GeoTIFF files for the Version 7 Final IMERG Climatology

Owen Kelley (owen.kelley@nasa.gov) for the Precipitation Processing System (PPS), 27 August 2024

1. Introduction

This document describes the GeoTIFF translation of a precipitation climatology based on Version 7 Final IMERG. IMERG stands for the Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM), and it is arguably NASA's most advance algorithm for estimating precipitation around the world. IMERG has provided 30-minute global 0.1-degree-resolution precipitation estimates for the past two decades based on observations by a constellation of Earth-observing satellites and several calibration datasets (Huffman 2023a,b,c). The term "Final" means that the climatology is based on the IMERG run that is created 3–4 months after satellite observations are collected so that more calibration data can be included.

In the spring of 2024, Jackson Tan created the Version V07B Final IMERG NetCDF4 climatology files, averaging by day of year or month of year (Tan and Huffman 2024). He also calculated a long-term average for the entire data period. Tan's NetCDF4 files, along with his Python 3 script that created them, can be downloaded from the following web page:

https://gpm.nasa.gov/data/imerg/precipitation-climatology

The IMERG NetCDF4 climatology uses as input the GES DISC 24-hour average IMERG NetCDF4 files. GES DISC stands for Goddard Earth Sciences Data and Information Service Center. These daily NetCDF files, in turn, use as input the half-hour IMERG HDF5 files that are generated directly by the IMERG algorithm at NASA Goddard's Precipitation Processing System (PPS). In January 2024, GPM released the Version 07B Final IMERG data product in the HDF5 format, which included the half-hour and monthly files covering June 2000 through a few months before the present.

In 2021, a NASA working group recommended that future NASA data products be released in one of two file formats: NetCDF4 and GeoTIFF (NASA 2021a). NetCDF4 is a file input/output library built on top of the HDF5 input/output library. From 2007 to 2020, NASA had recommend HDF5 for NASA data products (NASA 2007, 2016, 2020, 2021b). In addition to the IMERG climatology there are other kinds of IMERG files that have been translated to the GeoTIFF format. In addition, there is a GeoTIFF translation of the original half-hour and monthly IMERG files (Kelley 2023).

In general, a GeoTIFF file is a TIFF image file with geographic metadata fields that are encoded as TIFF tags and that follow conventions established in the GeoTIFF format specification. GeoTIFF files can be read by many Geographic Information System (GIS) applications. A variety of high-level programming languages, such as Python, IDL, and Matlab, can read and write GeoTIFF files (Kelley 2023, 2022). In contrast to the ease of reading GeoTIFF files, it can be challenging to write them with correctly formatted metadata. The text below describes IDL code developed at PPS that can translate the IMERG NetCDF climatology files to the GeoTIFF format.

To verify that an IMERG climatology GeoTIFF file contains valid metadata and the expected data, one can read it with a GIS application or a Linux command-line tool. Section 3.2 below describes how to test GeoTIFF files using the QGIS application, which is available for download at no cost from https://www.qgis.org/.

Some applications have trouble reading the geographic metadata stored inside GeoTIFF files (Stocker and Kelley 2007). One solution to this problem is to download an ESRI WorldFile along with the TIFF file (ESRI 2024). A WorldFile is a small text file containing six numbers. The WorldFile associated with IMERG's 3600×1800-element longitude-latitude grid contains the following text:

```
0.10000000
0.00000000
0.00000000
-0.10000000
-179.95000000
89.95000000
```

The filename for a WorldFile should be identical to the TIFF file with which you want to associate it, except that the WorldFile's filename extension should be "tfw" instead of "tif". That is the last three characters of the WorldFile's name should be "tfw" instead of "tif".

2. The IMERG Climatology Data Set

2.1. IMERG climatology NetCDF files

The IMERG climatology NetCDF files provides precipitation rate estimates on the same global 0.1º-resolution 3600-by-1800-element grid as the original HDF5 files created by the IMERG algorithm (Tan and Huffman 2024). At the time of writing, the day-of-month and month-of-year climatology files are averages of 22 years of data: 2001 through 2022. In contrast, the period for the overall average is 23 years: June 2000 through May 2023. The filenames of the IMERG climatology NetCDF files are as follows where *period* is "2001-2022" or "200006-202305" and *mm* and *dd* are the two-digit month and day of month:

period	filename pattern	units
day of year	IMERG-Final.CLIM.*period*.*mmdd*.V07B.nc4	mm day $^{-1}$
month of year	IMERG-Final.CLIM.*period*.*mm*.V07B.nc4	mm day $^{-1}$
overall average	IMERG-Final.CLIM.*period*.V07B.nc4	mm year $^{-1}$

