuality	n	N/A?	Float32	[4]
…/Geolocation	Latitude	n×29×4	Degrees_north	Float32	[4]
…/Geolocation	Longitude	n×29×4	Degrees_east	Float32	[4]
…/Geolocation	Time	n×29×4	Tai93	Float64	[4]

4.3 Metadata for Data Interoperability

4.3.1 CF Dimension Names

The Climate and Forecast (CF) convention requires that each dimension of a data array stored in a BF file must have a dimension name and the dimension name must be unique inside a file. Therefore, one dimension name can only be paired with one dimension size in one BF file.

Most of the dimension names provided in the original input granules for each Terra instrument are reserved in the BF granule metadata. For the interpolated latitude and longitude fields for MODIS and ASTER, we use the dimension names of the corresponding radiance fields. Since one BF file may have multiple HDF4 ASTER, MODIS, CERES and MOPITT granules, we have to change some dimension names to ensure that a dimension name is unique in one BF file. Although this complicates the dimension handling, we still adopt this approach primarily for the netCDF-4 users. The HDF5 users can simply ignore those attributes related to dimensions.

The following subsections provide detailed dimension information for each instrument.

4.3.1.1 ASTER

Table 4.7 Dimension names and sizes for ASTER where gsuffix represents each ASTER input granule. Suffix is in mmddyyyyhhxxss format. mmddyyyyhhmmss, stands for month (mm), date(dd), year(yyyy), hour(hh), minute(xx), and second(ss) of the starting time of data acquisition. This is consistent with the description listed in Table 4.2.

Category	Dimension Name	Dimension Size
TIR	ImageLine_TIR_Swath_gsuffix	Varies
	ImagePixel_TIR_Swath_gsuffix	Varies
	GeoTrack_TIR_Swath	11
	GeoXTrack_TIR_Swath	11
VNIR	ImageLine_VNIR_Swath_gsuffix	Varies
	ImagePixel_VNIR_Swath_gsuffix	Varies
	GeoTrack_VNIR_Swath	11
	GeoXTrack_VNIR_Swath	11
SWIR	ImageLine_SWIR_Swath_gsuffix	Varies
	ImagePixel_SWIR_Swath_gsuffix	Varies
	GeoTrack_SWIR_Swath	11
	GeoXTrack_SWIR_Swath	11
Pointing Angle	ASTER_PointingAngleDim	1
Solar Geometry	ASTER_Solar_GeometryDim	1

4.3.1.2 MODIS

4.3.1.2.1 General Information

Table 4.8 Dimension names and sizes. Except the non-typical dimension of the number of scans(listed in Table 4.9), all other dimensions provided by the MODIS input granules. The suffix '?' in the dimension name may be any number between 2 to 8 or character between 'a' and 'h'. The detailed information on these suffixes can be found in Table 4.9.

Category	Dimension Name	Dimension Size
1KM resolution	_40_nscans_MODIS_SWATH_Type_L1B(_?)	1950-2100
	Max_EV_frames_MODIS_SWATH_Type_L1B	1354
	Band_1KM_Emissive_MODIS_SWATH_Type_L1B	16
	Band_1KM_RefSB_MODIS_SWATH_Type_L1B	15
500mresolution	_20_nscans_MODIS_SWATH_Type_L1B(_?)	3900-4200
	_2_Max_EV_frames_MODIS_SWATH_Type_L1B	2708

	Band_500M_MODIS_SWATH_Type_L1B	5
250m resolution	_40_nscans_MODIS_SWATH_Type_L1B(_?)	7800-8400
	_4_Max_EV_frames_MODIS_SWATH_Type_L1B	5416
	Band_250M_MODIS_SWATH_Type_L1B	2
Geo-location		
	nscans_10_MODIS_Swath_Type_GEO(_?)	1950-2100
	mframes_MODIS_Swath_Type_GEO	1354

4.3.1.2.2 Number of Scans

The typical numbers of along track scans are 203 and 204. However, for a small percentage of MODIS granules, the number of scans doesn't hold the typical numbers. Considering all cases, the range is between 195 to 210 leading to 1950 to 2100 measurements for the 1km resolution; 3900 to 4200 measurements for the 500m resolution and 7800 to 8400 measurements for the 250m resolution, respectively. Since one BF file may include many MODIS granules and each dimension name must be unique, we have to provide different dimension names for the non-typical dimensions although in the input granule, they all share the same dimension name. To make it simple and reduce the unnecessary complex dimensions; we decide to add simple suffix after the original dimension names.

