

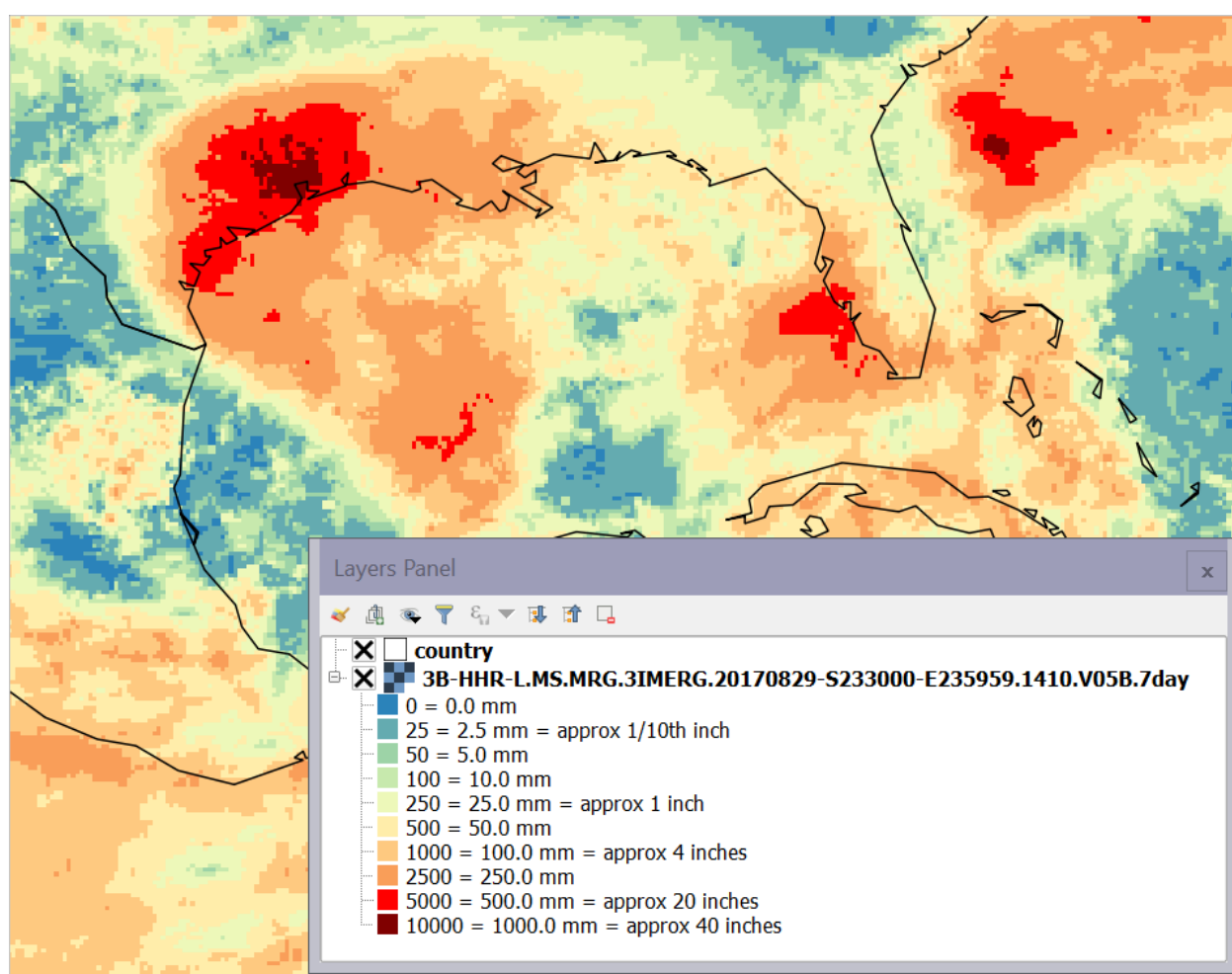
# The IMERG multi-satellite precipitation estimates reformatted as 2-byte GeoTIFF files for display in a Geographic Information System (GIS)

The Precipitation Processing System (PPS) at NASA Goddard

<https://pps.gsfc.nasa.gov/Documents/README.GIS.pdf>

Updated on 16 March 2022

by Owen Kelley (Owen.Kelley@nasa.gov)



**Caption:** Precipitation accumulation associated with Hurricane Harvey (2017) over Texas according to IMERG GIS. This near real-time data covers a 7-day period ending on 29 August 2017. The data has 0.1° geographic resolution and 0.1-millimeter precision. Image generated with the QGIS application.

# Contents

1. Introduction .....	3
1.1. The Early, Late, and Final versions of IMERG: latency and retention .....	5
1.2. IMERG HDF5 variables used to create IMERG GIS.....	6
1.3. Format specifications.....	7
2. How to Download IMERG GIS Files from PPS .....	8
2.1. Downloading IMERG GIS files .....	8
2.2. Deciding whether to download the WorldFile or just the GeoTIFF file .....	10
3. IMERG GIS Filenames and File Format.....	11
3.1. Scaling of precipitation rate (Final IMERG GIS only) .....	11
3.2. Scaling of precipitation accumulation .....	11
3.3. Precipitation-phase variables (liquid vs. solid precipitation) .....	14
3.4. Observation-count variables .....	18
3.5. Filenames for the GIS translation of Early and Late IMERG .....	19
3.6. Filenames for the GIS translation of Final IMERG .....	22
3.7. Missing data.....	24
3.8. Bias-correction related to missing data .....	24
3.9. Geographic boundaries .....	25
3.10. Format of geographic metadata in IMERG GIS GeoTIFFs and WorldFiles .....	25
4. How to Read the IMERG GIS Data Product.....	30
4.1. Read with ESRI ArcGIS.....	30
4.2. Read with QGIS.....	32
4.3. Read with Python.....	35
4.4. Read with Matlab.....	37
4.5. Read with IDL .....	38
6. References .....	39

# 1. Introduction

NASA has worked for decades to improve global high-resolution satellite-based estimates of surface precipitation. It is difficult to quantify how much precipitation falls around the Earth. This is especially true if one specifies a short time interval, high spatial resolution, or rapid dissemination (i.e., timeliness). The difficulty with obtaining such precipitation estimates arises because precipitation is often patchy and short lived and because observations from multiple kinds of instruments must be combined to produce the best global estimates.

NASA's most advanced algorithm for this purpose is IMERG, the Integrated Multi-satellite Retrieval for Global Precipitation Measurement (GPM), which was first released in 2014 (Huffman et al. 2019a, 2019b). The IMERG science algorithm writes its output in the Hierarchical Data Format Version 5 (HDF5) file format (<http://www.hdfgroup.org/HDF5/>).

The HDF5 format is widely used by the scientific community but is poorly supported by some Geographic Information Systems (GIS). Furthermore, GIS users often require periods of accumulation other than a half-hour or month, while the IMERG HDF5 files are produced only for these two periods.

For these reasons, NASA produces a GIS translation of IMERG for various accumulation periods. The GIS translation maintains the 0.1-degree spatial resolution of the original HDF5 data product. In April 2019, Version 6 of IMERG was released in both the HDF5 and the GIS formats, covering the over-20-year period from June 2000 to the present. This long-duration archive is a boon to scientific research and provides a training set for near real-time applications such as disaster monitoring. It is anticipated that Version 7 of IMERG will be released in 2022. With few exceptions, everyone should use the latest version of IMERG because it is likely to provide the most accurate precipitation estimates. All research using IMERG should state the version used.

Both the HDF5 and GIS versions of IMERG contain a latitude-longitude grid with 3600 columns for longitude and 1800 rows for latitude. This grid covers the globe (90°S to 90°N, 180°W to 180°E). IMERG Version 6 is the first version to morph microwave-based precipitation estimates that occur poleward of 60° latitude, rather than only report microwave-based precipitation estimates at the time of observation at high latitude. Still, IMERG Version 6 uses missing-data values over any snow-covered land surface and ice-covered ocean that lies north of 60°N or south of 60°S. Section 3.7 of the present document discusses missing-data values in IMERG GIS files.

The purpose of the present document is to describe the GIS translation of IMERG, a data product commonly called "IMERG GIS." IMERG GIS is a Tagged Image File Format (TIFF) image file that contains a two-dimensional array of two-byte integers and geographic metadata. Because of the metadata, an IMERG GIS file is also a Georeferenced Tagged Image File Format (GeoTIFF) file. Because some software has difficulty interpreting the geographic metadata stored in a GeoTIFF file, PPS also stores the geographic metadata in a separate file written in the Environmental Systems Research Institute (ESRI) WorldFile format.

An IMERG GIS file is somewhat different than a typical TIFF file that contains a photo. Many TIFF files contain an 8-bit or 24-bit digital image with pixels indexed to a 256-element color table or assigned true-color red-green-blue values. In contrast, an IMERG GIS file contains a grid of unsigned two-byte integers (16-bit) with values from 0 to 65535.<sup>1</sup> Some kinds of IMERG GIS files contain data that is scaled by 10 or 1000 and then rounded to the nearest integer. This scaling allows the TIFF file to preserve an additional digit or digits of accuracy.

Within a few hours of the satellite observations being collected, the Precipitation Processing System (PPS) at NASA Goddard creates near real-time IMERG, which is supplemented by the higher-accuracy research-quality IMERG that PPS creates several months after the satellite observations are collected. Near real-time IMERG is intended for applications such as disaster monitoring and flood and landslide risk assessment. Research IMERG is intended for meteorological and climatological research.<sup>2</sup> In the present document, the terms "real time" and "research quality" are used in an informal sense.

The official terms for the three latencies of IMERG are Early IMERG, Late IMERG, and Final IMERG, as described in Section 1.1 and Figure 1. Monthly HDF5 files are created only for Final IMERG, while 30-minute HDF5 files are created for all three latencies: Early, Late, and Final IMERG. Based on these HDF5 files, IMERG GIS files are created at additional durations for the convenience of users: 24-hour-duration for Final IMERG and 3-hour, 24-hour, 3-day, 7-day, and monthly duration for Early and Late IMERG.

The IMERG HDF5 files contain variables that describe the surface precipitation rate and the phase of the precipitation reaching the Earth's surface. In the HDF5 files, precipitation rate is expressed in millimeters per hour. In the 30-minute HDF5 files, the likelihood that the precipitation reaching the Earth's surface is in either liquid phase or mixed phase is expressed as a percentage (0% to 100%). See Section 3.3 of the present document for details about how precipitation phase is represented in IMERG GIS files.

The IMERG HDF5 files are generated from a constellation of polar-orbiting satellites that carry passive microwave sensors, from geosynchronous satellites that carry infrared sensors, and from calibration data derived from either satellite observations or rain gauges. From 1998 through 2019, PPS also generated a now-depreciated data product called the Tropical Rainfall Measurement Mission (TRMM) Multi-satellite Precipitation Analysis (TMPA). The design of

---

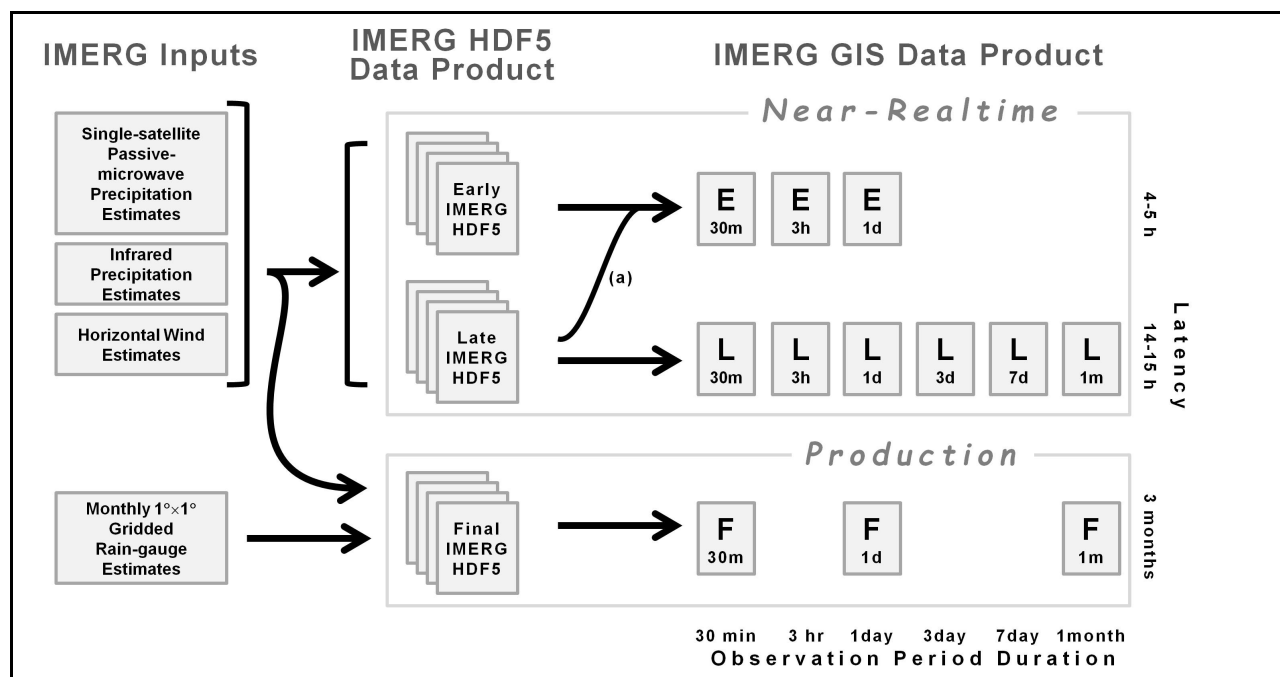
<sup>1</sup> In case a researcher accidentally reads the IMERG GIS data array into a signed rather than unsigned two-byte integer, IMERG GIS avoids using values greater than 32767, the maximum value of a signed two-byte integer. For example, IMERG GIS uses 29999 as the value for the missing data and scales precipitation accumulation (millimeters) by merely a factor of 10, so that even a seven-day accumulation is extremely unlikely to exceed 29999 after scaling millimeters by 10. Details about the IMERG GIS missing-data value and data-scaling factor are provided in Section 2 of the present document.