The IMERG Climatology NetCDF files contains three floating-point arrays: a 1-dimensional array of north latitude values ("lat"), a 1-dimensional array of east longitude values ("lon"), and a 2-dimenstional array of average precipitation rate ("precipitation"). The latitude and longitude arrays state the location of a grid box center. The first element of these two arrays identifies the location of the northwestern-most grid box, i.e., +89.95º north latitude and -179.95º east longitude. The precipitation array contains a missing-data value of -9999.9 if no precipitation estimate is available for that grid box.

There is no documentation to describe how the Goddard Earth Sciences (GES) Data and Information Services Center (DISC) handles missing values in the half-hour HDF5 files when the GES DISC averages them to create daily NetCDF4 files. The daily NetCDF4 file documentation is available at DOI 10.5067/GPM/IMERGDF/DAY/07, which resolves to https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGDF_07/summary?keywords=IMERG.

Jackson Tan provides his Python code for converting GES DISC daily NetCDF4 files into his IMERG climatology NetCDF4 files. This code is available at the URL mentioned in the introduction section of the present document. Tan uses the 90% rule when determining which grid boxes in the climatology should be filled with the missing-data value of -9999.9. In the day-of-year climatology file, this means that Tan reports missing data (-9999.9) for any grid box for which less than 90% of the available years have valid (i.e., non-missing) data in that grid box. In other words, if at least 20 of the 22 years have valid values for that day of year, then he reports the value that is the sum of the daily precipitation accumulation (mm day^{-1}) in valid years for that grid box divided by the number of years with valid data for that grid box. The algorithm of Kelley (2023, section 3.4) that creates the IMERG GIS data product follows the same 90% rule for missing data that Tan uses for the IMERG climatology NetCDF4 files.

2.2. IMERG climatology GeoTIFF files

The IMERG climatology GeoTIFF files have a "tif" filename extension, and they contain the same floating point values as the NetCDF4 files. The `tiffdump` and `tiffinfo` Linux command-line utilities can print out the geographic metadata within an IMERG climatology GeoTIFF File. Machines with Anaconda Python 3 installed may already have a copy of the `tiffdump` and/or `tiffinfo` utilities. To check for presence of the tiffdump utility, type "`whereis tiffdump`" or look in `/opt/anaconda3/bin/`. The three TIFF tags in IMERG GeoTIFF files that contain geographic metadata have tag numbers of 33550, 33922, and 34735, which are associated with the fields called ModelPixelScaleTag, ModelTiePointTag, and GeoKeyDirectoryTag, respectively.

On a Linux machine, an easy way to extract just the three relevant tags from a GeoTIFF file called *filename* is the following:

```
tiffdump filename | grep -E '(33550|33922|34735)' "
```

...which creates output similar to the following:

```
33550 (0x830e) DOUBLE (12) 3<0.1 0.1 0>
33922 (0x8482) DOUBLE (12) 6<0 0 0 -180 90 0>
34735 (0x87af) SHORT (3) 24<1 1 0 5 1024 0 1 2 1025 0 1 1 2048 0 1
 4326 2057 34736 1 0 2059 34736 1 1>
```

If the `tiffinfo` command is used instead of `tiffdump`, is formatted in the following way:

```
Tag 33550: 0.100000,0.100000,0.000000
Tag 33922: 0.000000,0.000000,0.000000,-180.000000,90.000000,0.000000
Tag 34735: 1,1,0,5,1024,0,1,2,1025,0,1,1,2048,0,1,4326,2057,34736,
 1,0,2059,34736,1,1
```

The meaning of these tags is explained in section 3.10 of Kelley (2023).

2.3. Displaying IMERG climatology GeoTIFF files

To display an IMERG climatology GeoTIFF file, you can use the free GIS application called QGIS. It can be downloaded from https://www.qgis.org/. After you start the QGIS application, select the Layer > Add Layer > Add Raster Layer menu option. In the Data Source Manager window, go do the Source section and select the GeoTIFF file that you want to open. Then, click the Add button.

