Precipitation Processing System (PPS)

File Naming Convention for Precipitation Products For the Global Precipitation Measurement (GPM) Mission PPS_610.2_P550, V1.4.4

Prepared By:

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
GODDARD SPACE FLIGHT CENTER
Code 610.2/PPS
Greenbelt, Maryland 20771

October 14, 2015

National Aeronautics and Space Administration

Goddard Space Flight Center
Greenbelt, Maryland

CM FOREWORD

This document is a Precipitation Processing System (PPS) Configuration Management (CM)-controlled document. Changes to this document require prior approval of the applicable Configuration Control Board (CCB) Chairperson or designee. Proposed changes shall be submitted to the PPS CM Office (CMO), along with supportive material justifying the proposed change. Changes to this document will be made by complete revision.

Questions or comments concerning this document should be addressed to:

PPS Configuration Management Office
Mail Stop 610.2
Goddard Space Flight Center
Greenbelt, Maryland 20771

SIGNATURE PAGE

Prepared by:

Erich F. Stocker Date
PPS Project Manager
GPM Deputy Project Scientist for Data
Code 610.2

Reviewed by:

John Stout Date
PPS Science Analyst
Code 610.2/George Mason University

Lawrence Woltz Date
PPS Systems Engineer
Code 610.2/Wyle Information Systems

John Kwiatkowski Date
PPS Science Analyst
Code 610.2/George Mason University

Yimin Ji Date
PPS Science Analyst
Code 610.2/George Mason University

Patty McCaughey Date
PPS Database Designer/Lead Programmer
Code 610.2/George Mason University

Approved by:

Erich F. Stocker Date
PPS Project Manager
GPM Deputy Project Scientist for Data
Code 610.2

CHANGE RECORD PAGE

Mod. #	Mod. Date	Description	Section Affected
1	2/16/2011	Version 1.0 modified to become Version 1.1	Sections 2.1.7, 2.2.1, 2.2.2
2	8/24/2011	Version 1.1 modified to become Version 1.2	Section 2.0; added new Appendix A
3	6/18/2012	Version 1.2 modified to become Version 1.3	Entire document
4	11/8/2012	Version 1.3 modified to become Version 1.4	Section 2.0
5	12/3/2012	Version 1.4 modified to become Version 1.4.2 (Version 1.4.1 was a draft version containing changes to Table A-1 only)	Appendix A, Tables A-1 and A-2
6	7/31/2013	Version 1.4 modified to become Version 1.4.3	Section 2.2.3; added new Sections 2.2.4 and 2.2.5
7	10/14/2015	Added a subfield example, latent heating examples, and corrections; now Version 1.4.4	Table 2; Sections 2.2.2, 2.2.3, 2.2.4

LIST OF TBDs/TBRs

Item No.	Location	Summary	Ind./Org.	Due Date

TABLE OF CONTENTS

LIST OF TABLES

1.0 INTRODUCTION

1.1 BACKGROUND

This document describes the file naming conventions that will be used to name data products produced by the Precipitation Processing System (PPS) for the Global Precipitation Measurement (GPM) Mission. The GPM Mission is an international collaboration that includes the National Aeronautics and Space Administration (NASA), the Japan Aerospace Exploration Agency (JAXA), and additional international partners. The objective is to advance precipitation measurements from space by deploying the space-borne GPM Observatory as a reference standard to unify a constellation of microwave sensors to provide next-generation global precipitation measurements for scientific research and societal applications. The GPM Observatory data will be combined on the ground with data from other constellation satellites participating as part of the program.

The primary GPM scientific objectives are 1) to establish new standards for precipitation measurement capabilities from space; 2) to advance understanding of the global water/energy cycle variability and freshwater availability; 3) to improve weather forecasting skills through frequent and accurate measurements of instantaneous rain rates and rain-affected radiances; 4) to advance climate modeling and prediction capabilities through improved knowledge of precipitation microphysics, atmospheric latent heat release, and surface water fluxes; and 5) to advance prediction capabilities for flood, drought, freshwater resources, and other hydrological applications through improved sampling and coverage of high-resolution precipitation estimates.