<sup>2</sup> IMERG may not be the best data product for climate research on precipitation over ocean and land if that research has low tolerance for artifacts due to satellite instruments entering and leaving over the 20+ year IMERG record. In such studies, it may be better to use the Global Precipitation Climatology Project (GPCP) 2.5-degree monthly precipitation estimates that cover 1979 to the present (<https://psl.noaa.gov/data/gridded/data.gpcp.html>).

the IMERG GIS data product is based on lessons learned when developing the GIS translation of TMPA.

Based on data downloads from PPS, IMERG is the most popular data product among more than a dozen data products associated with the GPM mission. Between April 2015 and 2019, about 10% of near real-time IMERG files that were downloaded were in the GIS format and the other 90% were in the HDF5 format (a total of over 100 million Early and Late IMERG file downloads). During the same period, approximately 1% of the approximately 100-million Final IMERG file downloads were in the GIS format, with the other 99% in the HDF5 format.

Questions related to the science of the IMERG algorithm should be sent to Dr. George Huffman ([george.j.huffman@nasa.gov](mailto:george.j.huffman@nasa.gov)) at NASA Goddard. Questions about the GIS reformatting of IMERG can be sent to the PPS Helpdesk ([helpdesk@mail.pps.eosdis.nasa.gov](mailto:helpdesk@mail.pps.eosdis.nasa.gov)). For more information about GPM, please visit the Global Precipitation Measurement (GPM) science-team website (<http://gpm.nasa.gov>).



**Figure 1.** Schematic diagram of the inputs that go into the IMERG HDF5 and IMERG GIS data products. Line (a) indicates that the Early IMERG GIS product uses the Early IMERG HDF5 product only for any 30-minute time period for which the higher-quality Late IMERG HDF5 product is not yet available.

### 1.1. The Early, Late, and Final versions of IMERG: latency and retention

Latency is the amount of time that passes between an observation being collected by the satellite and a data product being produced. There are three varieties of IMERG files, each with a different latency. These varieties are called Early, Late, and Final, and their associated latencies

are approximately 4–5 hours, 14–15 hours, and 3–4 months, respectively. In the Early IMERG data product, some regions that lack high-quality microwave estimates of precipitation are filled in with regions of microwave-estimated precipitation morphed in from previously observed high-quality observations. Where morphed-microwave information is not available, IMERG grid cells are filled with estimates that are based on geosynchronous infrared sensors. The longer latency of Late IMERG allows high-quality microwave-based precipitation estimates to be morphed both forward and backward in time to reach the time of interest. The much longer latency of Final IMERG allows the precipitation estimates to be calibrated by ancillary data sets (such as rain gauges) that become available months after the satellite and gauge observations were collected.

PPS retains the near real-time IMERG HDF5 and GIS products even after the higher-quality Final IMERG product is available. Near real-time IMERG HDF5 and GIS products (Early and Late) are retained until a new version of the Early and Late IMERG algorithm replaces them. This long retention period gives application developers a long-duration data set (over 20 years) for calibrating their real-time applications. Researchers, however, should use the Final IMERG product whenever possible instead of Early or Late IMERG because of the higher accuracy of Final IMERG.

PPS creates Early and Late IMERG GIS files going back to June 2000 whenever a new version of the IMERG algorithm is delivered, rather than just start Early and Late IMERG GIS processing with the most recent observations. This over 20-year-long data set is called a "retrospective" run of Early and Late IMERG. The retrospective run is a service to developers, to assist them with calibrating their near real-time applications.

## **1.2. IMERG HDF5 variables used to create IMERG GIS**

The IMERG GIS algorithm reads two variables from IMERG HDF5 files in order to create IMERG GIS files. In IMERG Version 6 (V6), these two HDF5 variables are called `/Grid/precipitationCal` and `/Grid/probabilityLiquidPrecipitation`. In 2022, it is anticipated that Version 7 (V7) of the IMERG algorithm will be generated. Version 7 will use the variable name `precipitation` instead of `precipitationCal`.

The `precipitationCal` (V6) and `precipitation` (V7) variables contain an estimate of the surface precipitation rate that is based on infrared satellite observations, microwave satellite observations, and any calibration data sets available at the time that the IMERG science algorithm is run. The `precipitationCal` (V6) and `precipitation` (V7) variables contain the best available estimate of surface precipitation possible at the latency of the file that contains them.

For a particular half-hour period, the `probabilityLiquidPrecipitation` variable is the probability that all of the precipitation in a single grid box reached the Earth's surface in either liquid phase or mixed phase. In other words, the stated probability is the complement of the probability that all of the precipitation in that grid box reached the surface in a completely frozen state. Precipitation phase is discussed in more detail in Section 3.3 of the present document.

### 1.3. Format specifications

The primary audience of the present document are researchers and applications developers. They will interact with IMERG GIS files using a graphical interface such as a GIS or an application programming interface (API) in the programming language of their choice. For this reason, the physical format of a GeoTIFF file on disk is usually irrelevant, but the logical structure of the contents of a GeoTIFF file may be relevant when designing or debugging software that reads GeoTIFF files.

The present section identifies the file-format specifications which describe both the physical and logical formats of GeoTIFF files and other relevant formats.

TIFF is an image format and TIFF stands for "Tag Image File Format" (Aldus 1992). Currently, Adobe owns the copyright to the 1992 document that defines the TIFF format but developers are not charged if they wish to write a software that implements the format described in that document. As of 2022, Adobe's website lists TIFF as an "open source standard" (<https://www.adobe.io/open/standards/TIFF.html>).

GeoTIFF is a format for storing geographic metadata inside of a TIFF file. The GeoTIFF specification defines the names and meaning of several TIFF tags and many TIFF GeoKeys. GeoKeys are small objects, many of which can be stored in a single TIFF tag. The GeoTIFF specification was introduced as a proposal in 1995, which declared itself to be GeoTIFF version 1.0 (Ritter and Ruth 1995). Subsequently, many software packages added partial or complete support for the GeoTIFF format. Many years later, the first formal standard for GeoTIFF was adopted in 2019 by the Open Geospatial Consortium (OGC). The OGC standard is designated GeoTIFF version 1.1. GeoTIFF version 1.1 is essentially identical to version 1.0 as far as the limited metadata needed to represent IMERG GIS. NASA claims that GeoTIFF version 1.1 is backwards compatible with version 1.0 (<https://earthdata.nasa.gov/esdis/eso/standards-and-references/geotiff>).

GDAL is a C library for reading many image formats, and it includes within it the libgeotiff C library for reading GeoTIFF files, which in turn includes the libtiff C library for reading TIFF files. The GDAL API is defined at <https://gdal.org/>. The GDAL C library is one of the ways that some high-level programming languages, including some Python libraries, read GeoTIFF files. Numerous people have reported difficulties installing GDAL-based solutions in high-level languages (e.g., Python Rasterio). Some developers distribute GDAL-free Python libraries for reading GeoTIFF files.

An ESRI WorldFile is a text file that contains six numbers, each on a separate line. The original specification document is not known, but the format is known to predate the 1995 release of GeoTIFF version 1.0. Definitions of the WorldFile format are ubiquitous online including in ESRI (2022) and on Wikipedia. Because this format is so simple, it is possible to write from scratch code in C or a high-level programming language to read and interpret a WorldFile.

HDF5 stands for Hierarchical Data Format Version 5 (HDF5). HDF5 is both a file format and an API. The file format specification (i.e., the "Design Specifications") and API can be found here: <https://portal.hdfgroup.org/display/HDF5/HDF5>. HDF5 is an open standard, and the C-language reference software that reads and writes HDF5 is currently maintained by a non-profit corporation called The HDF Group.

Suggesting the importance of the GeoTIFF format, a high-level NASA document recommended in 2020 that NASA Earth-observing data be stored in one of two formats: NetCDF4 or GeoTIFF (DPDG 2020, pp. 11–12). NetCDF4 is not really a file format, though. NetCDF4 is an Application Programming Interface (API) that internally calls the HDF5 C library to reading and writing data to and from disk (DPDG 2020, pg. 11).

Two documentation files are available for the IMERG science algorithm. The GPM data processing center, called the Precipitation Processing System (PPS), hosts the Algorithm Theoretical Basis Document for each version of IMERG (Huffman et al. 2019a). The URL for Version 6 IMERG is [https://arthurhou.pps.eosdis.nasa.gov/Documents/IMERG\\_ATBD\\_V06.pdf](https://arthurhou.pps.eosdis.nasa.gov/Documents/IMERG_ATBD_V06.pdf). PPS also hosts supplemental information about the IMERG algorithm in the *IMERG Technical Documentation* file (Huffman et al. 2019b).

The present document is the official description of the IMERG GIS data product. The URL for the present document is stated on its title page.

## 2. How to Download IMERG GIS Files from PPS

### 2.1. Downloading IMERG GIS files

Before downloading data files from the Precipitation Processing System (PPS), register an email address with PPS by visiting <http://registration.pps.eosdis.nasa.gov/>. After registering an email address, one can download Final IMERG GIS files using the PPS data-ordering system called STORM: <https://storm.pps.eosdis.nasa.gov/>. Other options for downloading Early, Late, and Final IMERG GIS files are given below.

As of 2021, files can be downloaded from the PPS archive using either the FTPS or HTTPS protocols. Between 1997 and 2021, the primary means of downloading PPS files had been FTP, the File Transfer Protocol. NASA mandated that organizations within NASA phase out the use of FTP.

For many researchers, the easiest way to download IMERG GIS may be to use HTTPS in a web browser. To do so, one would go to one of these URLs in a web browser to download Early, Late, or Final IMERG GIS:



<a href="https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/gis/early/2015/05/">https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/gis/early/2015/05/</a>	Early
<a href="https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/gis/2015/05/">https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/gis/2015/05/</a>	Late
<a href="https://arthurhouhttps.pps.eosdis.nasa.gov/gpmdata/2015/01/01/gis/">https://arthurhouhttps.pps.eosdis.nasa.gov/gpmdata/2015/01/01/gis/</a>	Final

Replace 2015/05/ or 2015/01/01/ with the date of interest in the format of yyyy/mm/ or yyyy/mm/dd/ where yyyy, mm, and dd are the four-digit year, two-digit month (zero padded), and two-digit day of month (zero padded). In these three URLs, note that the protocol is `https://` rather than `ftp://`. In addition, the string `https` appears in another location: the server name has the string `https` at the end of it, as in `jsimpsonhttps` or `arthurhouhttps`. Last, the directory tree on `jsimpsonhttps` is truncated, i.e., `data/` has been removed.

When one of these URLs is first visited, the web browser will present a pop-up window for one to type in a PPS-registered email address as both the username and password. This HTTPS interface to the archive presents a clickable list of directories, and eventually a list of files that one clicks on to download.

To avoid having the pop-up window appear, some web browsers will allow the username and password to be encoded within the URL. Instead of visiting `http://server/path`, one would visit `https://username:password@server/path`. The trick is that the `@` symbol in the email address would need to be encoded as `%40` when that email address as both the username and password. For example, one may use the following URL if the PPS-registered email address were `bob.smith@gmail.com`:

```
https://bob.smith%40gmail.com:bob.smith%40gmail.com@arthurhouhttps.pps.eosdis.nasa.gov/gpmdata/2015/01/01/gis/
```

Another approach for downloading IMERG GIS files is to use the `curl` or `wget` commands on the Linux command line. This is the best option if one has written a shell script to bulk download many files or to perform an automated search through the directory tree for specific files to download. PPS's experience is that the `curl` command works on both Linux and MacOS systems, while `wget` only works under Linux. Use the same three URLs mentioned above for the three kinds of IMERG: Early, Late, and Final.