Table 4.9 Dimension names and sizes for MODIS number of scan.

number of scan dimension	Dimension name	Dimension size
1kmtypical	_10_nscans_MODIS_SWATH_Type_L1B	2030
1km > typical		
	_10_nscans_MODIS_SWATH_Type_L1B_2	2040
	_10_nscans_MODIS_SWATH_Type_L1B_3	2050
	_10_nscans_MODIS_SWATH_Type_L1B_4	2060
	_10_nscans_MODIS_SWATH_Type_L1B_5	2070
	_10_nscans_MODIS_SWATH_Type_L1B_6	2080
	_10_nscans_MODIS_SWATH_Type_L1B_7	2090
	_10_nscans_MODIS_SWATH_Type_L1B_8	2100
1km< typical		
	_10_nscans_MODIS_SWATH_Type_L1B_a	2020
	_10_nscans_MODIS_SWATH_Type_L1B_b	2010
	_10_nscans_MODIS_SWATH_Type_L1B_c	2000
	_10_nscans_MODIS_SWATH_Type_L1B_d	1990
	_10_nscans_MODIS_SWATH_Type_L1B_e	1980
	_10_nscans_MODIS_SWATH_Type_L1B_f	1970
	_10_nscans_MODIS_SWATH_Type_L1B_g	1960
	_10_nscans_MODIS_SWATH_Type_L1B_h	1950

number of scan dimension	Dimension name	Dimension size
500m typical	_20_nscans_MODIS_SWATH_Type_L1B	4060
500m > typical		
	_20_nscans_MODIS_SWATH_Type_L1B_2	4080

	_20_nscans_MODIS_SWATH_Type_L1B_3	4100
	_20_nscans_MODIS_SWATH_Type_L1B_4	4120
	_20_nscans_MODIS_SWATH_Type_L1B_5	4140
	_20_nscans_MODIS_SWATH_Type_L1B_6	4160
	_20_nscans_MODIS_SWATH_Type_L1B_7	4180
	_20_nscans_MODIS_SWATH_Type_L1B_8	4200
500m< typical		
	_20_nscans_MODIS_SWATH_Type_L1B_a	4040
	_20_nscans_MODIS_SWATH_Type_L1B_b	4020
	_20_nscans_MODIS_SWATH_Type_L1B_c	4000
	_20_nscans_MODIS_SWATH_Type_L1B_d	3980
	_20_nscans_MODIS_SWATH_Type_L1B_e	3960
	_20_nscans_MODIS_SWATH_Type_L1B_f	3940
	_20_nscans_MODIS_SWATH_Type_L1B_g	3920
	_20_nscans_MODIS_SWATH_Type_L1B_h	3900

number of scan dimension	Dimension name	Dimension size
250m typical	_40_nscans_MODIS_SWATH_Type_L1B	8120
250m > typical		
	_40_nscans_MODIS_SWATH_Type_L1B_2	8160
	_40_nscans_MODIS_SWATH_Type_L1B_3	8200
	_40_nscans_MODIS_SWATH_Type_L1B_4	8240
	_40_nscans_MODIS_SWATH_Type_L1B_5	8280
	_40_nscans_MODIS_SWATH_Type_L1B_6	8320
	_40_nscans_MODIS_SWATH_Type_L1B_7	8360
	_40_nscans_MODIS_SWATH_Type_L1B_8	8400
250m< typical		
	_40_nscans_MODIS_SWATH_Type_L1B_a	8080
	_40_nscans_MODIS_SWATH_Type_L1B_b	8040
	_40_nscans_MODIS_SWATH_Type_L1B_c	8000
	_40_nscans_MODIS_SWATH_Type_L1B_d	7960
	_40_nscans_MODIS_SWATH_Type_L1B_e	7920
	_40_nscans_MODIS_SWATH_Type_L1B_f	7880
	_40_nscans_MODIS_SWATH_Type_L1B_g	7840
	_40_nscans_MODIS_SWATH_Type_L1B_h	7800

number of scan dimension	Dimension name	Dimension size
1kmgeolocation typical	nscans_10_MODIS_Swath_Type_GEO	2030
1km > typical		
	nscans_10_MODIS_Swath_Type_GEO_2	2040
	nscans_10_MODIS_Swath_Type_GEO_3	2050
	nscans_10_MODIS_Swath_Type_GEO_4	2060
	nscans_10_MODIS_Swath_Type_GEO_5	2070
	nscans_10_MODIS_Swath_Type_GEO_6	2080
	nscans_10_MODIS_Swath_Type_GEO_7	2090
	nscans_10_MODIS_Swath_Type_GEO_8	2100