The GPM Observatory consists of the NASA-provided GPM Spacecraft, the NASA-provided GPM Microwave Imager (GMI) Instrument, and the JAXA-provided Dual-frequency Precipitation Radar (DPR) Instrument consisting of the Ku-band radar (KuPR) and Ka-band radar (KaPR). The development of the DPR Instrument is based on the joint research between JAXA and the National Institute of Information and Communications Technology (NICT) of Japan.

The GPM Observatory is planned for launch in 2014 using an H-IIA launch vehicle for direct injection into a circular orbit at nominal initial altitude of 407 km with an inclination of 65 degrees. Launch will be from the Yoshinobu Launch Complex of the Tanegashima Space Center located on Tanegashima Island of Japan. The GPM Observatory has a planned launch readiness date of February 2014.

After the successful separation from the launch vehicle and completion of a checkout period, the GPM Observatory is planned to be operated for 3 years at a nominal altitude of 407 km. The GPM Observatory will be operated by NASA using its Tracking and Data Relay Satellite System (TDRSS) for command, control, and data acquisition. Science data received at the NASA Goddard Space Flight Center (GSFC) will be processed by the NASA Precipitation Processing System and JAXA GPM/DPR Mission Operation System.

1.2 APPLICABILITY AND SCOPE

The file naming conventions described in this document are applicable to all regular files intended for distribution to the public and routinely produced by the PPS. These file naming conventions are also intended to apply to files produced or reprocessed from the Tropical Rainfall Measuring Mission (TRMM) satellite during the period of GPM operations.

These file naming conventions are not required to be used by any other GPM partner. Each partner may adopt the file naming convention most appropriate for their needs and systems. PPS does not rename files provided by partners.

1.3 NAMING PHILOSOPHY

The primary objective was to develop a file naming convention that provides sufficient identification information about the products so users are able to catalog their retrieved products effectively. Secondly, the convention is intended to establish "smart" file names that can be parsed easily and incorporated into user systems even without reading the product metadata. Lastly, the convention approach accounts for the fact that in GPM there will be several satellites, some with the same instruments.

For ease of parsing and to maintain the order of directory listings, any field with digits is always normalized to the maximum number of digits that are expected. File name fields are separated by a period '.'. Regardless of the product level, parsing on the '.' will always return the same number of fields. Some file name fields have subfields, and those are separated by a hyphen '-'. While the file naming convention says digits, from the PPS perspective all file name fields are simply characters. However, users can be assured that fields documented as digits will contain only digits.

1.4 DOCUMENT ORGANIZATION

Section 2 contains the detailed file information. It begins with the general formatting approach, and then provides the individual data product level approach afterwards. While this creates replication within this document, it benefits users by making each of the sections independent so that if, for example, users are interested in understanding the file name composition of a Level 3 instrument accumulation product, they do not need to read any other section.

2.0 GPM FILE NAMING APPROACH

2.1 GENERAL FORMAT

This section describes the general format that has been adopted for naming files for GPM. In this section, the field names are generic in character. In subsequent sections, the generic name is replaced by a name appropriate to the level. The general format is as follows:

dataType.satellite.instrument.algorithmName.startDate-SstartTime-EendTime.
sequenceIndicator.VdataVersion.extension

2.1.1 Data Type

This field may be composed of two subfields. The first is the data level indicator, which is a varying alphanumeric value that contains a number to indicate the level of processing and a letter to indicate some sublevel. Table 1 presents the data level values for GPM at launch.

Table 1. Type Subfield

Type	Description
1A	Instrument count, geolocated, at instantaneous field of view (IFOV).
1B	Geolocated, calibrated T_b or radar power at IFOV.
1C	Intercalibrated brightness temperatures T_c at IFOV.
2A	Geolocated geophysical parameters at IFOV from a single instrument.
2B	Geolocated geophysical parameters at IFOV from multiple instruments.
3A	Space/time averaged geophysical parameters from a single instrument.
3B	Space/time averaged geophysical parameters from multiple instruments.
4	Combined satellite, ground and/or model data.

The second subfield for data type is optional and is an indication of accumulation. This is separated from the data level by a hyphen '-'. Table 2 lists the indicators currently supported.

Table 2. Indication of Accumulation Subfield Examples

Name	Description
HR	The product accumulates data for 1 hour.
HHR	The product accumulates data every half hour
DAY	The product accumulates data for a single day.
PENT	The product accumulates data for a 5-day period.
7DAY	The product accumulates data for a 7-day period.
MO	The product accumulates data for a designated month.
ORBIT	The product accumulates data for each orbit.