To use one of these URLs with `curl` or `wget`, the syntax may vary with the operating system, but a pattern to start with is the following. In these commands, `$url` represents the URL of the file to download and `$email` is set to a PPS-registered email address:

```
wget -O -q --user="$email" --password="$email" "$url"
curl -s -u "$email:$email" "$url"
```

In the `wget` command, `-O` means that the name of the file in the archive is the name given the downloaded copy of the file, `-q` means quiet mode. In the `curl` command, `-s` means quiet mode. Because the HTTPS server uses Apache, it may have extended downtimes during system maintenance days. In contrast, the option described below is independent of Apache, and therefore will have briefer service interruptions.

The last and perhaps most complicated download option to consider is to have a system administrator open the organization's firewall so that one can use the FTPS protocol to download IMERG GIS from PPS. The exact syntax for FTPS access to the PPS archive varies from one operating system to another. Here is an example that may serve as a starting point:

```
curl -O --ssl-reqd -u "$email:$email" "$url"  
ftp://jsimpsonftps.pps.eosdis.nasa.gov/data/imerg/gis/2016/06/3B-  
DAY-L.GIS.IMERG.20160602.V06B.zip
```

For further information about FTPS access to the PPS archive, see two documents available on the PPS web page <https://arthurhou.pps.eosdis.nasa.gov/GPMprelimdocs.html>. The first document describes FTPS access to near real-time products, and the second document describes FTPS access to research-quality products:

<https://arthurhou.pps.eosdis.nasa.gov/Documents/FTP-FTPS-Transition.pdf>

<https://storm.pps.eosdis.nasa.gov/storm/arthurhouFTP-FTPS-Transition.pdf>

If you experience difficulty downloading IMERG GIS, please contact the PPS Helpdesk at [helpdesk@mail.pps.eosdis.nasa.gov](mailto:helpdesk@mail.pps.eosdis.nasa.gov).

## 2.2. Deciding whether to download the WorldFile or just the GeoTIFF file

Each IMERG GIS variable is stored in a separate GeoTIFF file. A GeoTIFF file has a \*.tif file extension. When using a GIS application that can read the geographic metadata stored within the GeoTIFF file, then there is no need to download the WorldFile. A WorldFile is a small ASCII text file that contains geographic metadata (Wikipedia 2021).

If your software cannot read the geographic metadata in a GeoTIFF file, then download the \*.tfw WorldFile that corresponds to any TIFF file of interest. The name of the GeoTIFF file and its associated WorldFile are identical except for the extension of \*.tif vs. \*.tfw.

The precipitation GeoTIFF and WorldFile are stored separately in the PPS online data archive. If one wishes to examine other variables, such as the phase of the precipitation, then download the \*.zip compressed file for the time period of interest. An IMERG GIS \*.zip file contains the pairs of GeoTIFF and WorldFile files for that time period.

As a third alternative, one could hardcode the following information if one could not read the GeoTIFF metadata and did not wish to download the WorldFile. This information locates IMERG GIS data on a map: the IMERG GIS file contains a 3600-by-1800 element grid that covers the globe, 180°W to 180°E and 90°S to 90°N. Each grid box represents a rectangular area in latitude-longitude coordinates that has size 0.1° latitude by 0.1° longitude. The northwestern-most corner of the first grid box is 180°W, 90°N. The center of the northwestern-most grid box is 179.95°W, 89.95°N. Like many other satellite and global positioning system (GPS) data products, the IMERG spatial reference system (SRS) is the European Petroleum Survey Group

geodetic datum #4326 (EPSG 4326), which is another name for World Geodetic System 1984 (WGS84) (Wikipedia.org, "World Geodetic System" and "EPSG Geodetic Parameter Dataset").

To verify that your software is correctly reading and interpreting the geographic metadata of IMERG GIS, see Section 3.10 of the present document for details about the geographic metadata in the IMERG GIS GeoTIFF files and WorldFiles.

### **3. IMERG GIS Filenames and File Format**

#### **3.1. Scaling of precipitation rate (Final IMERG GIS only)**

IMERG GIS files store precipitation as one-byte or two-byte integers because some GIS applications historically have had difficulty reading TIFF files that contain floating-point values. When necessary to preserve accuracy, the precipitation estimates are scaled by 10 or 1000 before being rounded off and written as integers to the TIFF file, as described below and in Table 1.

Thirty-minute and 1-day Final IMERG GIS files contain precipitation rates that are scaled by 10 before they are written to the two-byte TIFF file. A 1-month Final IMERG GIS file contains monthly-average precipitation rates that were scaled by 1000 before being written to the two-byte TIFF file. The reason for the larger scaling factor for monthly averages is that the monthly-average precipitation rate is typically much smaller (near ~1 mm/h) than is the average precipitation rate of an individual storm (10 to 100 mm/h can occur).

For example, a rain rate of 1.0 mm/h in a 30-minute or 1-day Final IMERG HDF5 file would be stored as a two-byte integer value of 10 in the 30-minute or 1-day Final IMERG GIS TIFF file. A rain rate of 1.0 mm/h in a monthly Final IMERG HDF5 file would be stored as a two-byte integer value of 1000 in the monthly Final IMERG GIS TIFF file. In other words, the units of a 30-minute or 1-day Final IMERG GIS file are 0.1 millimeters per hour. The units of a monthly Final IMERG GIS file are 0.001 millimeters per hour.

For consistency with Early and Late IMERG GIS, PPS also stores precipitation accumulation in millimeters in Final IMERG GIS files, not just average precipitation rate.

#### **3.2. Scaling of precipitation accumulation**

Real-time IMERG (i.e., Early and Late IMERG) has precipitation accumulation stored in its GIS translation, not precipitation rate. For accumulation periods of 30 minutes to 7 days, the precipitation accumulation in millimeters is multiplied by 10 before being rounded off to make a two-byte integer that is written to the TIFF file. For example, an accumulation of 1 millimeter

would be written as 10 in the two-byte TIFF file of the Early or Late IMERG GIS data product. In other words, the units of an Early or Late IMERG GIS file are 0.1 millimeters. For full-month Late IMERG accumulations, the data values are stored unscaled, i.e., a 10 mm accumulation would be stored as "10" in the full-month TIFF file.

Precipitation accumulation calculated for various time periods can be useful for assessing flood risk or landslide hazards in real time (Hong et al. 2007), and this is the reason for converting rate to accumulation in Early and Late IMERG GIS files. Every time that PPS generates a 30-minute real-time IMERG HDF5 file, PPS immediately creates a pair of 30-minute GIS files. Every 3 hours, PPS also generates an updated accumulation for the 3-hour to 7-day accumulation periods listed for Early and Late IMERG in Table 1. In this way, real-time users have up-to-date accumulations for various durations. At the end of a month, a full-month accumulation is created for Late IMERG GIS.

**Table 1.** The precipitation variable and units of various IMERG GIS files.

Official name	GIS file's precipitation variable and period	Precipitation scaling factor (physical units)	Integers that may be stored in the GIS file <sup>a</sup>	Corresponding precipitation values <sup>b</sup>
Early IMERG	Accumulation in 30 minutes	$\times 10$ (0.1 mm)	0–29998	0.0–2999.8 mm
	3 hours 1 day <sup>c</sup>			
Late IMERG	Accumulation in 30 minutes 3 hours 1 day	$\times 10$ (0.1 mm)	0–29998	0.0–2999.8 mm
	3 days 7 days			
	Accumulation in 1 month	$\times 1$ (1 mm)	0–29998	0.0–29998. mm
Final IMERG	Rate averaged over 30 minutes	$\times 10$ (0.1 mm/h)	0–29998	0.0–2999.8 mm/h
	1 day	$\times 10$ (0.1 mm/h)	0–29998	0.0–2999.8 mm/h
	1 month	$\times 1000$ (0.001 mm/h)	0–29998	0.0–29.998 mm/h
	Accumulation in 30 minutes	$\times 10$ (0.1 mm)	0–29998	0.0–2999.8 mm
	1 day		0–29998	0.0–2999.8 mm
	1 month	$\times 1$ (1 mm)	0–29998	0.0–29998. mm

<sup>a</sup> In the 2-byte unsigned TIFF files, a value of 29999 is used to indicate "missing" data.

<sup>b</sup> These very large ranges for precipitation rate and accumulation can be represented in the file, although the ranges that are likely to be encountered are smaller, as discussed in Section 4 of this document.

<sup>c</sup> The 1-day, 3-day, and 7-day references in this table refer to periods that are exactly 24, 3 $\times$ 24, or 7 $\times$ 24 hours long.

### 3.3. Precipitation-phase variables (liquid vs. solid precipitation)

The precipitation rates and accumulations discussed in this document are all liquid-equivalent rates and depths. For example, an IMERG file that reports 1 inch of frozen precipitation (i.e., 25.4 mm) means that whatever the depth that this precipitation had in its original frozen form, this precipitation would have a 1-inch depth were it melted first before being measured. According to Section 3.10 of the IMERG Algorithm Theoretical Basis Document (ATBD), the current IMERG algorithm groups together mixed-phase and liquid-phase precipitation. Since hail is usually mixed with rain, IMERG categorizes all of the precipitation from hail events as liquid/mixed phase.

As described below, the precipitation-phase variables in a Final IMERG HDF5 file should be interpreted differently depending on whether that HDF5 file covers a 30-minute period or a 1-month period (Section 3.10, IMERG ATBD). Similar considerations cause the precipitation-phase accumulated from multiple 30-minute Early or Late IMERG HDF5 files to be handled differently for accumulation periods that are longer than 1 day vs. accumulation periods that are up to 1 day long.

The IMERG GIS algorithm handles precipitation-phase information in a way that follows logically from the precise meaning of the precipitation-phase information that is stored in a single 30-minute IMERG HDF5 file. In a 30-minute IMERG HDF5 file, the precipitation-phase variable is the percentage likelihood ( $PERCENT = 0\%$  to  $100\%$ ), at each  $0.1^\circ$  grid box, that all of the precipitation reaching the Earth's surface during that 30-minute period is in liquid phase or mixed phase. The complement of that percentage ( $100\% - PERCENT$ ) is the percentage likelihood that all of the precipitation that reached the Earth's surface in that  $0.1^\circ$  grid box was frozen during that 30-minute period. It follows from this definition of precipitation phase that one should use a 50% threshold (or a similar percentage) to decide if the precipitation in short-duration accumulations (30 minutes to 1 day) is 100% liquid/mixed phase or 100% frozen. This method of determining precipitation phase is subsequently called the "50%-threshold method."

In a monthly IMERG HDF5 file, the precipitation-phase variable ( $PERCENT$ ) is computed as the percentage of total monthly accumulation that fell in liquid/mixed phase. If a monthly IMERG HDF5 file reported a precipitation rate  $p$  (liquid-equivalent depth), then one should assume that a rate of  $p \cdot PERCENT$  of precipitation reached the Earth's surface in liquid/mixed phase and that  $p \cdot (1 - PERCENT)$  of precipitation reached the Earth's surface frozen. This method of calculating precipitation phase is subsequently called the "product method" because it takes the product of the probability and the total precipitation rate that are both stored in the IMERG HDF5 file.

If one is interested in the phase of the precipitation, one should download the compressed file with the \*.zip extension. The \*.zip file contains GeoTIFF and WorldFiles (\*.tif and \*.tfw files) for total precipitation ( $TP$ ), liquid/mixed precipitation ( $LP$ ), liquid-equivalent ice-phase-only precipitation ( $IP$ ), and the percent of liquid/mixed-phase vs. total precipitation ( $PERCENT$ ).

In the following discussion, TP, LP, and IP should be taken to have units of millimeters if one is discussing real-time or research values for precipitation accumulation during the period of an IMERG GIS file. In contrast, if one is working with the research values for average precipitation rate during the period of the IMERG GIS file, then the values of TP, LP, and IP should be taken to be millimeters per hour.

