VERSION 1.0

Algorithm Theoretical Basis Document (ATBD)

NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (I-MERG)

Prepared for:

Global Precipitation Measurement (GPM)
National Aeronautics and Space Administration (NASA)

Prepared by:

George J. Huffman
SSAI and NASA/GSFC
NASA/GSFC Code 613.1
Greenbelt, MD 20771

and David T. Bolvin, Dan Braithwaite, Kuolin Hsu, Robert Joyce, Pingping Xie, Soo-Hyun Yoo

30 November 2010

TABLE OF CONTENTS

Page

1. INTRODUCTION 1
   1.1 Objective 1
   1.2 Revision History 1

2. OBSERVING SYSTEMS 1
   2.1 Core Satellite 1
   2.2 Microwave Constellation 3
   2.3 IR Constellation 4
   2.4 Additional Satellites 4
   2.5 Precipitation Gauges 4

3. ALGORITHM DESCRIPTION 5
   3.1 Algorithm Overview 6
   3.2 Processing Outline 7
   3.3 Input Data 7
      3.3.1 Sensor Products 7
      3.3.2 Ancillary Data 7
   3.4 Microwave Intercalibration 8
   3.5 Microwave-Calibrated IR 9
   3.6 Kalman-Smoother Time Interpolation 9
   3.7 Satellite-Gauge Combination 10
   3.8 Post-Processing 10
   3.9 Error Estimates 11
   3.10 Algorithm Output 11
   3.11 Pre-Planned Product Improvements 12
      3.11.1 Addition/Deletion of Input Data 12
      3.11.2 Upgrades to Input Data 12
      3.11.3 Polar Sensors 12
      3.11.4 Upgrades for RT 12
      3.11.5 Use of Model Estimates 13
   3.12 Options for Processing 13
      3.12.1 Use of Multi-Spectral Geo-Data.....13
      3.12.2 Computing Propagation from Precipitation 14
      3.12.3 Incorporating Cloud Development Information.....14
      3.12.4 Use of Daily Gauges.....14

4. TESTING 14
   4.1 Algorithm Verification in the PPS Real-Time System 15
   4.2 Final Run Algorithm Verification 15
   4.3 Algorithm Validation 15

5. PRACTICAL CONSIDERATIONS 15
5.1 Module Dependencies 15
5.2 Files Used in I-MERG 16
5.3 Built-In Quality Assurance and Diagnostics 16
5.4 Exception Handling 19
5.5 Transitioning from TRMM- to GPM-Based Products 19
5.6 Timing of Reprocessing for Research Product 20

6. ASSUMPTIONS AND LIMITATIONS 20
6.1 Data Delivery 20
6.2 Assumed Sensor Performance 20

7. REFERENCES 20

8. ACRONYMS 22

LIST OF FIGURES

Page

Figure 1. High-level block diagram illustrating the major processing modules and data flows in I-MERG. The blocks are organized by institution to indicate heritage, but the final code package will be an integrated system. The numbers on the blocks are for reference in Section 5. Box 3 is computed at CPC as an integral part of I-MERG. 8

LIST OF TABLES

Page

Table 1. List of likely contributing data sets for I-MERG, broken out by sensor type. Data sets with start dates of Jan 98 extend before that time, but these data are not relevant to I-MERG. Square brackets ([ ]) indicate an estimated date. “M-T” stands for Megha-Tropiques. It is planned that the geosynchronous IR data will be processed into “even-odd” files at NESDIS. All “Precip” is Level 2 (scan/pixel) except for the precipitation gauge analyses. 2

Table 2. Notional requirements for I-MERG. TCI and GCI are TRMM Combined Instrument and GPM Combined Instrument, respectively. 6

Table 3. Notional lists of data fields to be included in each of the output datasets. 11

Table 4. Estimates of file counts and sizes to be used in I-MERG. The letters i, o, t, a, s in “Module Relation” indicate input, output, transfer (between modules or within a module), accumulator, and static, respectively. The numbers in “Module Relation” are keyed to the numbered boxes in Fig. 1. 17

1. INTRODUCTION

1.1 OBJECTIVE

This document describes the algorithm and processing sequence for the Integrated Multi-satellitE Retrievals for GPM (I-MERG). This algorithm is intended to intercalibrate, merge, and interpolate "all" satellite microwave precipitation estimates, together with microwave-calibrated infrared (IR) satellite estimates, precipitation gauge analyses, and potentially other precipitation estimators at fine time and space scales for the TRMM and GPM eras over the entire globe. The system is run several times for each observation time, first giving a quick estimate and successively providing better estimates as more data arrive. The final step uses monthly gauge data to create research-level products. Background information and references are provided to describe the context and the relation to other similar missions. Issues involved in understanding the accuracies obtained from the calculations are discussed. Throughout, a baseline Day-1 product is described, together with options and planned improvements that might be instituted before or after launch depending on maturity and project constraints.

1.2 REVISION HISTORY

Version	*Date*	*Author*	*Description*
1.0	30 November 2010	G. Huffman	Initial version

2. OBSERVING SYSTEMS

Historically and for the foreseeable future, passive microwave (PMW) sensors provide the lion's share of relatively accurate satellite-based precipitation estimates, and these are only available from low-Earth-orbit (leo) platforms. I-MERG is designed to compensate for the limited sampling available from single leo-satellites by using as many leo-satellites as possible, and then filling in gaps with geosynchronous-Earth-orbit (geo) infrared (IR) estimates. Additionally, at high latitudes the usual PMW imager channels are not useful, but satellite-based soundings have shown utility and numerical-model-based estimates have the potential to add value. Finally, precipitation gauge analyses are used to provide crucial regionalization and bias correction to the satellite estimates. None of the satellites, except the GPM Core satellite and one Microwave Constellation satellite are under GPM direction. Therefore, I-MERG uses as many satellites of opportunity as possible. Table 1 gives a listing of the likely data sources, the primary parameters, the date spans of useful operation, and the responsible institution. Note that we plan to provide a continuous record from the beginning of TRMM. In all cases except the geo-IR and the precipitation gauge analyses the input data are accessed as Level 2 (scan-pixel) precipitation.

2.1 CORE SATELLITE

The GPM Core Satellite, like the TRMM satellite before it, serves as both a calibration and an evaluation tool for all the PMW- and IR-based precipitation products integrated in I-MERG since it will provide match-ups with all other PMW-equipped leo-satellites and IR-equipped geo-satellites. Both the TRMM and GPM satellites provide multi-channel, dual-polarization PMW sensors and active scanning radars. Three critical improvements in GPM are that 1) the orbital inclination has been increased from 35° to 65°, affording coverage of important additional climate zones; 2) the radar has been upgraded to two frequencies, adding sensitivity to light

precipitation; and 3) "high-frequency" channels (165.5 and 183.3 GHz) have been added to the PMW imager, which are expected to facilitate sensing of light and solid precipitation.

Table 1. List of likely contributing data sets for I-MERG, broken out by sensor type. Data sets with start dates of Jan 98 extend before that time, but these data are not relevant to I-MERG. Square brackets ([ ]) indicate an estimated date. "M-T" stands for Megha-Tropiques. It is planned that the geosynchronous IR data will be processed into "even-odd" files at NESDIS. All "Precip" is Level 2 (scan/pixel) except for the precipitation gauge analyses.