The data will now appear on the QGIS map and in the Layer panel. If the Layers panel is not visible, select the View > Panels > Layers menu item. Right click on the name of the GeoTIFF file in the Layers panel and select Properties from the pop-up menu. In the Layer Properties window, perform the following steps working from top to bottom:

1. Change the "Render type" pull-down menu to "Singleband pseudocolor".
2. Change the Minimum field from -9999.9 to 0.0 so that missing data does not take up most of the range of colors in the color bar. Choose a Maximum value of something like 20 or 3000 depending on if you are looking at a daily/month average or the long-term average, respectively.
3. Select "Color ramp" pull-down menu "Invert Color Ramp" item so that the intense red is used for heavy rain not light rain.
4. Click the Apply button at the bottom of the Layer Properties window so that the QGIS map is refreshed with the new settings that you just selected.

3. Writing GeoTIFF files

3.1. Using IDL to write GeoTIFF files

While it is straightforward to write data to a TIFF file, it can be challenging to calculate and write the correct values for the GeoTIFF metadata. The present section provides some insight into GeoTIFF metadata. The GeoTIFF specification is incredible flexible in terms of the map projections it can store (OGC 2019). This complexity can make it difficult to figure out how to write a gridded dataset like IMERG that has a simple projection, i.e., a regular grid of equally sized grid boxes in latitude and longitude. For this projection, a useful example is buried in Appendix F of version 1.1 of the GeoTIFF specification (OGC 2019):

> **F.2.4. DMA ADRG Raster Graphic Map.** The U.S. Defense Mapping Agency (now NGA) produces ARC digitized raster graphics datasets by scanning maps and geometrically resampling them into an equirectangular projection, so that they may be directly indexed with WGS 84 geographic coordinates. The upper left corner is 120 degrees West, 32 degrees North. The scale for one map is 0.2 degrees per pixel horizontally, 0.1 degrees per pixel vertically. If stored in a GeoTIFF file it contains the following information:

```
ModelTiepointTag=(0.0, 0.0, 0.0, -120.0, 32.0, 0.0)
ModelPixelScaleTag = (0.2, 0.1, 0.0)
GeoKeyDirectoryTag:
 GTModelTypeGeoKey = 2 (ModelTypeGeographic 2D)
 GTRasterTypeGeoKey = 1 (RasterPixelIsArea)
 GeodeticCRSGeoKey = 4326 (Geographic 2D WGS 84)
```

Define the geographic metadata for the IMERG grid with the following IDL commands:

```
latMax = 90.0 ;degrees
lonMin = -180.0 ;degrees
```

```
gridboxSize =0.1 ;degrees
upperLeftRasterCoordinates = double( [0,0,0] )
upperLeftLonLatElev = double( [lonMin,latMax,0] )
modelTiePointTag = [upperLeftRasterCoordinates, $
 upperLeftLonLatElev ]
modelPixelScaleTag = double(gridboxSize) * [1,1,0]

GTModelTypeGeokey = 2 ;geographic coordinates
GTRasterTypeGeokey = 1 ;each value represents and area not a point
geographicTypeGeokey = 4326 ;lat/lon defined by WGS84
```

Pack the above defined variables in a structure and write the floating-point data and metadata with the following IDL commands:

```
metadata = { $
 latMax:latMax, lonMin:lonMin, gridboxSize:gridboxSize, $
 GTModelTypeGeokey: GTModelTypeGeokey, $
 GTRasterTypeGeokey: GTRasterTypeGeokey, $
 geographicTypeGeokey: geographicTypeGeokey $
}
write_tiff, outputFile, dataFloat, compression=1, /float, $
 geotiff=metadata
```

3.2. Using Python 3 to write GeoTIFF files

In Python 3, there are multiples possible libraries to use to write GeoTIFF files, but the majority of them are built on top of the GDAL library, which can make them challenging to install. By using a combination of the GDAL-free PIL and tifffile libraries, one can write both floating-point data and the required geographic metadata to a GeoTIFF file. Use tifffile to write the data because only tifffile can handle floating-point data. As a second step, use PIL to write metadata because only PIL correctly writes GeoTIFF-specific TIFF tags. The following Python 3 code illustrates this two-step process.

```
# -- write floating-point data to an ordinary TIFF file
import tifffile
tifffile.imwrite( outputFile, dataFloat )

# -- add GeoTIFF metadata to that TIFF file
from PIL import Image
image = Image.open( outputFile )
```

```
metadata = dict( )
metadata[33550] = (0.1, 0.1, 0.0) #--ModelPixelScaleTag
metadata[33922] = (0.0, 0.0, 0.0, -180.0, 90.0, 0.0) #--ModelTiepointTag
metadata[34735] = (1, 1, 0, 3, \
 1024, 0, 1, 2, 1025, 0, 1, 1, 2048, 0, 1, 4326) #--GeoKeyDirectoryTag
image.save( outputFile, tiffinfo=metadata )
```

References

ESRI, 2024: World files for raster datasets. web page, https://pro.arcgis.com/en/pro-app/latest/help/data/imagery/world-files-for-raster-datasets.htm.