To make the four IMERG GIS variables consistent, the following relationships always hold true. The total-precipitation file is always equal to the sum of the liquid/mixed-phase precipitation file and the liquid-equivalent ice-phase precipitation file. The liquid-precipitation percent file is always 100 times the ratio of the liquid/mixed-precipitation file and the total-precipitation file. These two relationships are expressed in Equation 1.

$$\begin{aligned}
 TP &= LP + IP \\
 PERCENT &= 100 \frac{LP}{TP}
 \end{aligned}
 \tag{1}$$

The developer of the IMERG science algorithm recommends that the product method generally should be used for "long" periods of time (longer than 1 day) while the previously-mentioned 50%-cutoff method should be used for "short" periods of time (up to and including 1-day long). The rest of this section describes the mathematical details that implement these ideas. In the following equations, the values read from the input HDF5 IMERG file are rate (mm/h) and percent (0% to 100%). The notations for these read-in input values are  $TP_0$  and  $PERCENT_0$  if one input HDF5 file is used and  $TP_i$  and  $PERCENT_i$  if more than one input HDF5 file is used. The output values written to the IMERG GIS files are total precipitation ( $TP$ ), liquid/mixed-phase precipitation ( $LP$ ), liquid-equivalent ice-phase precipitation ( $IP$ ), and percent-liquid precipitation ( $PERCENT$ ).

Equation 2 implements the 50%-threshold method. For 30-minute-duration Early and Late IMERG GIS accumulations, use Equation 2 with  $k = 0.5$  hour, a value for  $k$  that converts a precipitation rate (mm/h) into a half-hour accumulation (mm). For 30-minute-duration Final IMERG GIS averages,  $k = 1$  (unitless) in Equation 2 since both the input variable and output variables are precipitation rates (mm/h).

$$\begin{aligned}
TP &= k TP_0 \\
LP &= \begin{cases} TP & \text{if } PERCENT_0 \geq 50\% \\ 0 & \text{if } PERCENT_0 < 50\% \end{cases} \\
IP &= TP - LP \\
PERCENT &= \begin{cases} 255 & \text{if } TP_0 = 0 \\ 100 & \text{if } PERCENT_0 \geq 50\% \text{ and } TP_0 > 0 \\ 0 & \text{if } PERCENT_0 < 50\% \text{ and } TP_0 > 0 \end{cases}
\end{aligned} \tag{2}$$

An "undefined" value of 255 is used for *PERCENT* when there is zero precipitation.

Equation 3 is used for the case of several 30-minute IMERG HDF5 files as input, but few enough that the 50%-threshold method is still applicable. In Equation 3, terms with subscript *i* are the input grids read from the *i*<sup>th</sup> individual 30-minute IMERG HDF5 file from among the *n* 30-minute IMERG HDF5 files read. Terms without a subscript are the output grids written to a 3-hour or 1-day IMERG GIS file. In Equation 3, *k* = 0.5 hour for Early IMERG GIS and Late IMERG GIS. In contrast, *k* = 1/*n* (unitless) for Final IMERG. Final IMERG GIS files contain averages, and to calculate the average of *n* terms, one must divide the sum of the terms by *n*. For a 3-hour period, *n* = 6 and for a 1-day period, *n* = 48.



$$TP = k \sum_{i=1}^n TP_i \quad \text{for } n = 6 \text{ or } 48$$

$$b_i = \begin{cases} 1 & \text{if } PERCENT_i \geq 50\% \\ 0 & \text{if } PERCENT_i < 50\% \end{cases}$$

$$LP = k \sum_{i=1}^n b_i TP_i \quad \text{for } n = 6 \quad (3)$$

$$IP = TP - LP$$

$$PERCENT = \begin{cases} 255 & \text{if } TP = 0 \\ 100 \frac{LP}{TP} & \text{if } TP > 0 \end{cases}$$

Equation 4 is used when there are multiple 30-minute IMERG HDF5 files as input and the time period is long enough to use the product method rather than the 50%-threshold method. Equation 4 is only used for the 3-day, 7-day, and monthly accumulations calculated from Late IMERG HDF5 files. In Equation 4,  $n$  equals (48  $d$ ), where  $d$  is 3, 7, or 28–31, i.e., the number of days covered by the accumulation period. In Equation 4,  $k = 0.5$  hours because Late IMERG GIS contains accumulations, as described earlier in this section.

$$TP = k \sum_{i=1}^n TP_i$$

$$LP = k \sum_{i=1}^n \frac{PERCENT_i}{100} TP_i$$

$$IP = TP - LP \quad (4)$$

$$PERCENT = \begin{cases} 255 & \text{if } TP = 0 \\ 100 \frac{LP}{TP} & \text{if } TP > 0 \end{cases}$$

Equation 5 is for the case that there is a single, monthly IMERG HDF5 file as the input. A month is considered a long time-period for the purpose of this calculation, so the product method is used rather than the 50%-threshold method. Only Final IMERG (not Early or Late IMERG) has a monthly HDF file, so Eq. 5 is used to create only the monthly Final IMERG GIS file.

$$\begin{aligned}
 TP &= TP_0 \\
 LP &= \frac{PERCENT_0}{100} TP_0 \\
 IP &= TP - LP \\
 PERCENT &= 100 \frac{LP}{TP} = PERCENT_0
 \end{aligned} \tag{5}$$

### 3.4. Observation-count variables

In 2020, PPS added two diagnostic variables to IMERG GIS. The two variables are the number of half-hour periods with valid precipitation estimates in each of the 3600×1800 grid boxes, and the number of half-hour periods with greater than zero precipitation in each of the 3600×1800 grid boxes. The two variables are abbreviated "numValidHalfHour" and "numPrecipHalfHour" in the IMERG GIS filenames. The two variables are referred to as  $n_{valid}$  and  $n_{precip}$  in the equations in this section of the documentation.

When a grid box in an input IMERG HDF5 file contains a precipitation rate of exactly zero, that rate is considered equally "valid" as a greater-than-zero precipitation rate for the purpose of calculating IMERG GIS accumulations and average rates. In contrast, a value of -9999 in an input IMERG HDF5 file is considered a "not valid" precipitation rate.

In this discussion, another relevant quantity is the number of half-hour periods  $n_{max}$  that fit into the duration of the IMERG GIS file. The value of  $n_{max}$  depends only on the duration of the IMERG GIS file, and it does not vary with the presence or absence of missing-data values in the input IMERG HDF5 files. An IMERG GIS file with a duration of 30 minutes, 3 hours, 1 day, 3 days, or 7 days has a value of  $n_{max}$  of 1, 6, 48, 144, or 336, respectively. The value of  $n_{max}$  for a monthly IMERG GIS file varies from 1344 to 1488 depending on the number of days in that month because there are 28 to 31 days in a month and 48 half-hour periods in a day.

As described in subsequent paragraphs, there are some mathematical relations that exist among  $n_{precip}$ ,  $n_{valid}$ , and  $n_{max}$  and the values of other variables such as the accumulation or the average precipitation rate during the period covered by a grid box in an IMERG GIS file.

One of these relationships is that the number of valid half-hour periods  $n_{valid}$  for any grid box is between 0 and the number of half-hours  $n_{max}$  in the duration of the IMERG GIS file:

$$0 \leq n_{valid} \leq n_{max}$$

The number of half-hour periods  $n_{precip}$  with greater than zero precipitation in a grid box is between zero and the number of valid half-hour periods  $n_{valid}$  for that grid box:

$$0 \leq n_{precip} \leq n_{valid}$$

If the number of valid half-hour periods  $n_{valid}$  is less than 90% of  $n_{max}$ , then the precipitation accumulation is set to the IMERG GIS missing-data value (i.e., MISSING), while the values of both  $n_{valid}$  and  $n_{precip}$  may, nonetheless, be greater than zero. In this situation, the precipitation accumulation (ACCUMULATION) and average precipitation rate (RATE) are both set to the missing-data value because the values that could be calculated from the available half-hour periods are considered unreliable:

if  $n_{valid} < 0.90 n_{max}$   
then ACCUMULATION=MISSING and RATE=MISSING and  $n_{precip} \geq 0$

If the vast majority (90% or more) of half-hour periods have valid precipitation rates (zero or greater rates), then the accumulation and average rate calculated from the input IMERG HDF5 files is written to the IMERG GIS file, i.e., the missing-data value is not substituted:

if  $n_{valid} \geq 0.90 n_{max}$   
then ACCUMULATION  $\geq 0$  and RATE  $\geq 0$  and  $n_{precip} \geq 0$

The last interesting case is that, in grid boxes with very little accumulation, the accumulation written to the IMERG GIS file may be zero while  $n_{precip} > 0$ . This last case arises because floating-point accumulations and rates are rounded off before being written to an IMERG GIS file. More specifically, the accumulation in an IMERG GIS file is accurate to the nearest 0.1 mm or 1 mm, depending on whether the file's duration is under 1 month or equal to 1 month, respectively. For durations of under 1 month, an accumulation greater than zero but less than 0.05 mm for the entire period of the IMERG GIS file shall be round off to exactly zero before it is written to the IMERG GIS file. The value of  $n_{precip}$  in this same grid box is unaffected by this round-off error and may be greater than zero.

### 3.5. Filenames for the GIS translation of Early and Late IMERG

The filename convention is different for IMERG GIS files created in real-time (i.e., Early and Late IMERG) than for IMERG GIS files created months later (i.e., Final IMERG).

Based on the way that real-time IMERG GIS is used, it makes sense that the end datetime of the accumulation period is part of the filename of Early or Late IMERG GIS files. Real-time users often want precipitation estimates that are as close as possible to the present. Real-time users can easily estimate how "fresh" an IMERG GIS file is by looking at the datetime stamp in the GIS filename. For Early and Late IMERG GIS accumulation from 30 minutes to 7 days long,

the filenames state the start datetime and end time of the last 30-minute period included in that accumulation.

As soon as a new hour of Early or Late IMERG HDF5 file is created on the PPS Real-time System, PPS creates two 30-minute Early IMERG GIS accumulations and two 30-minute Late IMERG GIS accumulations. Every 3 hours, PPS also creates a set of Early and Late IMERG GIS accumulations for periods of 3 hours to 7 days. When the last Late IMERG HDF5 file of a month is created, PPS creates a Late IMERG GIS month accumulation. Table 2 describes the filename patterns for IMERG GIS files. Figure 2 is a schematic diagram of when the IMERG GIS files are created for various time periods near the end of a month.

**Table 2.** The filename patterns for IMERG GIS files.

Time-period boundary	Duration	Early and Late IMERG (i.e., "real-time" IMERG)	Final IMERG (i.e., "research" IMERG)
A 30-minute time period starting at 0000, 0030, 0100, ..., 2300, 2330	30 minutes	3B-HHR-E.* <b>A</b> - <i>Shhmm</i> *. <b>V</b> .30min. <b>X</b> <sup>a</sup> 3B-HHR-L.* <b>A</b> - <i>Shhmm</i> *. <b>V</b> .30min. <b>X</b> <sup>a</sup>	3B-HHR-GIS.* <b>A</b> - <i>Shhmm</i> *. <b>V</b> . <b>X</b> <sup>a</sup>
Any period that begins at 0, 3, 6, 9, 12, 18, or 21 UTC	3 hours		
	1 day	3B-HHR-E.* <b>B</b> - <i>Shh30</i> *. <b>V</b> .3hr. <b>X</b> 3B-HHR-E.* <b>B</b> - <i>Shh30</i> *. <b>V</b> .1day. <b>X</b>	
	3 hours	3B-HHR-L.* <b>B</b> - <i>Shh30</i> *. <b>V</b> .3hr. <b>X</b>	[not applicable]
	1 day	3B-HHR-L.* <b>B</b> - <i>Shh30</i> *. <b>V</b> .1day. <b>X</b>	
	3 days	3B-HHR-L.* <b>B</b> - <i>Shh30</i> *. <b>V</b> .3day. <b>X</b>	
	7 days	3B-HHR-L.* <b>B</b> - <i>Shh30</i> *. <b>V</b> .7day. <b>X</b>	
Day boundary, 0000 UTC	1 day	3B-DAY-L.* <b>A</b> -S0000*. <b>V</b> . <b>X</b> <sup>b</sup>	3B-DAY-GIS.* <b>A</b> -S0000*. <b>V</b> . <b>X</b>
Month boundary, 0000 UTC	1 month	3B-MO-L.* <b>A</b> -S0000*. <b>V</b> . <b>X</b>	3B-MO-GIS.* <b>A</b> -S0000*. <b>V</b> . <b>X</b>

<sup>a</sup> For 30-minute-long periods, A=B, i.e., the start date of the last 30-minute portion of the period is the same as the start date of the full period. The symbol *hh* represents 02, 05, 08, 11, 14, 17, 20, or 23 hours. The symbol *mm* represents 00 or 30 minutes.

<sup>b</sup> For Late IMERG GIS, at 0 UTC each day, there are two copies of the 1-day accumulation. One copy follows the PPS standard naming convention that uses the start date/time of the period of coverage. The other copy uses the GIS real-time convention, specifically that the filename gives the last 30-minute granule within the period of coverage.