1km < typical		
	nscans_10_MODIS_Swath_Type_GEO_a	2020
	nscans_10_MODIS_Swath_Type_GEO_b	2010
	nscans_10_MODIS_Swath_Type_GEO_c	2000
	nscans_10_MODIS_Swath_Type_GEO_d	1990
	nscans_10_MODIS_Swath_Type_GEO_e	1980
	nscans_10_MODIS_Swath_Type_GEO_f	1970
	nscans_10_MODIS_Swath_Type_GEO_g	1960
	nscans_10_MODIS_Swath_Type_GEO_h	1950

4.3.1.3 CERES

Table 4.10 Dimension name and size for CERES where gsuffix represents each CERESS input granule. Suffix is in yyyymmddhh format, where yyyymmddhh, stands for year(yyyy), ,month (mm), and hour(hh) of the starting time of data acquisition. This is consistent with the description in Table 4.3.

Category	Dimension Name	Dimension Size
FM1	Footprints_FM1_gsuffix	Varies
FM2	Footprints_FM2_gsuffix	Varies

4.3.1.4 MISR

Table 4.11 Dimension name and size provided in the MISR input granules. Note: we need to create dimension names of blue band, green band and nadir band for camera AN since the dimension sizes on this camera are different than those on other cameras. The prefix 'AN_" is added to the original dimension names for these bands for camera AN.

Category	Dimension Name	Dimension Size
Block Time	SOMBlock_Time	180
Block dimension for data	SOMBlockDim_RedBand	180
	SOMBlockDim_BlueBand	180
	SOMBlockDim_GreenBand	180
	SOMBlockDim_NIRBand	180
Block dimension for geolocation	SOMBlockDim_Standard	180
Block dimension for geolocation(high resolution)	SOMBlockDim	180
Block dimension for Geometry	SOMBlockDim_GeometricParameters	180
Block dimension for BRF conversion factors	SOMBlockDim_BRF_Conversion_Factors	180
Y dimension for red band	YDim_RedBand	2048
Y dimension for blue band	YDim_BlueBand	512
Y dimension for Green band	YDim_GreenBand	512
Y dimension for NIR band	YDim_NIRBand	512
Y dimension for geolocation	YDim_Standard	512
Y dimension for geolocation(high resolution)	YDimH	2048

Y dimension for Geometry	YDim_GeometricParameters	32
Y dimension for BRF conversion factors	YDim_BRF_Conversion_Factors	32
X dimension for red band	XDim_RedBand	512
X dimension for blue band	XDim_BlueBand	128
X dimension for Green band	XDim_GreenBand	128
X dimension for NIR band	XDim_NIRBand	128
X dimension for geolocation	XDim_Standard	128
X dimension for geolocation(high resolution)	XDimH	512
X dimension for Geometry	XDim_GeometricParameters	8
X dimension for BRF conversion factors	XDim_BRF_Conversion_Factors	8

Y dimension for blue band on the AN camera	AN_YDim_BlueBand	2048
Y dimension for green band on the AN camera	AN_YDim_GreenBand	2048
Y dimension for NIR band on the AN camera	AN_YDim_NIRBand	2048
X dimension for blue band on the AN camera	AN_XDim_BlueBand	512
X dimension for green band on the AN camera	AN_XDim_GreenBand	512
X dimension for NIR band on the AN camera	AN_XDim_NIRBand	512

Table 4.12 Dimension name and size for the variables that store MISR low accuracy(RDQI = 1) radiation spatial-index location. The first dimension is called "quality flag index dimension". It represents the number of reduced accuracy pixels. The dimension size varies from bands and cameras. The second dimension gives their indexed coordinates in the order of block, block-relative line and block-relative sample. The dimension size of the second dimension is always 3. For example, if the second dimension for a low accuracy pixel in the array contains the values of (57,9,316), the location of the pixel is block 57, line 9 and sample 316.