2.1.2 Satellite

This is a variable-size alphanumeric field that contains an identifier for one of the partner-provided satellites, or satellites accepted by the Science Team. As an example, TRMM would be a satellite identifier. No fixed set of satellite identifiers have yet been established. Those currently included will likely change before GPM launch and are included for examples only.

2.1.3 Instrument

This is a variable-size alphanumeric field that contains a recognized identifier for a particular instrument (e.g., GMI, DPR, Ku, Ka, Special Sensor Microwave Imager/Sounder [SSMIS], etc.). Some instruments fly on multiple satellites, so these are not unique identifiers. However, the satellite-instrument combination will always uniquely identify a data source. No fixed set of instrument identifiers have yet been established. Those currently included will likely change and are included for examples only.

2.1.4 Algorithm Name

This is a variable-size alphanumeric short name or acronym by which the developer wishes to identify the particular processing algorithm (e.g., Goddard profiling algorithm 2008 [GPROF2008], TRMM Multisatellite Precipitation Analysis [TMPA], etc.). This designator is provided by the Algorithm Development Team Leader. At the author's discretion and PPS's recommendation, some version indicator is to be included as part of the algorithm name. No fixed set of algorithm names are established. The algorithm version name is established and maintained by the algorithm team.

2.1.5 Start Date/Time

All files in GPM will be named using the start date/time of the temporal period of the data contained in the product. The field has two subfields separated by a hyphen.

2.1.5.1 Start Date

This field has a fixed format and follows the date format approved by the International Organization for Standardization (ISO). It contains a four-digit year, a two-digit month, and a two-digit day (YYYYMMDD). These parts are not separated.

2.1.5.2 Start Time

This field has a fixed format and follows the time format approved by ISO. It begins with a capital "S." It contains a two-digit hour, a two-digit minute, and a two-digit second (HHMMSS). These parts are not separated and are always present. Hours are presented in a 24-hour time format, with '00' indicating midnight. All times in GPM will be in Coordinated Universal Time (UTC).

2.1.5.3 End Time

This field has a fixed format and follows the time format approved by ISO. It begins with a capital "E." It contains a two-digit hour, a two-digit minute, and a two-digit second (HHMMSS). These parts are not separated and are always present. Hours are presented in a 24-hour time format, with '00' indicating midnight. All times in GPM will be in Coordinated Universal Time (UTC).

2.1.6 Sequence Indicator

This is an all-digit field that has a different maximum number of digits depending upon how it is used. Each product has a specification of the maximum number of digits. For example, if it is used for orbit as in L1 or L2 data, it would have six digits. The size is always fixed to the maximum number of digits specified for a product by including leading zeros as necessary.

2.1.7 Data Version

This is a fixed-size alphanumeric field that PPS assigns to indicate a particular processing state. This file name field always begins with a capital 'V' to indicate version. However, there is some logic built into the data version name.

The data version will always have two digits and a capital letter. The GPM 'at-launch' data version will be 00A. Each reprocessing during this checkout period will change the letter (i.e., 00B, 00C, etc.). Upon approval of the products by the Joint Precipitation Measurement Missions (PMM) Science Team (JPST), all of the products will be reprocessed to the designated start of mission science and will be labeled '01A.'

A change in the digits indicates that all of the mission data have been reprocessed to that level because of major changes in the algorithms. The JPST approves reprocessing, so a change in the digits always indicates an official reprocessing.

In some cases, bug fixes are made to algorithms but the JPST does not believe that a total reprocessing is appropriate or needed. In that event, the letter is incremented to indicate that a bug fix has been installed and that products are being forward-generated with that fix.

2.1.8 Extension

This is a variable-length alphanumeric field that indicates the file storage format. Table 3 provides the extension types that are currently supported.

Table 3. File Extensions

Extension	Description
HDF4	The file is in HDF4 format.
HDF5	The file is in HDF5v1.8 format.
NCDF	The file is in netCDFv4 format.
BIN	The file is in binary format.
TXT	The file is in American Standard Code for Information Exchange (ASCII) text format.