In Table 2, **E** refers to Early IMERG and **L** refers to Late IMERG. **A** stands for the start date of the period covered by the GIS file in the yyyyymmdd format. **B** stands for the start date of the last day in the multi-day period covered by the GIS file, in yyyyymmdd format. **V** stands for version number of the IMERG HDF5 data product such as "V06A."

In Table 2, **X** stands for the file suffix. The suffix is composed of one or more of the following: a duration, a variable, and a file extension. For durations of 30 minutes to 7 days, the duration of the string in the filename may be "30min," "3hr," "1d," "3d," and "7d." For GeoTIFF files, the variable and file extension is one of the following strings with the few exceptions discussed at the end of this section of the documentation:

tif	= Total precipitation accumulation
ice.tif	= Liquid-equivalent ice-phase precipitation accumulation
liquid.tif	= Liquid-phase and mixed-phase precipitation accumulation
liquidPercent.tif	= Percent chance that the precipitation is liquid or mixed phase in 30-minute to 1-day periods. For periods of 3 days or longer, the percent of precipitation that is liquid or mixed phase.
numValidHalfHour.tif	= During the 30-minute to 1-month duration of the IMERG GIS file, the number of 30-minute periods that have a valid precipitation estimate in a particular 0.1°×0.1° grid box.
numPrecipHalfHour.tif	= During the 30-minute to 1-month duration of the IMERG GIS file, the number of 30-minute periods that have greater than zero precipitation in a particular 0.1°×0.1° grid box.

A WorldFile has the same filename suffix as the associated GeoTIFF file except that it ends in \*.tfw. A zipfile likewise ends in \*.zip. A zip file can be uncompressed using the unzip Linux command or by double-clicking on the file in a Microsoft Windows system. The zip format is used to store groups of GIS files rather than the tar or gzip formats because all Microsoft Windows systems have native applications for uncompressing zip files while some Microsoft Windows systems do not have tar or gzip applications.

For Early and Late IMERG GIS only (not Final IMERG GIS), a text file (ending in \*.txt) is present when some of the expected 30-minute-duration HDF5 files were missing. In this situation, the text file states how many 30-minute-duration HDF5 files were used to generate the GeoTIFF file.

There are two additional kinds of Late IMERG GIS files: day and month files that start on day or month boundaries. First, the Late IMERG 24-hour accumulation that covers 0000 UTC to 2359 UTC is placed in a zip file with a name that follows this pattern: 3B-DAY-L.\*yyyyymmdd.Vxxx.zip, where xxx is the version number such as "06A." Second, the Late IMERG accumulation for a full month is placed in a zip file with a name that follows this pattern: 3B-MO-L.\*yyyymm01.Vxxx.zip.

### 3.6. Filenames for the GIS translation of Final IMERG

PPS creates overlapping accumulations for real-time IMERG GIS because real-time applications need up-to-date information, e.g., 24-hour accumulations updated more than once each day. However, for research IMERG, there has not been a demand for overlapping GIS files. For this reason, there is only one daily research IMERG GIS file containing total precipitation rate, and its averaging period starts at 0000 UTC on each day.

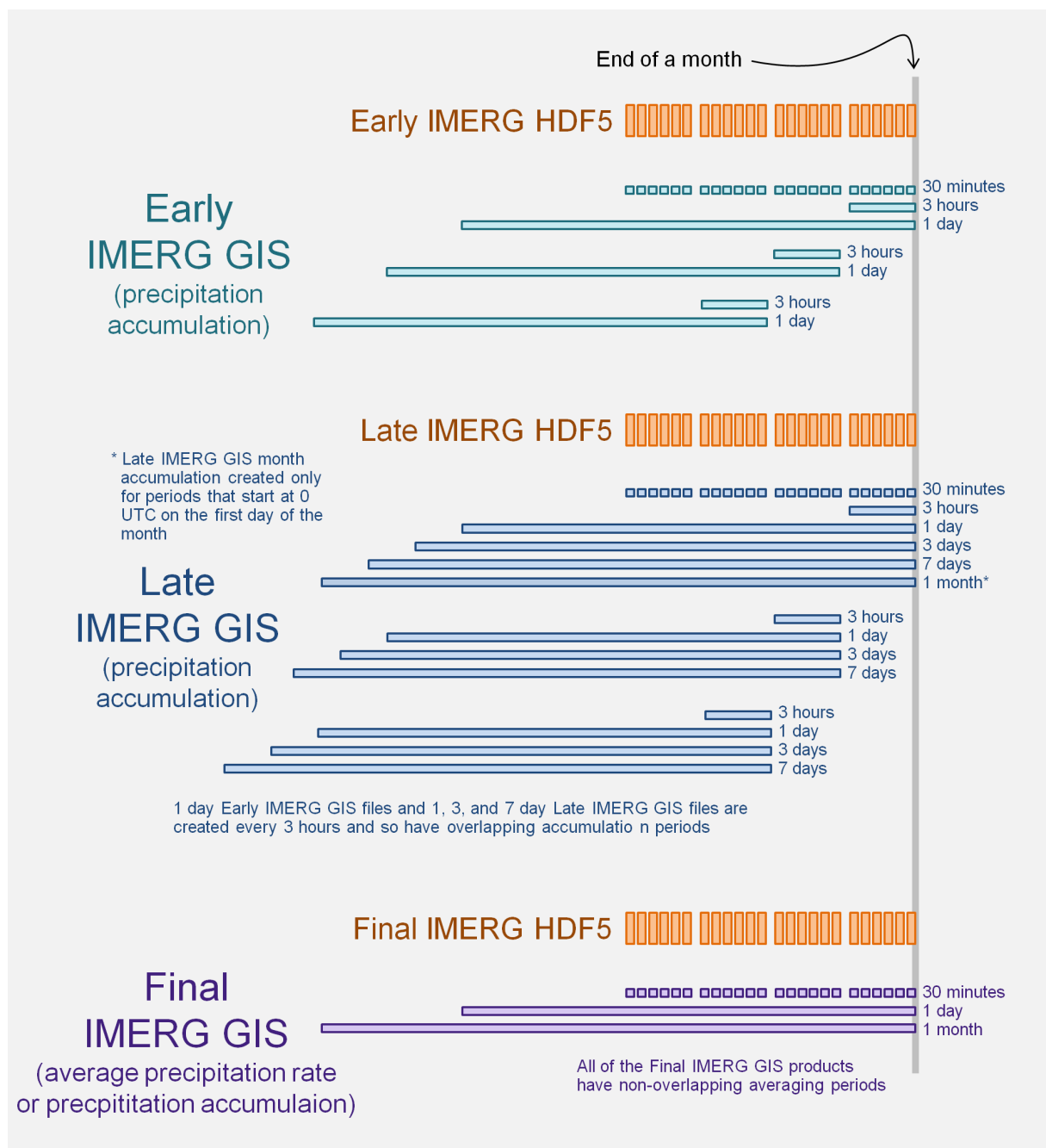
For research IMERG GIS files, the three different averaging periods (30 minutes, 1 day, and 1 month) are indicated by the filename prefix rather than by adding a duration suffix to the filename. See Figure 2 for a schematic representation of the time periods covered by Final IMERG GIS files. Below is the pattern for the filenames for the three averaging periods that begin on 0000 UTC on 1 June 2014:

<u>Filename root for GIS file</u>	<u>Duration of GIS file</u>
3B-HHR-GIS.*.20140601-S000000-E002959	30 minutes
3B-DAY-GIS.*.20140601-S000000-E235959	1 day
3B-MO-GIS.*.20140601-S000000-E235959	1 month

To view the total precipitation rate, download the \*.tif and \*.tfw files from the online archive. To view precipitation-phase information, download the \*.zip file. The \*.zip file contains nine GeoTIFF files with the following filename suffixes:

- total.accum.tif = accumulation during the stated duration of the file
- total.rate.tif = precipitation rate averaged over the stated duration of the file
- ice.accum.tif = solid-phase accumulation
- ice.rate.tif = solid-phase precipitation rate
- liquid.accum.tif = liquid and mixed-phase accumulation
- liquid.rate.tif = liquid and mixed-phase rate
- liquidPercent.tif = percent of total precipitation that falls as liquid or mixed phase
- numPrecipHalfHour.tif = During the 30-minute to 1-month duration of the IMERG GIS file, the number of 30-minute periods that have a valid precipitation estimate in a particular 0.1°×0.1° grid box.
- numValidHalfHour.tif = During the 30-minute to 1-month duration of the IMERG GIS file, the number of 30-minute periods that have greater than zero precipitation in a particular 0.1°×0.1° grid box.

Among these files, the first file contains the total precipitation rate, the \*.liquid.tif file contains the precipitation that reached the surface in liquid or mixed phase, the \*.ice.tif file contains precipitation that reached the surface in ice phase, and the \*.liquidPercent.tif extension indicates the fraction of the total precipitation that reached the surface in liquid or mixed phase.



**Figure 2.** A schematic diagram of some of the IMERG GIS files that are created near the end of a month. The IMERG GIS files are shown in green, blue, and purple for Early, Late, and Final IMERG, respectively. The 30-minute input files (HDF5 files) are shown in orange. The horizontal axis represents time, with the vertical gray line indicating the end of the month, and observations earlier than that to the left of the vertical gray line.

### 3.7. Missing data

A value of either 255 or 29999 is used to indicate missing data in IMERG GIS files depending on whether the data-type is 1 byte or 2 bytes. The two-byte IMERG GIS files that contain precipitation rate or precipitation accumulation use 29999 for the missing-data value in grid boxes for which the IMERG HDF5 files do not contain valid rainfall rates. The one-byte IMERG GIS files that contain the percentage of liquid precipitation use 255 for the missing value in grid boxes for which there are zero values or missing values for the total precipitation. The ratio of the liquid and mixed-phase precipitation to the total precipitation is undefined in this case (division by zero), so a missing data value is supplied in the percentage field of the IMERG GIS file.

The details of how IMERG GIS handles missing-data are based on the idea that IMERG GIS merely summarizes the precipitation estimates in either 30-minute-duration or monthly-duration IMERG HDF5 science files. The IMERG GIS algorithm itself is not a scientific algorithm, it is just a simple "accumulator" calculating sums and averages. The IMERG GIS algorithm avoids, to the extent possible, making scientific assumptions or using ancillary data.

There is only one option for handling missing-data when there is a single input file: any  $0.1^\circ \times 0.1^\circ$  grid box with missing data-values in the input file is set to missing in the output IMERG GIS file. This is the case with 30-minute-duration and 1-month-duration IMERG GIS files.

A different calculation method is used to generate IMERG GIS files that have multiple input 30-minute-duration HDF5 files. A  $0.1^\circ \times 0.1^\circ$  grid box in the output IMERG GIS file is given a missing-data value only if that grid box has missing data values in *all* of the input HDF5 files. If one or more of the input HDF5 files have a valid precipitation rate ( $\geq 0 \text{ mm h}^{-1}$ ) in a particular  $0.1^\circ \times 0.1^\circ$  grid box, then those values are multiplied by 0.5 h and summed to create the accumulation (mm) written to that grid box in the output IMERG GIS file.

Obviously, this calculation method is likely to give lower-than-expected accumulation in grid boxes where some of the 30-minute input periods have missing-data values. A more unstable but less-biased calculation method would be to multiply the observed average precipitation rate by the ratio of the duration of the output accumulation period and the sum of the durations of the 30-minute input HDF files that contain non-missing data values for this  $0.1^\circ \times 0.1^\circ$  grid box. A still more sophisticated algorithm would use climatological information, perhaps including seasonal and diurnal statistics to fill in greater than zero accumulation for the 30-minute periods that contain missing-data values in the input HDF5 files. The IMERG GIS algorithm, however, avoids making such scientific judgement calls.

### 3.8. Bias-correction related to missing data

This section describes a new bias-correction algorithm that PPS anticipates adding to IMERG GIS in 2022 when Version 7 of the IMERG HDF5 files begin being produced at PPS. In



2022, PPS will implement a bias correction suggested by the team of scientists who developed the IMERG algorithm. The bias correction is only relevant when creating IMERG GIS files from input files that are 30-minute-duration IMERG HDF5 files. The need for bias correction comes from the fact that some  $0.1^\circ \times 0.1^\circ$  grid boxes in a 30-minute HDF5 file may contain a missing-data value rather than a precipitation-rate estimate.

The estimated accumulation in an IMERG GIS file would be an undercount if precipitation had occurred during any half-hour period for which the 30-minute HDF5 file contains a missing-data value. If a large number of half-hour periods are filled with missing-data values then it is not safe to extrapolate from the available periods to estimate the total accumulation. In a typical month, for example, approximately 20% of grid boxes have missing-data values for all 30-minute periods during the month, primarily near the North Pole or South Pole. Approximately 4% of grid boxes have only a small number of half-hour periods (10% of the month's worth of half-hour periods) with missing data values.