Merged Radar – Passive Microwave Imager Products			
Product	*Param*	*Period of Record*	*Institution*
GPM DPR-GMI	Precip	[Jul 13] - [Jul 18]	PPS
TRMM PR-TMI	Precip	Jan 98 - [Jun 14]	PPS

Radar			
Sensor	*Param*	*Period of Record*	*Institution*
GPM DPR	Precip	[Jul 13] - [Jul 18]	PPS
TRMM PR	Precip	Jan 98 - [Jun 14]	PPS

Conically-Scanning Passive Microwave Imagers and Imager/Sounders			
Sensor	*Param*	*Period of Record*	*Institution*
Aqua AMSR-E	Precip	Jun 02 - [Jun 12]	Nat. Snow and Ice Data Cent.
DMSP F13 SSMI	Precip	Jan 98 - Nov 09	Colo. State Univ.; NOAA/NCDC
DMSP F14 SSMI	Precip	Jan 98 - Aug 08	Colo. State Univ.; NOAA/NCDC
DMSP F15 SSMI	Precip	Dec 99 - [Dec 11]	Colo. State Univ.; NOAA/NCDC
DMSP F16 SSMIS	Precip	Nov 03 - [Nov 15]	Colo. State Univ.; NOAA/NCDC
DMSP F17 SSMIS	Precip	Nov 06 - [Nov 18]	Colo. State Univ.; NOAA/NCDC
DMSP F18 SSMIS	Precip	Nov 09 - [Nov 21]	Colo. State Univ.; NOAA/NCDC
DMSP F19 SSMIS	Precip	[Nov 12] - [Nov 24]	Colo. State Univ.; NOAA/NCDC
DWSS [MI]	Precip	[Nov 12] - [Nov 24]	Colo. State Univ.; NOAA/NCDC
GCOMW1 AMSR2	Precip	[Jan 12] - [Jan 24]	Nat. Snow and Ice Data Cent.
GCOMW2 AMSR2	Precip	[Feb 16] - [Jan 28]	Nat. Snow and Ice Data Cent.
GPM GMI	Precip	[Jul 13] - [Jul 18]	PPS
GPM Constellation GMI	Precip	[Jul 14] - [Jul 19]	PPS
M-T MADRAS	Precip	[Jul 11] - [Jul 14]	CNES/ISRO
TRMM TMI	Precip	Jan 98 - [Jun 14]	PPS

Table 1, continued.

Cross-Track-Scanning Passive Microwave Sounders			
Sensor	*Param*	*Period of Record*	*Institution*
JPSS-1 ATMS	Precip	[Jun 16] - [Jun 21]	Colo. State Univ.; NOAA/NCDC
METOP-2/A MHS	Precip	Jun 07 - [Jun 17]	Colo. State Univ.; NOAA/NCDC
METOP-2/B MHS	Precip	[Apr 12] - [Jun 22]	Colo. State Univ.; NOAA/NCDC
METOP-2/C MHS	Precip	[Apr 16] - [Jun 26]	Colo. State Univ.; NOAA/NCDC
NOAA-15 AMSU	Precip	Jan 00 - Sep 10	Colo. State Univ.; NOAA/NCDC
NOAA-16 AMSU	Precip	Oct 00 - [Dec 12]	Colo. State Univ.; NOAA/NCDC
NOAA-17 AMSU	Precip	Jun 02 - Dec 09	Colo. State Univ.; NOAA/NCDC
NOAA-18 MHS	Precip	Jun 05 - [Jun 11]	Colo. State Univ.; NOAA/NCDC
NOAA-19 MHS	Precip	Apr 09 - [Apr 19]	Colo. State Univ.; NOAA/NCDC
NPP ATMS	Precip	[Sep 11] - [Sep 16]	NOAA

Geosynchronous Infrared Imagers			
Satellite	*Sub-sat. Lon.*	*Param*	*Agency*
MTSat-1R (old GMS)	140E	Tb	JMA
GOES-E (8, now 12)	75W	Tb	NESDIS
GOES-W (10, now 11)	135W	Tb	NESDIS
Meteosat-8 (old 7)	0E	Tb	EUMETSAT
Meteosat-5	63E	Tb	EUMETSAT

IR/Passive Microwave Sounders			
Sensor	*Param*	*Period of Record*	*Institution*
Aqua AIRS	Precip	Jun 02 - [Jun 12]	NASA/GSFC DISC
NOAA-14 TOVS	Precip	Jan 98 - April 05	Colo. State Univ.; NOAA/NCDC
NPP CrIS	Precip	[Sep 11] - [Sep 16]	NASA/GSFC DISC

Precipitation Gauge Analyses		
Param	*Period of Record*	*Institution*
Gridded Precip	Jan 98 - ongoing	DWD/GPCC

2.2 MICROWAVE CONSTELLATION

The constellation of PMW satellites is largely composed of satellites of opportunity. That is, their orbital characteristics, operations, channel selections, and data policies are outside the control of NASA. The exceptions are the GPM Core satellite and the GPM Constellation satellite, the latter of which is currently being negotiated with an international partner and would be under joint control by NASA. The notional orbital inclination for the GPM Constellation satellite is 40°, which is considered to best augment the polar orbits of the other PMW satellites (except Megha-Tropiques, which is at 20°). The imager channels are considered best for low- and mid-latitude use, while the sounding channels maintain some skill in cold and frozen-surface conditions. Future work with the high-frequency channels might demonstrate that the high-frequency channels on AMSU, GMI, MHS, and SSMIS are also useful at higher latitudes.

2.3 IR CONSTELLATION

Although three different organizations control the geo-IR satellites, long-standing international agreements ensure coordination of orbits and mutual aid in the event of an unexpected satellite failure. The basic requirement is for full-disk images every three hours at the major synoptic times (00, 03, …, 21 UTC). All satellite operators provide a great deal of imagery beyond that, although piecing it together can be somewhat challenging. These data are accessed as brightness temperatures (Tb) in the two-layer "even-odd" format developed at NOAA/CPC for CMORPH. The dataset assembly will be carried out at NOAA/CPC.

2.4 ADDITIONAL SATELLITES

Experience in creating fully global precipitation products for the Global Precipitation Climatology Project (GPCP) demonstrates that precipitation estimated from satellite soundings using the Susskind et al. (1997) algorithm has useful skill at scales as fine as 1° daily (Adler et al. 2003; Huffman et al. 1997). Even assuming that high-frequency channels on AMSU/MHS, GMI, and SSMIS eventually provide high-quality precipitation estimates at high latitudes, we suspect that the sounding-based estimates will still be needed to fill gaps in the collection of high-latitude estimates.

2.5 PRECIPITATION GAUGES

Work in GPCP and TRMM has shown that incorporating precipitation gauge data is important for controlling the bias that typifies satellite precipitation estimates. These projects show that even monthly gauge analyses produce significant improvements, at least for some regions in some seasons. Recent work at CPC shows substantial improvements in the bias correction using daily gauge analysis. The Deutscher Wetterdienst (DWD) Global Precipitation Climatology Centre (GPCC) was established over twenty years ago to provide high-quality precipitation analyses over land based on conventional precipitation gauges. We expect to use two GPCC products, the Full Data Analysis for the majority of the time (currently 1998-2007), and the Monitoring Product from 2008 to the present.

The Monitoring Product is available about two months after the month of observation (see Schneider et al. 2008; Rudolf and Schneider, 2005) and is based on SYNOP and monthly CLIMAT reports received in near-real time via GTS from ~7,000–8,000 stations world-wide reported in the following sources:

- monthly precipitation totals accumulated at GPCC from the SYNOP reports received at DWD, Offenbach,
- monthly precipitation totals accumulated at CPC/NOAA from the SYNOP reports received at NOAA, Washington D.C.,
- monthly precipitation totals from CLIMAT reports received at DWD, Offenbach, Germany,
- monthly precipitation totals from CLIMAT reports received at the UK Met. Office (UKMO), Exeter, UK, and
- monthly precipitation totals from CLIMAT reports received at Japan Meteorological Agency (JMA), Tokyo, Japan.

GPCC's Full Data Analysis is based on a significantly enlarged data base that covers the period 1901 up to 2007 (current V.4 released in 2008) and will be extended to include 2009 in V.5

(expected release at the end of 2010). Compared to the Monitoring Product, the Full Data Analysis includes additional data acquired from global data collections such as GHCN, FAO, CRU; data sets from the National Meteorological and/or Hydrological Services of about 190 countries of the world; and some data from GEWEX-related projects.

For both products, if data are available from more than one source, an "optimum" value – according to the quality of the different data sources – is selected for the precipitation analysis. The selected precipitation data undergo an automatic pre-screening and subsequently the data flagged as questionable are interactively reviewed by an expert. Based on the remaining quality-controlled station data, the anomalies from the background climatology are computed at the stations, interpolated using the SPHEREMAP objective analysis, and added to the background climatology to create the month's analysis.