2.2 PRODUCT FORMAT, LEVEL 1 TO LEVEL 3

2.2.1 Level 1 (L1) Products

Level 1 products are orbital in nature. The generic approach when applied to L1 uses the orbit as the sequence number. The start date/time is the date/time of the orbit start of the file. Obviously, some granules start in one day and end in the next day; the start time provides sufficient data to determine this from the file name alone.

Examples are as follows:

1A.GPM.GMI.L1AALG.20131101-S235152-E012400.000352.V01A.HDF5
1B.GPM.GMI.L1BALG.20131101-S235152-E012400.000352.V01A.HDF5
1C.F16.SSMIS.XCAL2007.20100101-S005934-E020340.032110.V01A.HDF5

2.2.2 Level 2 (L2) Products

Level 2 products are orbital in nature. The generic approach when applied to L2 uses orbit as the sequence number. The start date/time is the date/time of the orbit start of the file. As with L1, orbits may start on one day and complete during the next day.

Examples are as follows:

2A.GPM.GMI.GPROF2008.20131101-S235152-E012400.000352.V01A.HDF5
2B.GPM.DPRGMI.OLMAS.20131101-S235152-E012400.000352.V01A.HDF5
2B.GPM.DPRGMI.2HCSHv2-1.20131101-S235152-E012400.000352.V01A.HDF5

2.2.3 Level 3 (L3) Products From Single Instruments

Level 3 products are date/time gridded. For purposes of accumulations, the combined is considered a single instrument accumulation. The generic field sequence number is the two-digit month for the monthly products, the day of the year for the daily products, and the orbit number for the orbital products.

Examples are as follows:

3A-MO.GPM.GMI.GRIDG.20141101-S000000-E235959.11.V01A.HDF5
3B-DAY.GPM.DPRGMI.CMBG.20140201-S000000-E235959.032.V01A.HDF5

There are now separate ascending (ASC) and descending (DES) products for L3 daily. ASC and DES are minor attributes of the data level, just as DAY is a minor attribute of the 3A data level.

Examples for DPR are as follows:

3A-DAY-ASC.GPM.DPR.V1-20140922.20140201-S000000-E235959.032.V01A.HDF5
3A-DAY-DES. GPM.DPR.V1-20140922.20140201-S000000-E235959.032.V01A.HDF5

Examples for latent heating are as follows:

3B-ORBIT.GPM.DPRGMI.3GCSHv2-1.20131101-S235152-E012400.000352.V01A.HDF5
3B-MO.GPM.DPRGMI.3HCSHv2-1.20141101-S000000-E2359511.V01A.HDF5

2.2.4 Browse Products

For browse products, the file names will be as follows:

dataType-BR.satellite.instrument.algorithmName.startDate-SstartTime-EendTime.
sequenceIndicator.VdataVersion.extension

An example for a GMI browse product is as follows:

2A-BR.GPM.GMI.GPROF2008.20131101-S235152-E012400.000352.V01A.HDF5

2.2.5 Coincidence Products

Coincidence subset (CS) products are subsets of the full product, containing only the data within the coincident time and location. The coincidentSite is the designation of the ground site or satellite of the coincident event.

For coincidence products, the file names will be as follows:

dataType-CS-coincidentSite.satellite.instrument.algorithmName.startDate-SstartTime-EendTime.
sequenceIndicator.VdataVersion.extension

An example for a GMI coincidence subset product for the MELB ground site is as follows:

2A-CS-MELB.GPM.GMI.GPROF2008.20131101-S235152-E012400.000352.V01A.HDF5

2.2.6 Radiometer Level 3 Merged Products

This is a time/space gridded product that has a format similar to the other L3 products. However, the sequence number is the start of the time accumulation (currently every half hour). Half-hour accumulation times begin at 0000 for the first half hour and add 30 at each half-hour period. The merged product may have a third subfield in the Data Type field that indicates the latency of the product. The final run of the product never has the third subfield, but an earlier run more quickly produced and without the standard gauge adjustment might have an "L" (for example, 3B-HHR-L).