Specifically, the IMERG GIS bias correction attempts to improve the accumulation and average precipitation rates for grid boxes where no more than 10% of half-hour periods have missing-data values. The correction method has two parts. First, the average precipitation rate is calculated from available half-hour periods, i.e., periods without missing-data values. Second, the accumulation is estimated as the just-calculated average rate times the total duration of the period in hours.

### **3.9. Geographic boundaries**

IMERG HDF5 files and IMERG GIS files have geographic boundaries of  $90^\circ\text{S}$  to  $90^\circ\text{N}$  latitude and  $180^\circ\text{W}$  to  $180^\circ\text{E}$  longitude. Many grid boxes near the poles are filled with the missing-data value. IMERG has a  $0.1^\circ$  spatial resolution. The southernmost grid boxes cover areas bounded by  $90^\circ\text{S}$  to  $89.9^\circ\text{S}$  latitude. The westernmost grid boxes cover areas bounded by  $180^\circ\text{W}$  to  $179.9^\circ\text{W}$  longitude.

The first grid box, grid box 0 in both the x and y dimension, should be displayed at the upper left of the image, i.e., the northwest corner. This orientation of the grid matches the common convention for images, and is vertically flipped from the orientation that the Interactive Data Language (IDL) uses for displaying two-dimensional arrays of data values. IDL's default orientation is the mathematical convention that the origin of a two-dimensional coordinate system is at the lower left corner of the plot when plotting in the first quadrant.

### **3.10. Format of geographic metadata in IMERG GIS GeoTIFFs and WorldFiles**

Two geographic coordinate transformations are needed to display IMERG data on a high-resolution map. Only the first transformation is needed to make a lower-resolution map that covers the globe or a large portion of it. The first transformation converts from array indices of

the image to latitude-longitude coordinates. The parameters of this transformation may be specified by either the contents of the ESRI WorldFile or by a GeoTIFF file's ModelPixelScaleTag and ModelTiePointTag. The second transformation converts from latitude-longitude coordinates to physical features on the Earth's surface by means of a spatial reference system (SRS). The SRS used by IMERG is called World Geodetic System 1984 (WGS84). WGS84 is used by many satellite data products and also many global positioning systems (GPS).

In the language used in the GeoTIFF specification, the first of these two transformations is the raster-to-model transformation. The second transformation is the model-to-Earth transformation. See Section 2.7 of the GeoTIFF specification, version 1.0.1, <http://geotiff.maptools.org/spec/geotiff2.7.html>.

The first transformation, the raster-to-model transformation, can also be called a map projection. The IMERG grid the "Plate Carree" map projection which may also be called "Equidistance Cylindrical" (USGS 1989). For the map projection to be Plate Carree,  $r$  must be the same in the equations for latitude and longitude. In other words, a linear scaling occurs by a factor  $r$  to convert grid-box indices  $(x,y)$  to latitude and longitude  $(lat,lon)$ :  $lon = r x + x_0$  and  $lat = r y + y_0$  (Bugayevskiy and Snyder 1995, pg. 55). Here  $(x_0,y_0)$  is the upper left corner of the first grid box, which is the absolute northwestern edge of the area covered by the entire grid. For an array whose elements represent areas (i.e., a grid whose elements are grid boxes), as is the case for IMERG, the GeoTIFF specification defines the raster coordinate as going from 0 to 1 as one passes through the first grid box. The center of the first grid box has a coordinate of 0.5. Elaborating on this point, the version 1.0 GeoTIFF specification is quoted here at Section 2.5.2.2 "Raster Space" (<http://geotiff.maptools.org/spec/geotiff2.5.html#2.5.2>):

"PixelIsArea" Raster Space

The "PixelIsArea" raster grid space R, which is the default, uses coordinates I and J, with (0,0) denoting the upper-left corner of the image, and increasing I to the right, increasing J down. The first pixel-value fills the square grid cell with the bounds:

top-left = (0,0), bottom-right = (1,1)

and so on; by extension this one-by-one grid cell is also referred to as a pixel. An N by M pixel image covers an area with the mathematically defined bounds (0,0), (N,M).

```
(0,0)t
+---+---+> I
| * | * |
+---+---+ Standard (PixelIsArea) TIFF Raster space R,
| (1,1) (2,1) showing the areas (*) of several pixels.
|
J
```

Following this convention, the raster coordinate (0,0) is the upper left corner of the first grid box, i.e., the absolute extreme of the overall coverage of the image. The GeoTIFF convention is consistent with the convention used in C and IDL that the first index of an array is zero. In

contrast, Matlab and a few other languages use 1 as the first index of an array, not zero. In such one-based languages, GeoTIFF metadata may be represented somewhat differently after being read into memory than it is represented in zero-based languages.

The Plate Carree map projection is one example of a simple-cylindrical map projection. Plate Carree is the specific simple-cylindrical map projection in which the location at which the grid boxes are most approximately square is along the Equator. This property is a natural consequence of coefficient  $r$  being identical in the definitions of latitude and longitude. The term "Equirectangular" is sometimes used to describe the Plate Carree map projection, but the term Equirectangular is best avoided because Equirectangular may also refer for a different map projection. In this other map projection, the minimal distortion occurs at 30°N and 30°S latitude rather than at the Equator (Snyder 1989, pp. 22 and 24), i.e., a different value of  $r$  is used in the definitions of latitude and longitude.

IMERG GIS represents the raster-to-model transformation in two ways: with an ESRI WorldFile and with tags and GeoKeys in the GeoTIFF file. An ESRI WorldFile identifies the latitude and longitude of the center of a single grid box, the northwesternmost grid box. The WorldFile also specifies the size of each grid box in degrees latitude and longitude. For IMERG, the center of the northwesternmost grid box is at 89.95°N 179.95°W and each grid box is 0.1° by 0.1°, as shown in Table 3.

**Table 3.** Contents of an IMERG GIS WorldFile.

Contents of File	Meaning
0.1	The increase in degrees longitude moving east one grid box
0	Rotation
0	Rotation
-0.1	The increase in degrees latitude moving one grid box to the south
-179.95	Degrees east longitude of the center of the northwesternmost grid box
89.95	Degrees north latitude of the center of the northwesternmost grid box

The tags in the GeoTIFF file are intended to be interpreted automatically by GIS or other software. Should it be necessary to manually interpret these three tags, they are described here. Two TIFF tags (`ModelPixelScaleTag` and `ModelTiePointTag`) define the raster-to-model transformation while another TIFF tag (`GeoKeyDirectoryTag`) defines the model-to-Earth transformation. Using pseudocode, the values for the raster-to-model transformation are as follows based on the northwesternmost extent of the grid (`latMax = 90°N`, `lonMin = -180°E`) and a resolution of `gridboxSize = 0.1°`:

```

upperLeftRasterCoordinates = double( [0,0,0] )
upperLeftLonLatElev = double( [lonMin,latMax,0] )
modelTiePointTag = [upperLeftRasterCoordinates, $
    upperLeftLonLatElev ]
modelPixelScaleTag = double(gridboxSize) * [1,1,0]

```

For the model-to-Earth transformation, Table 4 describes the three GeoKeys stored within the TIFF tag called GeoKeyDirectoryTag.

**Table 4.** Model-to-Earth transformation parameters stored in the TIFF tag called GeoKeyDirectoryTag.

GeoKey Name	Value	Meaning stated in GeoTIFF specification
GTModelTypeGeokey	2	The general category of latitude/longitude coordinates is "geographic" which is identical to "geodetic"
GTRasterTypeGeokey	1	Each value in the array represents an area, i.e., the image consists of grid boxes not points
geographicTypeGeokey	4326	The specific kind of latitude/longitude is WGS84

In the IDL language, one creates a GeoTIFF file by placing all of the raster-to-model and model-to-Earth parameters in a single data structure passed to IDL's `write_tiff` routine. The two-dimensional array of two-byte integer precipitation rate is stored in the data variable, the filename is stored in the string `fileName`, and the variables `pixelSize` and `ModelTiePointTag` are defined above.

```

metadata = { $
    ModelPixelScaleTag: double(gridboxSize) * [1,1,0], $
    ModelTiePointTag: ModelTiePointTag, $
    $
    $ ; -- values stored in the GeoKeyDirectoryTag Tiff tag
    GTModelTypeGeokey: fix(2), $
    GTRasterTypeGeokey: fix(1), $
    geographicTypeGeokey: fix(4326) $
}
write_tiff, outputFile, data, compression=1, /short, $
    geotiff=metadata

```

In Python, one of the ways to create a GeoTIFF file is to use the PIL and numpy libraries. Data is a two-dimensional numpy array, and the `fileName` is string variable. A Python 3 code snippet is below:

```

from PIL import Image
import numpy
image = Image.fromarray( data )
metadata = { 'ModelPixelScaleTag': (0.1, 0.1, 0.0), \
    'ModelTiePointTag': (0.0, 0.0, 0.0, -180.0, 90.0, 0.0) \

```

```
'GeoKeyDirectoryTag': (1, 1, 0, 3, \
    1024, 0, 1, 2, 1025, 0, 1, 1, 2048, 0, 1, 4326) }
image.save( fileName, format='TIFF', \
    compression='tiff_adobe_deflate', tiffinfo=metadata )
```

MacOS installations of Anaconda Python 3 may come with a `tiffdump` Linux command-line utility that shows the geographic information in a TIFF file. The first column of the `tiffdump` output is the TIFF tag name, the second column is the TIFF tag number, which is useful occasionally for reading a specific tag using Python. The output of the `tiffdump` utility for a Version 7 IMERG GIS file includes the following lines:

```
ModelPixelScaleTag      33550      8-byte float      0.1 0.1 0
ModelTiePointTag        33922      8-byte float      0 0 0 -180 90 0
GeoKeyDirectoryTag      34735      2-byte integer    1 1 0 3 1024 0 1 2 1025 0 1
    1 2048 0 1 4326
```

Another way to interrogate the metadata of a GeoTIFF file on MacOS is to use the Anaconda Python 3 installer (`conda`) to install the GDAL C library and Linux command-line utilities. To do so, use the following command:

```
conda install -c conda-forge gdal
```

After this installation, the `gdalinfo` utility becomes available on the Linux command line. While the `gdalinfo` utility's output format may change, it is something like the following, combining information actually read from the GeoTIFF file and other information that `gdalinfo` deduces:

```
Size is 3600, 1800
Coordinate System is: GEOGCRS["WGS84",
    DATUM["World Geodetic System 1984",
        ELLIPSOID["WGS84",6378137,298.257223563,
            LENGTHUNIT["metre",1]]],
    PRIMEM["Greenwich",0, ANGLEUNIT["degree",0.0174532925199433]],
    CS[ellipsoidal,2],
    AXIS["geodetic latitude (Lat)",north, ORDER[1],
        ANGLEUNIT["degree",0.0174532925199433]],
    AXIS["geodetic longitude (Lon)",east, ORDER[2],
        ANGLEUNIT["degree",0.0174532925199433]],
    USAGE[ SCOPE["unknown"], AREA["World"], BBOX[-90,-180,90,180]],
    ID["EPSG",4326]]
Data axis to CRS axis mapping: 2,1
Origin = (-180.00000000000000,90.00000000000000)
Pixel Size = (0.100000001490116,-0.100000001490116)
Metadata: AREA_OR_POINT=Area
```

In an IMERG GIS file, the elements of the `GeoKeyDirectoryTag` TIFF tag are described in the following table.

**Table 5.** For a Version 7, IMERG GIS file, the GeoKeyDirectoryTag TIFF Tag contents, i.e., its header and GeoKeys.

Value in Tag	Meaning
<i>Header</i>	
1	KeyDirectoryVersion
1	KeyRevision
0	MinorRevision
3	NumberOfKeys, i.e., the number of subsequent blocks of 4 integers
<i>GeoKey 1</i>	
1024	KeyID (ID 1024 represents GTModelTypeGeoKey)
0	TIFFTagLocation (0 = stored in this tag)
1	Number of elements
2	Value of GeoKey is 2 which means "Geographic latitude-longitude System" model
<i>GeoKey 2</i>	
1025	KeyID (ID 1025 represents GTRasterTypeGeoKey)
0	TIFFTagLocation (0 = stored in this tag)
1	Number of elements
1	The value of 1 represents RasterPixellIsArea as opposed to RasterPixellIsPoint
<i>GeoKey 3</i>	
2048	KeyID (ID 2048 represents GeographicTypeGeoKey)
0	TIFFTagLocation (0 = stored in this tag)
1	Number of elements
4326	The value of 4326 represents GCS_WGS_84

## 4. How to Read the IMERG GIS Data Product



### 4.1. Read with ESRI ArcGIS

This section describes how to display IMERG GIS files in ESRI's ArcGIS Pro application. For decades, ArcMap was ESRI's entry-level GIS application, but ArcMap was superseded in

2015 by ArcGIS Pro. The following instructions were developed for ArcGIS Pro running on a Microsoft Windows system.