CPC collects daily precipitation gauge data from ~16,000 stations around the world through the GTS, and from enhanced national networks over the U.S., Mexico, and a few other countries. They analyze global daily precipitation on a quasi real-time basis by interpolating quality-controlled station reports. Note that the "day" in this analysis is defined region by region, not at a uniform UTC time. These data are the basis for the daily satellite-gauge option in Subsection 3.12.4.

3. ALGORITHM DESCRIPTION

Given the available diverse, changing, uncoordinated set of input precipitation estimates, with various periods of record, regions of coverage, and sensor-specific strengths and limitations, we seek to compute the longest, most detailed record of "global" precipitation. To do this, we combine the input estimates into a "best" data set. Although we wish to maintain reasonable homogeneity in the input datasets, for example by using consistently processed archives for each sensor, we are not striving to compute a Climate Data Record dataset.

What are the requirements for the Day-1 Combined Satellite product? No formal survey has been done, but in fact a fairly firm set of notional requirements can be condensed, as summarized in Table 2. The space-time resolution is roughly the microwave spatial scale and the IR temporal scale. The space-time domain represents the PMM goal of covering the whole globe starting with TRMM. Multiple products are specified to satisfy different classes of users, summarized in Subsection 3.2. The best TRMM, and then GPM estimate of precipitation should be taken as the calibration standard, and gauge data are clearly important for anchoring the satellite estimates. Error estimates and embedded auxiliary data fields are key to giving users (and developers) the information needed to assess quality by time and region over the life of the dataset. Finally, as a quasi-operational system, the code must "take a licking and keep on ticking."

Table 2. Notional requirements for I-MERG. TCI and GCI are TRMM Combined Instrument and GPM Combined Instrument, respectively.

Resolution	0.07° [i.e., roughly the resolution of microwave estimates]
Time interval	30 min. [i.e., the geo-satellite interval]
Spatial domain	global, initially filled with data 50°N-50°S
Time domain	1998-present; later explore entire SSM/I era (1987-present)
Product sequence	early sat. (~4 hr), late sat. (~18 hr), final sat.-gauge (~2 months after month) [more data in longer-latency products]
Instantaneous vs. accumulated	instantaneous for half-hr, accumulation for monthly
Sensor precipitation products intercalibrated to TCI before GPM launch, later to GCI	
Global, monthly gauge analyses including retrospective product; explore use in submonthly-to-daily and near-real-time products	
Error estimates; final form still open for definition	
Embedded data fields showing how the estimates were computed	
Operationally feasible, robust to data drop-outs and (constantly) changing constellation	
Output in HDF5 v1.8 (compatible with NetCDF4)	
Archiving and reprocessing for all RT and post-RT products	

3.1 ALGORITHM OVERVIEW

A great deal of expertise in merged precipitation algorithms has been developed in the U.S. during the TRMM era, funded mainly by PMM and by NASA NEWS, NOAA programs (CPO, USWRP), NSF SAHRA, and UNESCO GWADI. In many ways the challenge in the current effort is to identify the strengths of the various groups on the PMM Science Team and merge their efforts to create a unified U.S. algorithm. Specifics identified to date are:

- Perform careful intercalibration of microwave estimates
  - GSFC group has a strong background
- Provide finer time and space scales to get adequate sampling
  - CPC has strong experience with Lagrangian time interpolation using Kalman filters – CMORPH-KF
- Provide microwave-calibrated IR estimates to fill “holes” in the PMW constellation
  - CPC group operationally produces the 4-km Merged and “Even-Odd” IR Tb products
  - U.C.-Irvine group has strong experience in computing IR estimates
- Incorporate gauge data to control bias
  - GSFC group has a strong background
  - CPC group has developed a test system to perform daily bias correction
- Provide error estimates
  - both GSFC and CPC groups bring strengths
- Deliver and support a code package that runs in the PMM Precipitation Processing System (PPS) environment
  - GSFC group has a strong track record

The high-level block diagram that results from this analysis is shown in Fig. 1, which identifies the institutions that provide the heritage code for the various blocks.

3.2 PROCESSING OUTLINE

The block diagram for I-MERG is shown in Fig. 1. In words, the precipitation estimates computed from the various satellite PMW sensors are assembled, mostly received as level 2 precipitation estimates from the relevant producers, but potentially computed in a few cases at PPS as is now done for the TRMM Multi-satellite Precipitation Analysis (TMPA). These estimates are gridded and intercalibrated, then combined into half-hourly fields and provided to both the CPC Morphing-Kalman Filter (CMORPH-KF) Lagrangian time interpolation scheme and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Cloud Classification System (PERSIANN-CCS) re-calibration scheme. In parallel, CPC will assemble the zenith-angle-corrected, intercalibrated "even-odd" geo-IR fields and forward them to PPS for use in the CMORPH-KF Lagrangian time interpolation scheme and the PERSIANN-CCS computation routines. The PERSIANN-CCS estimates are computed (supported by an asynchronous re-calibration cycle) and sent to the CMORPH-KF Lagrangian time interpolation scheme. The CMORPH-KF Lagrangian time interpolation (supported by an asynchronous KF weights updating cycle) uses the PMW and IR estimates to create half-hourly estimates. The system will be run twice in near-real time

- "Early" multi-satellite product ~4 hr after observation time and
- "Late" multi-satellite product ~18 hr after observation time,

and once after the monthly gauge analysis is received

- "Final" satellite-gauge product ~2 months after the observation month.

The baseline is for the real-time estimates to be calibrated with climatological coefficients that vary by month and location, while in the "Final" post-real-time run the multi-satellite estimates are adjusted so that they sum to a monthly satellite-gauge combination following the TMPA. In all cases the output contains multiple fields that provide information on the input data, selected intermediate fields, and estimation quality.

3.3 INPUT DATA

3.3.1 Sensor Products

The sensor products are detailed in Section 2 as part of the discussion of the various sensors. For the most part, the datasets listed in Table 1 from previous and current sensors are already archived at PPS as part of the TMPA work under TRMM, but we anticipate that new versions will become available before GPM launch. As such, we will work closely with PPS to ensure a smooth transition of their archive to the new versions.

3.3.2 Ancillary Products

No additional ancillary products are required on a routine basis for the baseline I-MERG algorithm.

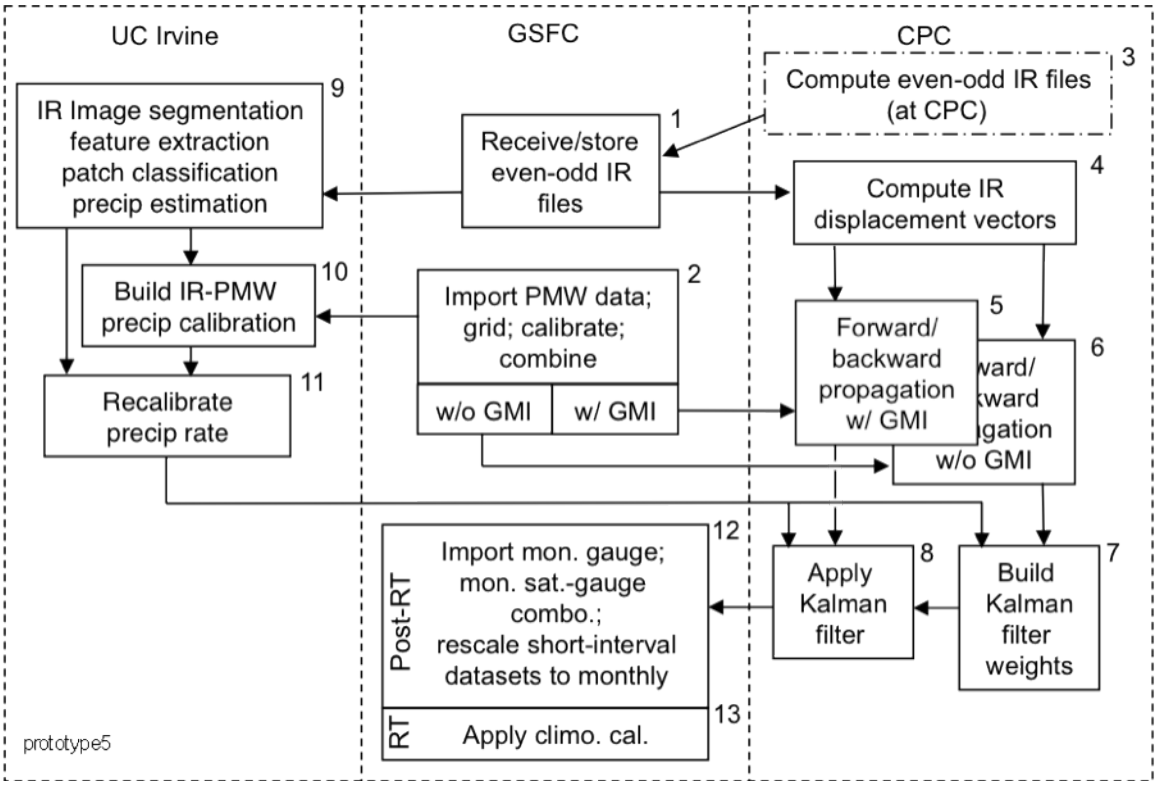

Figure 1. High-level block diagram illustrating the major processing modules and data flows in I-MERG. The blocks are organized by institution to indicate heritage, but the final code package will be an integrated system. The numbers on the blocks are for reference in Section 5. Box 3 is computed at CPC as an integral part of I-MERG.