An example is as follows:

3B-HHR.MS.MRG.iMERG.20101130-S013000-E015959.0090.V01A.HDF5

3.0 ABBREVIATIONS AND ACRONYMS

ASC	Ascending (Product)
ASCII	American Standard Code for Information Exchange
BIN	Binary
CCB	Configuration Control Board
CM	Configuration Management
CMO	Configuration Management Office
CS	Coincidence Subset
DES	Descending (Product)
DPR	Dual-Frequency Precipitation Radar
GMI	GPM Microwave Imager
GPM	Global Precipitation Measurement
GPROF	Goddard Profiling Algorithm
GSFC	Goddard Space Flight Center
HDF	Hierarchical Data Format
IFOV	Instantaneous Field of View
ISO	International Standards Organization
JAXA	Japan Aerospace Exploration Agency
JPST	Joint PMM Science Team
KaPR	Ka-Band Precipitation Radar
KuPR	Ku-Band Precipitation Radar
L1A – L1C	Level 1A – Level 1C (Algorithms)
NASA	National Aeronautics and Space Administration
NCDF	netCDFv4 (File Format)
NICT	National Institute of Information and Communications Technology
PMM	Precipitation Measurement Missions
PPS	Precipitation Processing System
SSMIS	Special Sensor Microwave Imager/Sounder
T_b	Brightness Temperature
TBD/TBR	To Be Determined/To Be Resolved
T_c	Intercalibrated Brightness Temperature
TDRSS	Tracking and Data Relay Satellite System
TMPA	TRMM Multisatellite Precipitation Analysis
TRMM	Tropical Rainfall Measuring Mission
TXT	Text
UTC	Coordinated Universal Time

Note: For a comprehensive list of acronyms, please see the TRMM GPM Acronym Dictionary, PPS_610.2_P580.

APPENDIX A. REALTIME FILE NAMES

Realtime file names will follow the standard file naming conventions as much as possible. This appendix describes the conventions to be used in naming realtime files.

A-1. SWATH PRODUCT CONVENTIONS

Table A-1. Swath Product Realtime File Naming Conventions

Component Name	Component Type	Separator	Component Number
Data Type	Major	.	1
Satellite	Major	.	2
Instrument	Major	.	3
Algorithm Name	Major	.	4
Start Date	Major	.	5
S	Literal/Constant	None	5
Start Time	Subcomponent	-	Part of 5
E	Literal/Constant	None	Part of 5
End Time	Subcomponent	-	Part of 5
V	Literal/Constant	.	6
Data Product Version	Major	None	Part of 6
RT-H5	Literal/Constant Major	.	7

Major components are separated by the '.' character.
Some fields have subcomponents; the parts of subcomponents within the major components are separated by the '-' character.

The data type is as described in the main document. However, except for the merged half-hourly product, no Level 3 products are produced in realtime.

In realtime there may be times when a start time is NOT known. If this happens, then the start time will be set to 000000. There will always be a start time subgroup. Also, realtime products are not orbital, so for swath products there is no sequence number.

Examples are as follows:

1B.GPM.GMI.ALG1B11v2.20140102-S235624-E012400.V01R.RT-H5
2A.NPP.ATMS.GPROF2010r22.2014013-S002346-E012400.V01R.RT-H5

A-2. GRIDDED PRODUCTS

The near-realtime system will also produce merged half-hourly gridded products. These will be produced more than once for each half hour. As a result, the merged near-realtime data type has a latency subcomponent.

Table A-2. Gridded Product Realtime File Naming Conventions

Component Name	Component Type	Separator	Major Component Count
Data Type	Major	.	1
Data Period	Subcomponent	-	Part of 1
Data Latency	Subcomponent	-	Part of 1
Satellite	Major	.	2
Instrument	Major	.	3
Algorithm Name	Major	.	4
Start Date	Major	.	5
Start Time	Subcomponent	-	Part of 5
Half-Hour Sequence	Major	.	6
V	Literal/Constant	.	7
Data Product Version	Major	None	Part of 7
RT-H5	Literal/Major	.	8

Major components are separated by the '.' character.
Subcomponents are separated by the '-' character.

The data period in near-realtime is currently always HHR for half-hour time gridding. The latency subcomponent may be:

E – Early (6 hours or less) OR L – Late (16 hours or less)

The number of times it is run might be variable for a particular half hour, but there will always be an E and an L. The number of times will be fixed at launch. However, it may be changed during any reprocessing cycle.

The half-hour sequence starts at 0000, and increments by 30 for each half hour of the day.

An example is as follows:

3B-HHR-E.MS.MRG.iMERGEv21.20150203-S000000-E01595.0060.V01R.RT-H5