First, download an IMERG GIS \*.tif from the PPS online archive, as described in Section 2.1 of this document. It appears that ArcGIS Pro is able to read the geographic metadata stored in a GeoTIFF file (\*.tif) so it is not necessary to download the ESRI WorldFile (\*.tfw) associated with a GeoTIFF file.

Start ArcGIS Pro by typing "ArcGIS" into the Windows search bar. Select Open in the search results. In the ArcGIS window, you will be prompted to open or create an object before you will be allowed to read data into ArcGIS. You may choose to open a Blank Template consisting of a map. Give a name to that map and select a directory where ArcGIS will write that map.

In the blank map, click on the Map tab to make the Map ribbon visible across the top of the ArcGIS window. In the Map ribbon, find the Layer section. Click the Add Data button () in the Layer Section. If you are unfamiliar with appearance of this icon, hover over the various buttons in the Layer Section until one of the buttons reveals itself to be the Add Data button. Along the left side of the Add Data window, navigate the tree menu to locate the IMERG GIS \*.tif file that you previously downloaded from PPS. This may involve you selecting Computer, then This PC, then Downloads if the Downloads folder is where you downloaded the \*.tif file. If you do not see the name of the \*.tif file in the directory that you know contains this file, you may have to refresh ArcGIS. Unlike most Windows and MacOS programs, ArcGIS does not update its knowledge of what files are in a directory unless you tell ArcGIS to do so. Specifically, click the button with the circular arrow () that is located along the top of the Add Data window.

After you select a file to open, click No if ArcGIS asks if you want to build pyramids or calculate statistics. After opening and reading from the IMERG GIS TIFF file, ArcGIS will display this data on top of and obscuring the ESRI default basemap layer.

If for some reason ArcGIS cannot determine the coordinate system or appropriate map projection for a particular IMERG GIS \*.tif file, use the following procedure to establish this information. In the ArcGIS search bar, search for "Define Projection", and select the search result called "Define Projection (Data Management Tools)". In the "Geoprocessing: Define Projection" panel that appears, select the name of the IMERG GIS file in the Input Dataset pull-down menu. In the Coordinate System pull-down menu, select GCS\_WGS\_1984. If this option is not visible in the pull-down menu, then click on the globe icon, which is the Select Coordinate System button. In the Coordinate System pop-up window, select Geographic > World > WGS 1984.

To make the colors of the IMERG GIS file more informative, locate the IMERG GIS file in the Contents panel of the ArcGIS window. Double click on the IMERG GIS file to display that file's Symbology information in the Symbology panel of the ArcGIS window. In the Symbology panel, you may wish to change the Color Scheme from grayscale to rainbow colors. You may wish to click on the Mask tab near the bottom of the Symbology panel. In the Mask tab, check

the checkbox next to Display Background Value. Choose a data value of 0 to be displayed with a transparent color. Making this selection will allow the default basemap layer to show through the IMERG GIS layer where the IMERG GIS layer has zero precipitation.

By default, ArcGIS might choose a global Mercator map projection. If you prefer a global map projection that has degrees of latitude and longitude scaled equally, then use the following procedure to switch ArcGIS's map projection. In the Contents panel, right click on the Map object. Select the Properties menu item from the pop-up menu. In the Map Properties window that appears, select the Coordinate Systems menu item from the menu on the left-hand side of the window. If ArcGIS is currently using the "WGS 1984 Web Mercator" map projection, then that projection will be already selected in the list under XY Coordinate Systems Available. To select a map projection that uses equally scaled latitude and longitude, click on either "WGS 1984" or "Projected Coordinate System > World > Plate Carree."

## **4.2. Read with QGIS**

It is possible to display the IMERG GIS data product using the QGIS (formerly Quantum GIS) application. QGIS is an open-source GIS application that can be downloaded at no cost from <http://qgis.org>. The use of QGIS, and ArcMap for that matter, in this document is not an endorsement. These two GIS applications are mentioned here merely to illustrate the kinds of steps that one may need to take to display the IMERG GIS files in a GIS application.

To begin, download the \*.tif and \*.tfw file from the online archive, as described in Section 2.1 of this document. Start QGIS. Select the Layer>AddLayer>AddRasterLayer menu option. Browse to the \*.tif file of interest and click the Open button. QGIS will locate the accompanying \*.tfw file that you placed in the same directory. If the Coordinate Reference System Selector window appears, click OK to dismiss it. Once QGIS reads the GeoTIFF file, an entry for this file will appear in the data set list on the left side of the QGIS window and the data itself will appear in the main display portion of the QGIS window.

Most likely the default color table shown by QGIS will be far from optimal and perhaps just a black block covering the globe.

This paragraph gives a quick way to improve the QGIS color table, and the rest of the section provides a more detailed way to improve the color table. This initial example assumes that one is looking at Early or Late IMERG GeoTIFF files, although the procedure is similar for looking at Late IMERG GeoTIFF Files. Begin by calling up the Layer Property window for this GeoTIFF File. To do so, find the name of this GeoTIFF File in the Layer Panel that usually appears on the left side of the QGIS window. On this filename, right click and select Properties from the pop-up menu that appears. In the Layer Properties window, select the "Style" tab from the vertical list along the left side of the Layer Properties window. Then do four things within the Style tab. First, find the pull-down menu next to the Render Type label and select "Singleband pseudocolor." Second, click the box next to the "Invert" label so that an X symbol appears in that

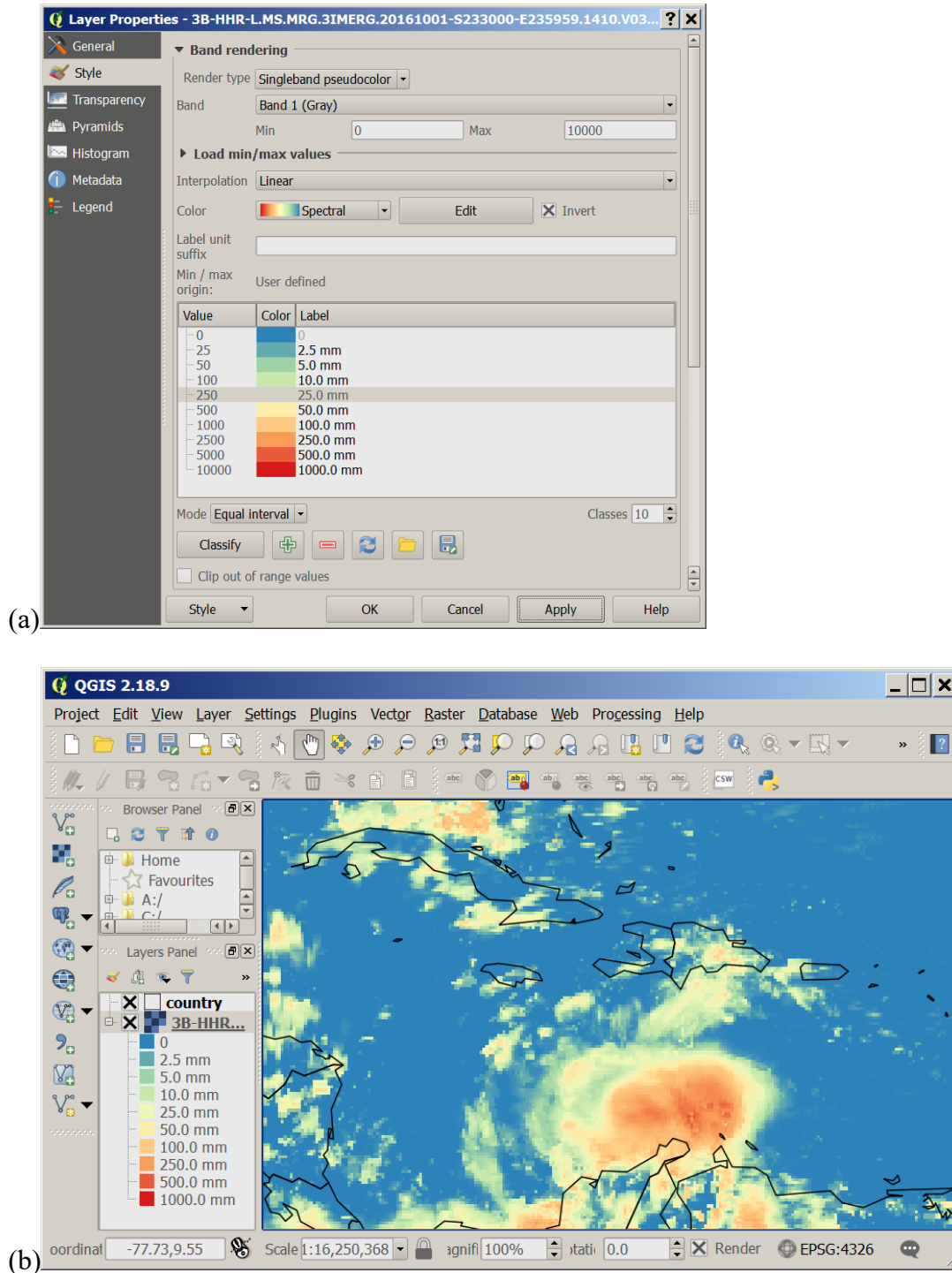


box. Third, type in an appropriate maximum value in the field next to the "Max" label. For 30-minute to 7-day accumulations (which are scaled by 10 in the GeoTIFF file), start off with a maximum of 10000; for 1-month accumulations (which are not scaled in the GeoTIFF file), start off with a maximum of 1000; and for percent liquid, start off with a maximum of 255 (since they are stored as unsigned 1-byte integers in the GeoTIFF file). Fourth, click the Apply button at the bottom of the Layer Properties window. The result should be an improved display of the data in the QGIS window.

One can, optionally, create a more refined color table by manually selecting the limits of each color in the color table. To do so, first change the Mode from "Continuous" to "Equal Interval" so that it becomes possible to increase the number of colors in the color table. Increase the Classes field to 10 colors. At this point, one could click the Apply button to preview what the display will look like. From this point on, this example will be for displaying 30-minute to 7-day accumulations, although a similar procedure could be followed for monthly accumulations or for percent liquid data.

Most locations will have precipitation near the low end of the color table. Precipitation rates or accumulations are most numerous near the zero end of the color table with a few widely-scattered large values because precipitation is approximately logarithmically distributed. In other words, precipitation is usually light (clustered near ~1 mm/h) but a few places on Earth at any given time might have precipitation that is one or two orders of magnitude higher (10 to 100 mm/h). The color table will better match the data set if most of the graduations are located near the lower end of the scale, as described here. Double-click on the number to the left of the second lightest precipitation rate (the rate one step heavier than the zero color) and change the number to 25. Double-click on the next lowest entry in the color table and change it to 50. Continue this process, choosing values of 100, 250, 500, 1000, 2500, 5000, and 10000 for the rest of the color table entries. After making these changes, the Style tab of the Layer Properties window will look something like Figure 5. Click the Apply button to see the updated, unequally-spaced color table.

To speed up the process the next time that you want to display an IMERG GIS file, select the Style menu at the bottom-left of the Layer Properties window and choose the Save Style menu item. In the file-selector window that pops up, type in a filename and click the Save button to dismiss the file-selector window. Note that the Style menu at the bottom-left of the Layer Properties window has a Load Style menu item. This menu item is the one to select in the future to reload the color table just created. In the Layer Property window, select the OK button to dismiss the Layer Property window. To view the color-table legend in the main QGIS window, click on the plus symbol to the left of the filename in the list of open files on the left side of the QGIS window.



**Figure 5.** Screen captures of QGIS displaying IMERG GIS files. (a) The Style tab of the QGIS Layer Properties window after an unequally-spaced color table is defined for displaying precipitation scaled by 10 and stored as a 2-byte integer in an Early or Late IMERG GIS accumulation for a period of 30 minutes to 7 days. (b) The QGIS main window with a 1-day Late IMERG GIS accumulation for 8 October 2016 showing Hurricane Matthew in the Caribbean.

Table 7 states the physical meaning of the color table cutoffs suggested in the previous paragraph (25, 50, ... 5000, 10000). The table also provides the physical meaning of these cutoffs, depending on what kind of IMERG GIS file is being displayed. Optionally, type in these physical meanings as color-table annotations so that they appear to the right of each color of the color table. To do so, double-click on the color table or use another means to open the Layer Properties window. Then, type the desired text labels in the field to the right of each color (Figure 5a). Click the Apply button and then the OK button to apply the labels (Figure 5b) and dismiss the Layer Properties window.