3.4 MICROWAVE INTERCALIBRATION

As in TRMM, the I-MERG precipitation estimates will be calibrated to the deemed highest-quality GPM single- or combined-sensor estimates following Huffman et al. (2007). Leading candidates for this calibrator are the DPR and the GCI estimates. Candidate calibrators will be analyzed and compared with TRMM-based estimates and ground validation to determine the most suitable choice. During the "ramp-up" period in the GPM mission it is planned to use a TRMM-based calibrator until a GPM-based calibrator is available. The microwave intercalibration technique will be based on quantile-quantile matching, similar to Miller (1972) and Krajewski and Smith (1991). The temporal and spatial scale of the histogram matching for any given sensor will depend upon the unique orbit and individual sensor characteristics. Vastly different orbits, leading to fewer data overlaps, may require a longer calibration period to ensure representative geographic and diurnal sampling. Similarly, radically different sensors may require higher spatial and temporal resolution sampling. Climatological (fixed) calibrations will be used when possible, with dynamically-computed (monthly, say) calibrations utilized when necessary. Comprehensive analysis will determine the exact nature of the calibration for each individual sensor's estimates.

CMORPH-KF, PERSIANN-CCS, and TMPA-RT all use various lengths of trailing calibration in which updating is considered necessary, and this is the intended approach for the real-time I-MERG runs. The post-real-time TMPA uses a calendar-month calibration, but for consistency in

the Day-1 code we plan to routinely update the calibration such that each day is approximately in the middle of its calibration period. [This might be termed a "displaced trailing calibration."]

One improvement in GPM over TRMM is that both DPR and GCI will be available in real time, whereas TCI is not in TRMM. This will allow us to have the same calibrating sensor for all the runs.

3.5 MICROWAVE-CALIBRATED IR

Geo-satellites give frequent sampling, but the resulting IR Tb data are related to cloud top features (temperature and albedo) rather than directly to surface precipitation. This indirect relationship is best captured if the IR Tb-precipitation relationship is improved using texture and patch classification as well as applying routine updates using leo-PMW based precipitation estimates. Here, following the PERSIANN-CCS (Hong et al. 2004), the 60°N-S latitude belt is subsetted into 24 overlapping sub-regions (six in longitude by four in latitude) to allow for regional training and parallel processing. For each sub-region, the full-resolution IR Tb field is segmented into separable cloud patches using a watershed algorithm. Cloud patch features are extracted at three separate temperature levels: 220K, 235K, and 253K, which are chosen to demonstrate the existence of the cloud patches at different altitudes in the atmosphere. An unsupervised clustering analysis (Self-Organizing Feature Map) is used to classify cloud patches into a number of cloud patch groups based on the similarities of patch features. Precipitation is assigned to each classified cloud patch group based on a training set of leo-PMW precipitation samples. These initial precipitation estimates are then adjusted using coefficients based on a trailing backlog of matched leo-PMW precipitation and cloud-patch precipitation. The backlog is sized to ensure sufficient sampling to generate a stable estimate.

3.6 KALMAN-SMOOTHER TIME INTERPOLATION

Under the Kalman Smoother framework as developed in CMORPH-KF and applied here, the precipitation analysis for a grid box is defined in three steps (Joyce et al. 2011). First, PMW estimates of instantaneous rain rates closest to the target analysis time in both the forward and backward directions are propagated from their observation times to the analysis time using the cloud motion vectors computed from the geo-IR images. The "prediction" of the precipitation analysis is then defined by averaging the forward- and backward-propagated PMW estimates with weights inversely proportional to their error variance. The IR-based precipitation estimates are incorporated if the gap between the two PMW observations is longer than 90 minutes. This threshold is due both to the natural timescale of precipitation at these fine scales and to the retrieval errors in the microwave algorithms.

The cloud motion vectors used to propagate the PMW estimates are calculated by computing the pattern correlation between spatially lagged geo-IR Tb arrays from two consecutive images. The spatial displacement with the highest correlation is used to define the cloud motion vector. The cloud motion vectors are defined for each 2.5° lat/lon grid box using IR data over a 5° lat/lon domain centered on the target grid box. Over mid-latitudes, precipitation systems present slightly different movements than the cloud systems that we are tracing with the geo-IR Tb. To account for the differences, the PDFs of the zonal and meridional components of the cloud motion vectors were compared against those of the precipitation systems observed by the Stage II radar over the contiguous U.S. (CONUS). A static correction table was then established for adjusting the geo-IR-based cloud motion vectors in both hemispheres' mid-latitudes to better

represent precipitation motion. Interpolation in time, and then space is used to provide spatially complete propagation fields.

Errors for the individual satellite estimates are calculated by comparison against the Stage II radar estimates of precipitation over CONUS, and against TMI estimates over the rest of the globe. Error functions for the TMI are taken to be the same as those for the AMSR-E, based on an early comparison against the Stage II radar observations over CONUS. Expressed in the form of correlation, the error for the propagated PMW estimates are defined as regionally dependent and seasonally changing functions of sensor type and the length of propagation time. Over land, the error functions are computed for each 10° latitude band using data collected over a 30°-wide latitude band centered on the target band. No zonal differences in the error are considered due to the limited sampling of the data. Over ocean, the error functions are defined for each 20°x20° lat/lon box using data over a 40°x40° lat/lon region centered on the target box. Over both land and ocean, the error functions are calculated for each month using data over a five-month period centered on the target month, to account for the seasonal variations. The comparisons against Stage II were done once, while those against TMI are updated monthly.

3.7 SATELLITE-GAUGE COMBINATION

For the baseline post-real-time I-MERG package, we will follow the TMPA approach for infusing monthly gauge information into the fine-scale precipitation estimates (Huffman et al. 2007). All of the full-resolution multi-satellite estimates in a month are summed to create a monthly multi-satellite-only field. This field is combined with the monthly GPCC precipitation gauge analysis (over land) in a two-step process. First, the multi-satellite estimate is adjusted to the large-scale bias of the gauges, and then the adjusted multi-satellite and gauge fields are combined using weighting by inverse estimated error variance. The satellite-gauge combination is a product in its own right. In addition, the field of ratios between the original multi-satellite and satellite-gauge fields is computed, then each field of multi-satellite precipitation estimates in the month is multiplied by the ratio field to create the final fine-time-resolution I-MERG estimates.

3.8 POST-PROCESSING

The baseline I-MERG real-time products will follow the TMPA procedure in providing both the original multi-satellite estimate and a climatologically calibrated field. The climatological calibration is intended to make the real-time products as consistent as possible with the Final product. One important simplification compared to the TMPA is that we anticipate that both the DPR and GCI will be computed in real time for GPM. This contrasts to the situation in TRMM where the TCI is not computed in real time and we have had to substitute TMI as the RT calibrator. Accordingly, in GPM only a straightforward calibration to the Final product will be computed. If the sub-monthly precipitation gauge combination option is incorporated in the "Late" (18-hr) real-time product, presumably the need for post-processing will have to be re-assessed, but the "Early" (4-hour) real-time product is certain to require the climatological calibration to the Final product.