Category	Dimension Name	Dimension Size
Quality flag index dimension	MISR_AA_GR_LA_INX_DIM	Varies
	MISR_AA_RR_LA_INX_DIM	Varies
	MISR_AF_GR_LA_INX_DIM	Varies
	MISR_AF_RR_LA_INX_DIM	Varies
	MISR_AN_BR_LA_INX_DIM	Varies
	MISR_AN_GR_LA_INX_DIM	Varies
	MISR_AN_NR_LA_INX_DIM	Varies
	MISR_AN_RR_LA_INX_DIM	varies
	MISR_BA_GR_LA_INX_DIM	varies
	MISR_BA_NR_LA_INX_DIM	varies
	MISR_BA_RR_LA_INX_DIM	varies
	MISR_BF_NR_LA_INX_DIM	varies

	MISR_BF_RR_LA_INX_DIM	varies
	MISR_CA_NR_LA_INX_DIM	varies
	MISR_CA_RR_LA_INX_DIM	varies
	MISR_CF_NR_LA_INX_DIM	varies
	MISR_CF_RR_LA_INX_DIM	varies
	MISR_DA_NR_LA_INX_DIM	varies
	MISR_DF_BR_LA_INX_DIM	varies
	MISR_DF_GR_LA_INX_DIM	varies
	MISR_DF_RR_LA_INX_DIM	varies
Quality flag position dimension	MISR_LA_POS_DIM	3

4.3.1.5 MOPITT

Table 4.13 Dimension names and sizes provided by MOPITT input granules. Note: since there may be two MOPITT input granules in one orbit, we use ntrack_1 and ntrack_2 to distinguish these two granules.

Category	Dimension Name	Dimension Size
	ncalib	8
	Nchan	8
	Neng	2
	Nengpoints	34
	Npchan	2
	Npixels	4
	Nposition	5
	Nsector	4
	Nstare	29
	Nstate	2
The dimension of the number of track for the first granule	ntrack_1	varies
The dimension of the number of track for the second granule	ntrack_2	varies

4.3.2 Other CF-related Metadata

4.3.2.1 _FillValues and valid_min,valid_max

CF conventions strongly recommend having the attributes valid_min and valid_max or the equivalent valid_range for the data variables. Valid_min stores the smallest valid value of a variable and valid_max stores the largest valid value of a variable. For the BF product, we set the valid_min for all the radiance variables be zero. The valid_max for individual instrument can be found in Table 4.12.

Table 4.14 The largest valid value(valid_max) of a variable of radiance variables for each instrument

Radiance fields	valid_max
ASTER	569
CERES	The input granule has the equivalent valid_range attribute.
MISR	800
MODIS radiance	100
MOPITT	20

MODIS reflectance	900

Besides valid_min and valid_max, CF conventions also require _FillValue if filled values are used in the measurement. Table 4.13 lists the _FillValue information as well as other special values for each instrument.

Table 4.15 The radiance filled values for each instrument

Instrument	_FillValue	Description
ASTER	-999.0	The radiance values for pixels not containing valid data, as indicated in Section 2.4 of ASTER Level 1T Product User\'s Guide(Version 1.0), are set to -999.0, which is also used as a filled value. For saturated pixels, their radiance values are set to -998.0.
CERES	3.402823e+38f	Provided by the input granule, the BF just keeps them.
MISR	-999.0	The radiance value for a pixel is set to -999.0, if the value of its RDQI is 2 or 3 or if its original dn value is either 16378 or 16380.
MODIS	-999.0	The reserved dn values for uncalibrated data ranging between 65501 and 65535, as listed in Table 5.6.1 of MODIS Level 1B Product User\'s Guide(MOD_PR02 V6.1.12(TERRA)), are proportionally mapped to the floating point numbers between -964.0 and -999.0, when being converted to radiance.
MOPITT	-9999.0	According to the original MOPITT granule attribute, -8888.0 is used to represent the invalid data. -9999.0 is used as the FillValue.

4.3.2.2 Coordinates and Geo-location Units

We provide the CF *coordinates* attributes for the radiance fields of ASTER, MODIS and MISR, MOPITT according to the CF conventions and Dataset Interoperability Recommendations for Earth Science approved by NASA ESDIS Standards Office(ESO)(https://cdn.earthdata.nasa.gov/conduit/upload/5098/ESDS-RFC-028v1.1.pdf). We also make the units of latitude and longitude CF-compliant.

4.4 Other Metadata

The representation for the data acquisition time may vary for different instruments. The BF product provides an attribute called GranuleTime to describe how individual instrument represents the data acquisition time. Table 4.16 lists the description of the GranuleTime for each instrument.

Table 4.16 The granule time for each instrument

Instrument	Granule Time example	Description
ASTER	01112010002054	The GranuleTime attribute represents the time of data acquisition in UTC with the MMDDYYYYhhmmss format. D: day. M: month. Y: year. h: hour. m: minute s:second. For example, 01112010002054 represents January 11th, 2010, at the 0 hour, the 20th minute, the 54th second UTC.