**Table 7.** For IMERG GIS files, the physical meaning of the suggested color-table cutoffs for precipitation variables displayed in QGIS.

Value in the *.tif file	Physical meaning			
	Precipitation accumulation		Precipitation rate	
	30-minute to 7-day	Monthly	30-minute or 1-day	Monthly
0	0 mm	0 mm	0 mm/h	0 mm/h
25	2.5 mm	25 mm	2.5 mm/h	0.025 mm/h
50	5.0 mm	50 mm	5.0 mm/h	0.050 mm/h
100	10.0 mm	100 mm	10.0 mm/h	0.100 mm/h
250	25.0 mm	250 mm	25.0 mm/h	0.250 mm/h
500	50.0 mm	500 mm	50.0 mm/h	0.500 mm/h
1000	100.0 mm	1000 mm	100.0 mm/h	1.000 mm/h
2500	250.0 mm	2.5 meters	250.0 mm/h	2.500 mm/h
5000	500.0 mm	5.0 meters	500.0 mm/h	5.000 mm/h
10000	1000.0 mm	10.0 meters	1000.0 mm/h	10.000 mm/h

#### 4.3. Read with Python

TIFF files are easy to read in Python. It is more difficult to have Python interpret the geographic metadata in a GeoTIFF file so that the data can be placed on a map. The difficulty arises that one has to either install a Python library that includes GDAL, which has a reputation for being hard to install, or one has to use a less-well-established non-GDAL alternative. Instructions for installing GDAL Python support is given here: <https://opensourceoptions.com/blog/how-to->

[install-gdal-with-anaconda/](#). As of 2022, GDAL is not part of default install of Anaconda Python 3. In other words, one cannot successfully type "import gdal" at Python interactive prompt of a default Anaconda Python 3 installation.

A simple TIFF file reader gives access to all tags including those that store geographic metadata, although such a reader does not manipulate these tags. For example, PIL library may be used, which is part of Anaconda Python 3.

```
from PIL import Image
import numpy
fileHandle = Image.open(fileName)
data = numpy.array( fileHandle )
data.shape
tagNumberList = [ oneNum for oneNum in fileHandle.tag ]
fileHandle.tag[33550]
fileHandle.tag[33922]
fileHandle.tag[34735]
```

The trick with using the PIL library is that one needs to know the numerical ID of the TIFF tag of interest. As described above in Section 3.10, the `tifinfo` Linux utility can be used to determine the IDs of the tags of interest in an IMERG GIS TIFF file.

Another non-GDAL approach in Python 3 is to use the `tifffile` library. Unlike PIL, the `tifffile` library can access TIFF tags by name, as shown below:

```
import tifffile, numpy
with tifffile.TiffFile(fileName) as fileHandle:
    tifTagDict = {}
    for tagObj in fileHandle.pages[0].tags.values():
        tifTagDict[ tagObj.name ] = tagObj.value
    data = numpy.array( fileHandle.pages[0].asarray() )
    tifTagDict.keys()
    tifTagDict[ 'ModelPixelScaleTag' ]
    tifTagDict[ 'ModelTiepointTag' ]
    tifTagDict[ 'GeoKeyDirectoryTag' ]
```

A third non-GDAL alternative in Python 3 is the `geotiff` library. This library is in the early stages of development. With Anaconda Python 3 installed on a MacOS system, one would type `pip install geotiff` to install the package. If one has not already added these Python packages to the Anaconda install, installing the `geotiff` library will also install them: `numcodecs`, `fasteners`, `asciitree`, `zarr`, and `pyproj`. More information on the `geotiff` Python library here:

<https://kipcrossing.github.io/2021-01-04-geotiff-python-package/> and <https://github.com/KipCrossing/geotiff>. The Python `geotiff` library can be used in the following way:

```
# -- read data from file
import numpy
from geotiff import GeoTiff
fileHandle = GeoTIFF( fileName, as_crs=None )
# -- print metadata
```

```

fileHandle.crs_code
fileHandle.tif_shape
fileHandle.tif_bBox
data = numpy.array( fileHandle.read() )
data.shape
# -- plot data
import matplotlib.pyplot as pyplot
pyplot.imshow( data )
pyplot.show()

```

One operation that is frequently useful is to subset the global IMERG grid to retain a rectangular region in which the grid boxes have cells centers fall within a user-specific latitude range and longitude range. The following code snippet shows one way to perform such geographic subsetting in Python.

```

from PIL import Image
import numpy

# Specify IMERG GIS filename and lat/lon region of interest
file = '3B-DAY-GIS.MS.MRG.3IMERG.20210101-S000000-E235959.0000.V06B.tif'
latDegree = [-1,0]
lonDegree = [105,105.5]

# Read the tiff file, convert to numpy array, and unscale to
# get millimeters. This will give you a 1800 x 3600 element array
# that covers the globe at 0.1-degree latitude-longitude resolution.
scale = 10.0
dataGlobal = numpy.asarray( Image.open(file), dtype='float' ) / scale

# Determine the lat/lon integer indices for the region of interest
latDegree = numpy.asarray( latDegree, dtype='float' )
lonDegree = numpy.asarray( lonDegree, dtype='float' )
iLat = numpy.int16( numpy rint( (latDegree +90) * 10 ) )
iLon = numpy.int16( numpy rint( (lonDegree +180) * 10 ) )

# Diagnostic print out of lat/lon values of cell centers in region
latCenter = numpy.fromiter(range(1800),dtype='float')/10.0 -89.95
lonCenter = numpy.fromiter(range(3600),dtype='float')/10.0 -179.95
print( 'latitudes of cell centers... ' )
latCenter[ iLat[0]:iLat[1] ]
print( 'longitude of cell centers... ' )
lonCenter[ iLon[0]:iLon[1] ]

# Geographically subset data grid to region of interest
dataSubset = dataGlobal[ iLat[0]:iLat[1] , iLon[0]:iLon[1] ]
print( 'shape of geographic subset... ' )
dataSubset.shape

```

#### 4.4. Read with Matlab

In Matlab, it is best to use a TIFF image reader function called `imread()`, to read the data array from an IMERG GIS GeoTIFF file. To obtain the associate geographic metadata, one has

two options: define a structure stating the grid box center locations using Matlab's `georefpstings()` function or read the WorldFile associated with the GeoTIFF image.

```
% Read 1800 x 3600 element data array
data = imread( fileName ) ;

% Read metadata: Method #1
metadata=georefpstings([-90+0.05,90-0.05],[-180+0.05,180-0.05],0.1,0.1);

% Read metadata: Method #2
metadata = worldfileread(worldFileName,'geographic',size(data)) ;

% Display data
mapshow( data, metadata )
```

The `worldfileread()` function is only available to Matlab users who have purchased the optional Matlab Mapping Toolbox. While working with the data array, keep in mind that the first index of a Matlab array is 1, whereas, in most programming languages (C, IDL, or Python), the first array index is 0.

As of Matlab version 2021a, the Matlab high-level function for reading the image and metadata from a GeoTIFF file does not work on IMERG GIS GeoTIFF files. In other words, the `[data,metadata] = readgeoraster(fileName)` function call will return a blank metadata structure. Furthermore, the `readgeoraster()` function is only available to Matlab users who have purchased the optional Matlab Mapping Toolbox.

The source of the difficulty appears to be that `readgeoraster()` is unable to interpret GeoTIFF metadata for images in which their pixels represent grid boxes, i.e., values that are an average over an area. The IMERG precipitation rates do, in fact, represent an estimate of the area-averaged precipitation rate over the full extent of a  $0.1^\circ \times 0.1^\circ$  grid box. The `readgeoraster()` function appears to be able to process metadata for images only where the image pixels represent point values, not areas. Specifically, Matlab can read GeoTIFFs with a value of 2 instead of 1 for the `GTRasterTypeGeoKey` TIFF GeoKey. According to the GeoTIFF format specification, a value of 2 for this GeoKey would signify `RasterPixelsPoint` while a value of 1 signifies `RasterPixelsArea`.

## 4.5. Read with IDL

In the Interactive Data Language (IDL), it is easy to read the data and metadata from a GeoTIFF File, but one must manually interpret the metadata.

```
; -- read information from file
data = read_tiff( fileName, geotiff=metadata)
help, /structure, data, metadata
; -- calculate map limit from metadata in file
deltaDeg = metadata.modelPixelScaleTag[0]
```

```

latMax = metadata.modelTiePointTag[4]
lonMin = metadata.modelTiePointTag[3]
dimSize = size(data,/dimension)
latRange = float( latMax -[0,dimSize[1]]*deltaDeg )
lonRange = float( lonMin +[0,dimSize[0]]*deltaDeg )
mapset_limit = [ latRange[0], lonRange[0], latRange[1], lonRange[1] ]
; -- display it with continent outlines
image = alog10( data > 0.1 )
window,0, xsize=2*360,ysize=2*180
tvsc1, congrid(image,2*360,2*180), order=1
map_set, limit=mapset_limit, /noborder, position=[0,0,1,1], /noerase
map_continents, color=255

```

## 6. References

- Aldus Corporation, 1992: *TIFF, Revision 6.0, Final – June 3, 1992*.  
<https://www.adobe.io/content/dam/udp/en/open/standards/tiff/TIFF6.pdf>. Copyright to this document now held by Adobe Systems Incorporated.
- Data Product Development Group (DPDG) Working Group, July 2020: *Data Product Development Guide for Data Producers*. NASA, report ESDS-RFC-041,  
<https://cdn.earthdata.nasa.gov/conduit/upload/14909/ESDS-RFC-041.pdf>.
- ESRI, 2022: What is the format of the world file used for georeferencing images? web page,  
<https://support.esri.com/en/technical-article/000002860>.
- Hong, Y., R. F. Adler, A. Negri, and G. J. Huffman, 2007: Flood and landslide applications of near real-time satellite rainfall products. *Nat. Hazards*, **43**, 285–294.
- Huffman, G. J., D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, S. Sorooshian, J. Tan, and P. Xie, 2019a: *NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG), Algorithm Theoretical Basis Document (ATBD)*. [https://pps.gsfc.nasa.gov/Documents/IMERG\\_ATBD\\_V06.pdf](https://pps.gsfc.nasa.gov/Documents/IMERG_ATBD_V06.pdf).
- Huffman, G. J., D. T. Bolvin, E. J. Nelkin, and J. Tan, 2019b: *Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation*. White paper.  
[https://pps.gsfc.nasa.gov/Documents/IMERG\\_doc\\_190313.pdf](https://pps.gsfc.nasa.gov/Documents/IMERG_doc_190313.pdf).
- Huffman, G. J., R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM Multi-satellite Precipitation Analysis (TMPA). Chapter 1 of *Satellite Rainfall Applications for Surface Hydrology*. F. Hossain and M. Gebremichael, Eds. Springer Verlag, ISBN: 978-90-481-2914-0, 3–22.

- Huffman, G. J., R. F. Adler, D. T. Bolvin, G. Gu, E. J. Nelkin, K. P. Bowman, Y. Hong, E. F. Stocker, and D. B. Wolff, 2007: The TRMM Multi-satellite Precipitation Analysis: Quasi-global, multi-year, combined-sensor precipitation estimates at fine scale. *J. Hydrometeor.*, **8**, 38–55. [https://link.springer.com/chapter/10.1007/978-90-481-2915-7\\_1](https://link.springer.com/chapter/10.1007/978-90-481-2915-7_1).
- Kelley, O., 2022: *The IMERG multi-satellite precipitation estimates reformatted as 2-byte GeoTIFF files for display in a Geographic Information System (GIS)*. NASA Goddard Precipitation Processing System (PPS), 43 pp, <https://arthurhou.pps.eosdis.nasa.gov/Documents/README.GIS.pdf>.
- Open Geospatial Consortium (OGC), 2019: *OGC GeoTIFF Standard*. Version 1.1, <http://www.opengis.net/doc/IS/GeoTIFF/1.1>.
- Ritter, N., and M. Ruth, 1995: *GeoTIFF Format Specification: GeoTIFF Revision 1.0*. JPL, 95 pp., <https://gis-lab.info/docs/geotiff-1.8.2.pdf>.
- Snyder, J. P., and L. M. Bugayevskiy, 1995: *Map Projections: A reference Manual*. CRC Press, 328 pp.
- U.S. Geological Survey, 1989: *An Album of Map Projections*. USGS, professional paper #1453. 249 pp.
- Wikipedia, 2021: "World file." Entry in online encyclopedia. [http://en.wikipedia.org/wiki/World\\_file](http://en.wikipedia.org/wiki/World_file).