The final post-processing step that we might choose to take is to use the half-hourly "instantaneous" multi-satellite precipitation fields to create daily accumulation products. If this is done, it would be defined as the sum of all half-hourly estimates available in a UTC day, scaled up to represent full coverage of the UTC day.

3.9 ERROR ESTIMATES

Error estimates are a required item in the output datasets. The baseline fine-scale datasets will have errors estimated as part of the Kalman filter methodology. The baseline monthly post-real-time datasets will have error estimates that are computed as part of the optimal estimation of the satellite-gauge combination. It is possible that more sophisticated error fields can be incorporated as part of I-MERG, for example providing additional information on the time-space error structure. In such a case, the critical problem is to limit the number of time/space-varying parameters that consequently require the insertion of additional parameter fields in each dataset.

3.10 ALGORITHM OUTPUT

All output data files have multiple fields with project-mandated metadata and written in HDF5 v1.8, which is compatible with NetCDF4. The choice of fields varies, depending on which data run is creating the output. Table 2 contains the notional lists of data fields. Recall that PMM will provide an on-the-fly data subsetting by time, region, and parameter, so users are not required to download the entire file. Alternatively, we might explore the option of providing two files for the half-hourly output, one "simple" file with precipitation alone and another "detail" file with the rest of the fields.

As listed in Table 1, the notional requirement is that the output be on a global 0.07° grid. However, there is a strong argument that a fully global grid should be (approximately) equal-area. This issue is under discussion within the project and the actual Day-1 grid will conform to the final decision. Also, the IR data are actually available on a 0.035° grid, and the question has been raised whether that ought to be the notional grid size. At present the baseline is left at 0.07° because there are scientific questions about downscaling microwave footprints to the finer scale and operational questions about data volume.

Table 3. Notional lists of data fields to be included in each of the output datasets.

Half-hourly data file (Early, Late, Final)	
1	Instantaneous precipitation - calibrated
2	Instantaneous precipitation - uncalibrated
3	Precipitation error
4	PMW precipitation
5	PMW source 1 identifier
6	PMW source 1 time
7	PMW source 2 identifier
8	PMW source 2 time
9	IR precipitation
10	IR KF weight
Monthly data file (Final)	
1	Sat-Gauge precipitation
2	Sat-Gauge precipitation error
3	Gauge relative weighting

3.11 PRE-PLANNED PRODUCT IMPROVEMENTS

Throughout the useful life of the Day-1 I-MERG we plan for the code to be reasonably robust to errors, drop-outs, and changes in the make-up of the satellite constellation. However, the team considers it helpful to pre-plan certain enhancements to the code that we are fairly certain will be required at some point.

3.11.1 Addition/Deletion of Input Data

Satellites come and go over time. For the most part, satellite drop-outs, other than of the GPM Core itself, simply result in a smaller amount of input data for the system. Addition of data, on the other hand, is potentially complicated by a range of possible priorities and calibration needs of the new sensor. In I-MERG we follow the work pioneered in the Version 7 TMPA, where extra satellite slots are programmed in, some with the priority of conical scanners and some for sounder priority. When a new sensor comes on-line, it can be assigned to an appropriate-priority slot and start contributing from that point forward, once the calibration coefficients are determined, which can require several months of data. However, including the new sensor's data from before the date/time on which it is instituted in the dataset requires reprocessing (next Subsection).

3.11.2 Upgrades to Input Data

When an existing sensor's data record is reprocessed, or a new sensor is introduced that has an archive not previously used, it is necessary to reprocess the archive of I-MERG data to preserve consistent statistical behavior (to the extent possible) across the entire record. While reprocessing should not be undertaken lightly, given the computing demands on PPS and the disruption to the users, hard practical experience shows that we need to be more aggressive about this issue than has been the case previously for the TMPA. For example, the second version of NESDIS AMSU, introduced in 2004, resulted in an underestimate of light rain. The result in the TMPA was a low bias in fractional coverage and rain amount over ocean. When an upgraded version of the NESDIS AMSU was introduced in early 2007 these biases were greatly reduced, but we allowed the inhomogeneity to persist in the TMPA archive. As a result, users have to be continually reminded that the relatively low values are a known problem, a problem that could have been fixed by reprocessing.

3.11.3 Polar Sensors

The Multi-Satellite team intends to extend I-MERG to the polar regions, consistent with GPM's fully global focus. This requires estimating displacement vectors at higher latitudes from the asynchronous assemblage of LEO-IR satellites. We will then use these vectors to displace the available high-latitude precipitation estimates and apply the backward/forward Kalman filter to compute the output estimates. Available high-latitude estimates include the TOVS and AIRS estimates computed using the Susskind et al. (1997) algorithm, AMSU and MHS estimates computed by Surussavadee and Stalin (2008a,b), and numerical model estimates (Subsection 3.11.5). This development work will require close cooperation with the experts in high-latitude GV.

3.11.4 Upgrades for RT

It is likely that the RT will require modifications to create the most useful output. For example, it might be necessary to tune the latency of the Early and/or Late runs to accommodate the realities of the data reception. Tuning for the Late run will also have to balance the useful time

range of backward-propagated microwave data against the latency of the following microwave overpass. In addition, work to improve the calibration of the RT to the Final post-real-time production dataset might require a reprocessing to implement the improvement. If the daily gauge option is instituted for the "Late" run, we believe we can fit it in to the latency structure of the baseline scenario. That is, if the daily gauge analysis has a latency that is much longer than the Late run satellites require, the daily gauge computation might be able to use the PDF of data up through the previous day.

3.11.5 Use of Model Estimates

Validation work by Ebert et al. (2007) among others, demonstrates that numerical model estimates of precipitation can out-perform observational estimates at daily 0.25°x0.25° scale in the cool season over land. This stands in contrast to the poor performance by model estimates in tropical and subtropical conditions for day-to-day variations, diurnal cycle, and seasonal variation. The Multi-Satellite team's experience in isolating bias in input datasets and the flexible, error-sensitive behavior of the Kalman filter concept seem to suggest that the I-MERG is a natural platform for testing the joint use of observational and model-base precipitation estimates. This is particularly true given the expectation that the team will be exploring extension into polar regions (Subsection 3.11.3). It is absolutely clear that the team intends to maintain a robust observation-only capability throughout GPM to support a variety of applications, not the least being validation of model estimates. However, a parallel joint observation-model product is a worthy contribution to the project and to advancing scientific understanding.

3.12 OPTIONS FOR PROCESSING

Since the clear mandate for the Day-1 algorithm is driven by a very aggressive schedule (ATBD, working code, and final code in November of 2010, 2011, and 2012, respectively), the baseline algorithm described in this Subsection is designed around code that is already running and tested. At the same time, the team has several concepts in research that might become sufficiently mature in this timeframe that one or more of them might be added to the baseline system. Alternatively, these concepts might be the prime targets for upgrading the Day-2 version.

3.12.1 Use of Multi-Spectral Geo-Data

Besides the thermal IR channel discussed above, geo-satellites also provide other channels, usually visible and one or more spread across the IR spectrum. Historically, these channels have not been used due to apparent modest improvements in skill, difficulties in handling the higher data volumes, and limitations to daylight hours (for visible). However, our on-going dependence on geo-satellite data to fill in between PMW overpasses and the increasing number of channels makes it important to reconsider this aspect. Recent studies seem to indicate reasonable increases in skill using modern neural net approaches, particularly when visible data are used (Behrangi et al. 2009). Several important steps must be taken to capitalize on this apparent benefit in using multi-spectral data. First the scientific development must be advanced to operational status. Second, we must work with the data providers to arrange for routine delivery of the data in a useful format, including a complete archive. Third, choices must be made on the selection of channels, recognizing that previous generations of geo-satellites had less capable sensors than those now appearing.

3.12.2 Computing Propagation from Precipitation

The propagation vectors currently used in the CMORPH-KF Lagrangian time interpolation are computed from IR-based cloud motions. As noted in Subsection 3.6, these differ from the motion of precipitation systems, creating a source of error. Accordingly, the team is working on alternative schemes to more accurately estimate motion from precipitation data.