Huffman, G. J., D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, S. Sorooshian, J. Tan, and P. Xie, 2023a: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Version 07, Algorithm Theoretical Basis Document (ATBD). https://arthurhou.pps.eosdis.nasa.gov/Documents/IMERG_V07_ATBD_final.pdf.

Huffman, G. J., D. T. Bolvin, R. Joyce, O. A. Kelley, E. J. Nelkin, and J. Tan, D. C. Watters, and B. J. West, 2023b: Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation. white paper, https://arthurhou.pps.eosdis.nasa.gov/Documents/IMERG_TechnicalDocumentation_final.pdf.

Huffman, G. J., D. T. Bolvin, R. Joyce, O. A. Kelley, E. J. Nelkin, and J. Tan, D. C. Watters, and B. J. West, 2023c: IMERG V07 Release Notes. white paper, https://arthurhou.pps.eosdis.nasa.gov/Documents/IMERG_V07_ReleaseNotes_final.pdf.NASA, 2021a: *Data Product Development Guide for Data Producers*. white paper ESDS-RFC-041v1.1, https://www.earthdata.nasa.gov/s3fs-public/2022-06/ESDS-RFC-041-DPDG_V1.1-20220516_0.pdf.

Kelley, O. A., 2023: The IMERG multi-satellite precipitation estimates reformatted as 2-byte GeoTIFF files for display in a Geographic Information System (GIS). white paper, https://pps.gsfc.nasa.gov/Documents/README.GIS.pdf.

Kelley, O. A., 2022: Using Python to Read Data Products from the Global Precipitation Measurement (GPM) Mission. white paper, https://arthurhou.pps.eosdis.nasa.gov/Documents/gpmPythonNotes.pdf.

NASA, 2021b: HDF5 Data Model, File Format and Library—HDF5 1.6. web page, https://www.earthdata.nasa.gov/esdis/esco/standards-and-practices/hdf5. Quote text that appears to have been written in 2007, concurrent with NASA (2007): "HDF and HDF-EOS data formats, software libraries and application programming interfaces (APIs), have been widely used for NASA earth observation mission data for many years. The latest version of HDF, HDF5 is the current or planned data format for missions including.... Overall, HDF5 is a widely used data format with a well-defined specification that provides a standard way of storing and working with science data. The ESDS-RFC-007 TWG [e.g., NASA 2007] thus recommends its endorsement by the SPG as an Earth Science Data Systems Standard."

NASA, 2020: *Dataset Interoperability Recommendations for Earth Science: Part 2*, version 1.2. white paper ESDS-RFC-036v1.2, https://www.earthdata.nasa.gov/s3fs-public/imported/ESDS-RFC-036v1.2.pdf.

NASA, 2016: *Dataset Interoperability Recommendations for Earth Science*, version 1.3. white paper, ESDS-RFC-028v1.3, https://www.earthdata.nasa.gov/s3fs-public/imported/ESDS-RFC-028v1.3.pdf.

NASA, 2007: *HDF5 Data Model, File Format and Library – HDF5 1.6*, version 1. white paper ESDS-RFC-007v1, https://www.earthdata.nasa.gov/s3fs-public/imported/ESDS-RFC-007v1.pdf.

OGC, 2019: OGC GeoTIFF standard. Open Geospatial Consortium (OGC), https://www.ogc.org/standard/geotiff/.

Stocker, E. F., and O. A. Kelley, August 2007: Improving access to precipitation data for GIS users: Designing for ease of use. *IASTED International Conference on Environmental Modeling and Simulation* (EMS 2007), paper 579-012, Honolulu, Hawaii, https://www.iasted.org/conferences/pastinfo-579.html.

Tan, J., and G. J. Huffman, 2024: IMERG V07 Daily and Grand Climatology Products, https://gpm.nasa.gov/sites/default/files/data/climatologies/2024/IMERG.V07.Climatology.Release.240502.pdf.