CERES	2007070316	the value of the GranuleTime attribute is time of data acquisition in UTC with the YYYYMMDDhh format. Y: year. M: month. D: day. h: hour. For example, 2007070316 represents July 3rd, 2007 at the 16th hour UTC.
MISR	040110	The attribute GranuleTime is represented by orbit numbers. For example, the value of 040110 indicates the data was acquired for orbit 40110.
MODIS	2007184.1610	The integer portion of the GranuleTime attribute value represents the Julian Date of acquisition in YYYYDDD form. The fractional portion represents the UTC hours and minutes of the Julian Date. For example, 2007184.1610 indicates the data acquisition time is at the 16th hour and the 10th minute (UTC) on July 3rd, 2007.
MOPITT	20070703	The value of GranuleTime attribute is the calendar date of data acquisition with the YYYYMMDD format. Y: year. M: month. D: day. For example, 20070703 represents July 3rd, 2007.

Appendix A: Missing input granules

Not all of the Terra instruments have valid radiance data for the same time period due to various reasons including but not limited to instrument anomalies, spacecraft maneuvers, instrument calibration activities, and software failures. For some orbits, no radiance data for all of the five Terra instrument are available and hence BF granules are not created. Table ?.? lists all of the orbits between Orbit 1000 (Feb 25, 2000) and Orbit 85302(December 31, 2015) for which the BF granules were not created.

Some input granules staged on the DAACs' servers are found corrupted and unreadable and We reported them to the DAACs. These input granules are not incorporated into the BF prod

Appendix B: MODIS scan number arrangement explanation

The number of MODIS along-track scans in some of the original 5-mintue granules are smaller than 203 or larger than 204, which has not been documented in the MODIS officially published documentations. The explanation for this is given as follows based on the personal contact with James Kuyper at the NASA Goddard Space Flight Center.

Data packets collected by the MODIS instrument during that transmission occasionally suffer a bit flip which affects a random field. If the bit flip affects the image data, it won't match the checksum for that packet, and it will be filtered out. However, the checksum only covers the image data. The primary and secondary packet headers are not covered, and they contain a wide variety of important information. If a bit lip gives a field an invalid value, it will generally cause that packet to be skipped. However, it's very common for the field to still contain a valid but incorrect value after the bit-flip. For a 5-minute MODIS granule with as scana number larger than 204, the relevant fields are the packet time stamp and the scan count field. The packets get sorted by time stamp before being processed which means that a corrupt time stamp will cause a packet to be moved to a different location in the file. A bit-flip in a time stamp can cause a huge change if it hits a high-order bit, and such packets generally get dropped.

However, it can also cause a small change if it hits a low-order bit. Any packet with a time stamp that is in error by less than 2 hours has a good chance of being mistaken for a valid packet collected at a different time. Scans were identified by looking at the scan count field. It holds the same value for all packets that belong to same scan. It increases by 1 with each scan. It's only 3 bits long, so when the scan count reaches 7, the next scan has a scan count of 1. If a packet is in the wrong location in the file due to a corrupted time stamp, it therefore has only about 1 chance in 8 of having the same scan count value as it's neighboring packets. If a packet has a corrupted scan count, it will also generally not match it's neighboring packets. In either case, the earliest versions of our code would see. For example, a consecutive bunch of packets with a scan count of 5, and treat them as a single scan. Then a packet would have a scan count of 3, and the Level 1code would assume that a new scan had started. This would be followed by many additional packets that have a scan count of 5, which our code would assume belonged to yet a third scan. The net result was that a single scan would be split up into three scans, the first of which would contain a large fraction of the data from the real scan, the second of which would contain only a single packet, and the third of which would contain the rest of the data from that same real scan. The Level 1 code has since been modified to look for packets which have time stamp and scan count values which are inconsistent with those of their neighboring packets, and filters them out. However, it can't do so perfectly. For instance, if multiple consecutive corrupt data packets happen to have the same scan count value, it's harder to be sure that they're corrupt. Any corrupt packet that escapes our current filters has a chance of causing split scans, just like the simpler case described above. Therefore, MODIS L1A processing is designed to allow as many as 210 scans, which can happen it runs into sufficiently many split scans. If so, any remaining unprocessed packets are discarded.

The cases where the scan number is less than 203 can be caused for any of a number of reasons: data transmission can be interrupted, individual data packets can get lost, and corrupted packets were detected and filter out.

Appendix C: CDL output of a sample BF

Appendix E: Sample metadata(Collection-level and granule-level)