3.12.3 Incorporating Cloud Development Information

Precipitation develops and decays over time periods that are short compared to the typical revisit time of the leo-PMW constellation. Experience shows that the correlation of observed and propagated precipitation fields may drop from 1.0 to ~0.6 within 30-minutes and further fall to ~0.4 or lower after an hour of propagation. Instantaneous geo-IR precipitation estimates are notoriously poor, but nonetheless provide a minimum floor of skill when a gridbox lacks recent propagated leo-PMW estimates. Taking a different approach, capturing the dynamic evolution of geo-IR cloud images may help to identify cloud systems in various stages of development. This approach to addressing the "cloud development problem" is a relatively new area of research and requires further investigation to determine the best strategies for capturing the development process. One possibility is to drive a highly simplified conceptual cloud model with parameters computed from the geo-IR Tb data, as in the Bellerby et al. (2009) Lagrangian Model (LMODEL). Another is to modify the propagated leo-PMW precipitation estimates with time based on parameters computed from the geo-IR Tb data, as in the Behrangi et al. (2011) Rain Estimation using Forward Adjusted-advection of Microwave Estimates (REFAME).

3.12.4 Use of Daily Gauges

The biases discussed previously vary on sub-monthly time scales, of course. To address this problem, we will examine the possibility of refining the bias correction approach described in Subsection 3.7 through the use of daily gauge analysis. Recently, CPC has developed a new technique to correct the bias in high-resolution satellite precipitation estimates through matching the PDF of the satellite estimates against that of the daily gauge analysis (Xie et al. 2010). The PDF bias correction is carried out in two steps using historical, and then real-time data. First, PDF tables are constructed for each 0.25° lat/lon gridbox over the global land and for each calendar day using co-located satellite and gauge data pairs over a spatial domain centered on the target grid box and over a sliding window of 31 days centered on the target calendar day for the 13-year period from 1998 to the present. The spatial domain is expanded until a sufficient number of data pairs are collected. After the correction using historical data, the satellite estimates are further calibrated against the real-time data to remove the year-to-year variations in the bias. To this end, PDF tables are created using co-located data collected over a 30-day period ending at the target date. The least numbers of co-located data pairs used to create PDFs are 500 and 300 for the corrections using historical and real-time data, respectively.

4. TESTING

The CMORPH-KF and PERSIANN-CCS systems will be brought up in the GSFC development environment in GSFC Code 613.1 with the minimum number of changes possible to ensure that the code as originally presented is functional. The TMPA code already satisfies this requirement. Thereafter, the I-MERG code will be built on the development system.

4.1 ALGORITHM VERIFICATION IN THE PPS REAL-TIME SYSTEM

As the baseline I-MERG code is developed it will be validated on the development system. Once the entire package is assembled, it will be transferred to PPS for integration and testing. We expect that the first "production" tests will be carried out on the PPS real-time system in parallel with the then-current TMPA-RT code. This will allow us to make detailed and summary comparisons between the two sets of products on a routine basis. The new I-MERG real-time products will be compared against coincident CMORPH-KF, PERSIANN-CCS, and (post-real-time) TMPA fields. The goal in this stage is to shake out as many bugs and conceptual difficulties as possible, applying corrections to all versions of the I-MERG code.

4.2 FINAL RUN ALGORITHM VERIFICATION

The background of RT system testing should ensure that the production code is relatively mature by the time its production testing begins. The main features to be validated are the use of somewhat different input data sets and the addition of monthly gauge calibration. As before, it is important to compare the results to the other estimates and validation data listed above.

4.3 ALGORITHM VALIDATION

The more formal algorithm validation will examine various aspects of the I-MERG results. At the instantaneous level, comparison to the fine-scale NOAA Next-Generation Multi-Sensor Quantitative Precipitation Estimates (Q2) and to the PMM Kwajalein and Melbourne radar archives are considered key. As part of this effort, we expect to carry out similar comparisons against the gridded Level 2 input data. The performance at larger space-time scales will be assessed using accumulations of these three datasets, as well as the CPC daily gauge analysis, the IPWG validation sites (Australia, CONUS, Japan, South America, Western Europe), the GPCC global monthly gauge analysis, the Pacific atoll data, and the ATLAS II buoy data. For higher-latitude validation, the GPCC data can be used to validate the satellite-only products. The team already has access to Finnish Meteorological Institute precipitation gauge data, and will work with the P. Groisman project to access precipitation gauge data and metadata over much of the high-latitude Northern Hemisphere. At a minimum, metrics should include bias, root-mean-square error, mean absolute error, correlation, and skill scores. Decompositions into hit error, miss error, etc. following Tian et al. (2009) should be considered as well. We will invite the validation teams, such as the W. Krajewski project, to examine the test datasets with the detailed validation approaches that they are developing. Finally, we will invite selected users, particularly hydrologists, to use the test datasets and report their experiences to help determine what I-MERG's level of skill is for their applications. Ideally, the PMM Ground Validation (GV) team will establish an on-going monitoring activity to detect dataset quality problems.

5. PRACTICAL CONSIDERATIONS

5.1 MODULE DEPENDENCIES

The baseline structure of the I-MERG is shown in Fig. 1. We have not enforced consistency on the various boxes in the sense that some boxes might be programmed as multiple modules, while others will be computed in a single module. As summarized in Subsection 5.2, the data flow between modules, and between executions of the same module, is carried out using files, which

typically have fixed names. Input and output datasets necessarily have names that reflect the time sequencing of the data that they contain.

The satellite-satellite calibrations, which include the PMW intercalibrations to a GPM standard (block 2), IR-PMW precip calibration for the IR estimates (block 10), and the Kalman filter weights (block 7), are conceptually asynchronous with the actual half-hourly precipitation dataset processing. It is a matter for discussion with PPS as to whether the calibrations will be run sequentially or in parallel, but the system is designed to be very forgiving of occasional missed calibration match-ups – without significant loss of skill it can run with the then-current calibration files, as long as the dropouts do not become too severe. The heritage TMPA system has computed the PMW intercalibration on a calendar month basis, while the PERSIANN-CCS and CMORPH-KF have run the IR-PMW and KF weights, respectively, on trailing accumulations of match-ups. For I-MERG we plan to run all the calibrations on trailing accumulations of match-ups. In the case of the post-real-time Final analysis we must wait for the GPCC precipitation gauge analysis, so we plan to accumulate the match-ups with a sufficient delay after real time that the Final calibrations are approximately centered.

The only important difference between real- and post-real-time runs comes in the last calibration, which will be computed for the real-time as climatological adjustments to the post-real-time product, and for the post-real-time as calendar-month adjustments to and combination with monthly gauge analyses. It is a matter for discussion with PPS as to how far the "Final" processing will be carried through the system pending the delivery of the monthly precipitation gauge analysis about two months after the end of the month.

As noted above, we plan to compute three runs of the algorithm, namely the "Early", "Late", and "Final" runs at about 4 hr, 18 hr, and 2 months after observation time. The simplest approach is to maintain three entirely separate sets of files and to compute everything in each run. It is a matter for discussion with PPS whether the savings in processing that is achieved by sharing some files between runs, particularly the griddings of the input Level 2 precipitation datasets and the accumulations for the calibrations, warrant the additional computational complexity of creating dependencies between runs. The most likely target would be to share such files between the two real-time runs (Early and Late).

5.2 FILES USED IN I-MERG

Input, output, inter-module data transfer, and inter-run/static data storage is accomplished through files in I-MERG. Table 3 displays our best estimate of what the file sizes and count might be, but as the code is developed these numbers will certainly change. Also, it is important to note that the granularity of the input data implies that two of each of the input types will have to be used in each half-hour with fair regularity. On the other hand, the precipitation gauge analysis only provides one file in a month, which is also true for the monthly Merged Satellite/Gauge product. Several of the options and planned upgrades will require the use, transport, and accumulation of data in additional files.

5.3 BUILT-IN QUALITY ASSURANCE AND DIAGNOSTICS

To the extent possible, every effort will be made to incorporate quality assurance checks in the I-MERG system. This includes quality checks of all input data, and selected intermediate and output data based on metrics developed for TRMM. The goal of these metrics is to capture

Table 4. Estimates of file counts and sizes to be used in I-MERG. The letters i, o, t, a, s in "Module Relation" indicate input, output, transfer (between modules or within a module), accumulator, and static, respectively. The numbers in "Module Relation" are keyed to the numbered boxes in Fig. 1.

INPUT					
ID	*# Granules*	*Granularity*	*Granule Size (MB)*	*Total Size (MB)*	*Module Relation*
TMI	1	One orbit	56	56	i2
TCI	1	One orbit	328	328	i2
SSMI x 3	3	One orbit	11	33	i2
SSMIS x 4	4	One orbit	11	44	i2
AMSR x 2	2	One orbit	11	22	i2
AMSU x 3	3	One orbit	5	15	i2
MHS x 6	6	One orbit	5	30	i2
Megha-Tropiques	1	One orbit	10	10	i2
GMI	1	One orbit	56	56	i2
GCI	1	One orbit	328	328	i2
ATMS x 2	2	One orbit	11	22	i2
AIRS	1	One orbit	11	11	i2
TOVS	1	One orbit	11	11	i2
CrIS	1	One orbit	11	11	i2,i4
IR	2	One hour	65	130	i1,t1-4,t1-9
GPCC	1	One month	8	8	i12

INTERMEDIATE/STATIC				
ID	*# Files*	*Size (MB)*	*Total Size (MB)*	*Module Relation*
GCI-other cal	21	1	21	s2
Surface Type	1	52	52	s2
GCI-TMI cal	1	230	230	a2
GCI-GMI cal	1	230	230	a2
GCI-TCI cal	1	230	230	a2
Cloud classification	24	0.04	1	s9
IR/rainrate	24	0.4	10	s9

Table 4, cont.

TRANSFER					
ID	*# Granules*	*Granularity*	*Granule Size (MB)*	*Total Size (MB)*	*Module Relation*
Gridded HQ	14*	One orbit	78	1092	t2
GCI-TMI cal	1	One orbit	260	260	t2
GCI-GMI cal	1	One orbit	260	260	t2
GCI-TCI cal	1	One orbit	260	260	t2
IR stats	1	One hour	12	12	i4
Cloud Motion Vectors	4	One hour	0.2	0.8	t4-5,t4-6
Kalman Filter weights	1	One month	435	435	t7-8
PMW (with GMI and without GMI)	2	30 minutes	182	364	t2-5,t2-6,t2-10
PMW forward and backward prop (no GMI)	2	30 minutes	182	364	t6-7
PMW forward and backward prop (GMI)	2	30 minutes	182	364	t5-8
Intermediate IR	1	30 minutes	54	54	t9
Intermediate HQ	1	30 minutes	41	41	t10
IR sub areas	1	30 minutes	3	3	t9
HQ sub areas	1	30 minutes	2	2	t10
CCS precip sub areas (unadjusted)	1	30 minutes	5	5	t9-11
CCS precip sub areas (adjusted)	1	30 minutes	2	2	t11
Sub area adjustment coefficients	1	30 minutes	2	2	t10-11
CCS global precip (unadjusted)	1	30 minutes	54	54	t11
CCS global precip (adjusted)	1	30 minutes	54	54	t11-7,t11-8
Merged PMW/IR (uncal)	1	30 minutes	383	383	t8-12,t8-13

* Although 28 sources of "high-quality" satellite data are listed under "INPUT", it is assumed that no more than 14 such satellites will be available at any given time.

OUTPUT					
ID	*# Granules*	*Granularity*	*Granule Size (MB)*	*Total Size (MB)*	*Module Relation*
Merged MW/IR (cal)	1	30 minutes	383	383	o12,o13
Merged MW/IR	1	One month	65	65	o12

discrepancies before they propagate into the downstream processing. PPS toolkit warning and error messages will be the primary mechanism used to flag potential problems. Optional diagnostic information will be available to the operator when requested. It is possible that a separate, post-processing algorithm will be used to extend the quality assurance procedure to the final product. The goal of this post-processing algorithm is to capture more subtle issues than observable during production.

5.4 EXCEPTION HANDLING

Like the TRMM predecessor, the I-MERG system will be quite robust in handling exceptions, including input file existence and integrity, command-line consistency, and routine data checks. The algorithm developers will work with PPS to specify a minimum set of required inputs for each algorithm, and the conditions under which the I-MERG system should report issues using the PPS toolkit. It is the responsibility of the Multi-Satellite Team to create the toolkit error messages. When issues are flagged by the toolkit, additional diagnostic output is integrated into the code by the developers to assist in isolating the problem when requested by the operator. Error reporting will be used when exceptions are significant enough to halt execution. Warning reporting will be used when exceptions should be noted but processing can continue. In both cases, PPS will contact the algorithm developers to determine the severity of the exception and how best to address it.

5.5 TRANSITIONING FROM TRMM- TO GPM-BASED PRODUCTS

As noted in Section 4, the vision is that the I-MERG RT products will be brought up first, debugged, and run in parallel with the current TMPA-RT system. We favor providing routinely computed I-MERG RT products to successively more users as soon as practical in order to 1) gain critical feedback from key user groups as early as possible, and 2) give users the maximum time possible to make the transition to the new processing paradigm. This might occur in early 2012, clearly still using TRMM as the calibrator to generate “GPM proxy” products. It is open for discussion whether we can run and disseminate the legacy TMPA-RT and I-MERG RT products in full parallel, or whether the project will consider one to be an off-line product. In addition, the issue of when to first create a multi-year archive of I-MERG RT proxy products must be addressed.

Meanwhile, as noted in Section 4, we envision the implementation of the “final” production system lagging the RT to allow lessons learned in the new RT to be infused before the production code is frozen. The actual Day-1 production system will still depend on TRMM-based calibrations at launch. It is open for discussion whether it makes sense to create a multi-year archive of I-MERG Final run datasets for all or just the latter part of the TRMM era.

After launch, the Day-1 Level 2 GPM algorithms will be finalized after “enough” GPM data are accumulated. After that, the I-MERG calibrations using GPM data can be computed and instituted. This will allow PPS to compute the archive for the first GPM-calibrated I-MERG products, both RT and Final.

This nominal scenario assumes that TRMM persists and provides data until the acceptance of GPM. Should TRMM either exhaust its fuel or suffer a major failure in either the PR or TMI, the legacy TMPA-RT, and by extension the I-MERG RT proxy products would be relatively unaffected: the TRMM-based RT calibrations are actually climatological at the present time, so the only actual effect would be the loss of the TMI precipitation estimates in the combined

microwave field. The impact on the final TMPA and I-MERG fields would be more serious, since they routinely use (will use) either the PR or the TCI product as a calibrator. In such a case, we would likely implement a "fail-safe" climatological calibration and accept the lower quality result in order to maintain continuity.

5.6 TIMING OF REPROCESSING FOR I-MERG PRODUCTS

As hinted in the previous Subsection, the decision to reprocess the production I-MERG archive as the result of algorithm changes in one or more input products critically depends on the availability of a completely reprocessed archive of the affected input product(s). In particular, when a general reprocessing is called for in the GPM suite of products, the I-MERG products can be started only after the requisite Level 2 products have been finalized and substantially reprocessed, allowing I-MERG to use the data for calibration and routine use in the products. Furthermore, such a general reprocessing probably implies the need to update the I-MERG RT products and a consequent need to reprocess the I-MERG RT archives as well.

6. ASSUMPTIONS AND LIMITATIONS

6.1 DATA DELIVERY

In general, the I-MERG package is extremely forgiving of dropouts in individual sensors, including the calibrating sensor products and the geo-IR data. Our experience with the TMPA is that extended drop-outs are rare for the TMI and PR (and so TCI). When the PR anomaly occurred in late June 2009, we completed the TMPA for June, then held production. Eventually, roughly 15 days of PR was declared missing, so we instituted a manual process to create calibrations that covered 30 days of data that adjoined the missing period. The TMPA-RT was unaffected, since the calibrations are climatological, as explained above. Extended geo-IR dropouts in the TMPA-RT have been back-filled as a service to users, but this is not required to enable continued RT production.

6.2 ASSUMED SENSOR PERFORMANCE

The implicit assumption in the I-MERG code is that the various PMW datasets are either stable or unavailable. There is a time-dependent calibration update for the PMW-IR calibration in both real and post-real time, and for the calibration of GMI to GCI in the post-real-time. So, if the IR is drifting the time-dependent calibrations should account for the problem. However, assuming that we use climatological GMI-to-everything-else calibrations, a drifting GMI cannot be accommodated. We would be more flexible if we decide to routinely update these GMI-to-everything-else calibrations, since drift in the GMI would be automatically accommodated. For all sensors, including geo-IR and gauge, variations in the amount of unbiased noise should not automatically bias the results, although the resulting random errors will fluctuate correspondingly.

7. REFERENCES

Adler, R.F., G.J. Huffman, A. Chang, R. Ferraro, P. Xie, J. Janowiak, B. Rudolf, U. Schneider, S. Curtis, D. Bolvin, A. Gruber, J. Susskind, P. Arkin, E.J. Nelkin, 2003: The Version 2 Global Precipitation Climatology Project (GPCP) Monthly Precipitation Analysis (1979-

Present). *J. Hdrometeor.*, **4**(6), 1147-1167.

Behrangi, A., K. Hsu, B. Imam, S. Sorooshian, G.J. Huffman, R.J. Kuligowski, 2009: PERSIANN-MSA: A Precipitation Estimation Method from Satellite-based Multi-spectral Analysis. *J. Hydrometeor.*, **10**(6), 1414-1429.

Behrangi, A., B. Imam, K. Hsu, S. Sorooshian, T.J. Bellerby, G.J. Huffman, 2011: REFAME: Rain Estimation Using Forward Adjusted-Advection of Microwave Estimates. *J. Hydrometeor.*, accepted.

Bellerby, T., K. Hsu, S. Sorooshian, 2009: LMODEL: A Satellite Precipitation Methodology Using Cloud Development Modeling. Part I: Algorithm Construction and Calibration. *J. Hydrometeor.*, **10**, 1081-1095.

Ebert, E.E., J.E. Janowiak, C. Kidd, 2007: Comparison of Near-Real-Time Precipitation Estimates from Satellite Observations and Numerical Models. *Bull. Amer. Meteor. Soc.*, **88**, 47-64.

Hong, K.L. Hsu, S. Sorooshian, X. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. Appl. Meteorol.* **43**, 1834–1852.

Huffman, G.J., R.F. Adler, D.T. Bolvin, G. Gu, E.J. Nelkin, K.P. Bowman, Y. Hong, E.F. Stocker, D.B. Wolff, 2007: The TRMM Multi-satellite Precipitation Analysis: Quasi-Global, Multi-Year, Combined-Sensor Precipitation Estimates at Fine Scale. *J. Hydrometeor.*, **8**(1), 38-55.

Huffman, G.J., R.F. Adler, M. Morrissey, D.T. Bolvin, S. Curtis, R. Joyce, B McGavock, J. Susskind, 2001: Global Precipitation at One-Degree Daily Resolution from Multi-Satellite Observations. *J. Hydrometeor.*, **2**(1), 36-50.

Joyce, R.J., P. Xie, J.E. Janowiak, 2011: Kalman Filter Based CMORPH. *J. Hydrometeor.*, to be submitted.

Krajewski, W.F., J.A. Smith, 1991: On the Estimation of Climatological Z–R Relationships. *J. Appl. Meteor.*, **30**, 1436–1445.

Miller, J.R., 1972: A Climatological Z-R Relationship for Convective Storms in the Northern Great Plains. *15th Conf. on Radar Meteor.*, 153-154.

Rudolf, B., U. Schneider, 2005: Calculation of Gridded Precipitation Data for the Global Land-Surface Using In-Situ Gauge Observations. *Proc. of the 2nd Internat. Precip. Working Group Workshop*, 25-28 October 2004, Monterey, CA, 231-247.

Schneider, U., T. Fuchs, A. Meyer-Christoffer, B. Rudolf, 2008: Global Precipitation Analysis Products of the GPCC. GPCC Internet Publication, DWD, 12 pp. Available online at ftp://ftp262-anon.dwd.de/pub/data/gpcc/PDF/GPCC_intro_products_2008.pdf.

Surussavadee, C., D.H. Staelin, 2008: Global Millimeter-Wave Precipitation Retrievals Trained with a Cloud-Resolving Numerical Weather Prediction Model, Part I: Retrieval Design. *IEEE Trans. on Geosci. and Rem. Sens.*, **46**, 99-108.

Surussavadee, C., D.H. Staelin, 2008: Global Millimeter-Wave Precipitation Retrievals Trained with a Cloud-Resolving Numerical Weather Prediction Model, Part II: Performance Evaluation. *IEEE Trans. on Geosci. and Rem. Sens.*, **46**, 109-118.

Susskind, J., P. Piraino, L. Rokke, L. Iredell, A. Mehta, 1997: Characteristics of the TOVS Pathfinder Path A dataset. *Bull. Amer. Meteor. Soc.*, **78,** 1449–1472.

Tian, Y., C. Peters-Lidard, J. Eylander, R. Joyce, G. Huffman, R. Adler, K.-L. Hsu, F. J. Turk, M. Garcia, J. Zeng, 2009: Component Analysis of Errors in Satellite-Based Precipitation Estimates. *J. Geophys. Res, - Atmos.*, **114**, D22104, doi:10.1029/2009JD011949.

Xie, P., S.-H. Yoo, R. Joyce, Y. Yarosh, 2010: A High-Resolution Gauge-Satellite Merged Global Precipitation Analysis and Its Applications for Model Verifications. *WMO 3rd Internat. Conf. on QPE / QPF and Hydrol.*, 18-22 October, 2010, Nanjing, China.

8. Acronyms

AIRS	Advanced Infrared Sounder
AMSR-E	Advanced Microwave Scanning Radiometer – Earth Observing System
ATLAS	Autonomous Temperature Line Acquisition System
ATMS	Advanced Technology Microwave Sounder
CLIMAT	Monthly Climatological Data
CMORPH	CPC Morphing technique
CONUS	contiguous U.S.
CPC	Climate Prediction Center
CPO	Climate Program Office
CrIS	Cross-track Infrared Sounder
CRU	Climate Research Unit
DISC	Data and Information Services Center
DMSP	Defense Meteorological Satellite Program
DPR	Dual-frequency Precipitation Radar
DWD	Deutscher Wetterdienst
DWSS	Defense Weather Satellite System
EUMETSAT	European organization for the exploitation of Meteorological Satellites
FAO	Food and Agriculture Organization
GCI	GPM Combined Instrument algorithm
GCOMW	Global Change Observation Mission - Water
geo	geosynchronous Earth orbit
GEWEX	Global Energy and Water cycle Experiment
GHCN	Global Historical Climatology Network
GMI	GPM Microwave Imager
GMS	Geosynchronous Meteorological Satellite
GOES	Geosynchronous Operational Environmental Satellite
GPCC	Global Precipitation Climatology Centre
GPCP	Global Precipitation Climatology Project
GPM	Global Precipitation Measurement mission
GSFC	Goddard Space Flight Center
HDF	Hierarchical Data Format
I-MERG	Integrated Multi-satellitE Retrievals for GPM
JMA	Japanese Meteorological Agency
JPSS	Joint Polar Satellite System
KF	Kalman filter version
leo	low Earth orbit
LMODEL	Lagrangian Model
MADRAS	Microwave Analysis and Detection of Rain and Atmospheric Structures
Meteosat	Meteorological Satellite
METOP	Meteorological Polar Orbit satellite

MHS	Microwave Humidity Sounder
MTSat	Multi-functional Transport Satellite
NASA	National Aeronautics and Space Administration
NCDC	National Climatic Data Center
NESDIS	National Environmental Satellite Data Information Service
NetCDF	Network Common Data Format
NOAA	National Oceanic and Atmospheric Administration
NPP	National Polar Operational Environmental Satellite System (NPOESS) Preparatory Project
PDF	Probability Density Function
PERSIANN-CCS	Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks – Cloud Classification System
PMM	Precipitation Measurement Missions
PMW	Passive Microwave
PPS	Precipitation Processing System
PR	Precipitation Radar
Q2	Next-Generation Multi-Sensor Quantitative Precipitation Estimates
REFAME	Rain Estimation using Forward Adjusted-advection of Microwave Estimates
RT	Real Time
SSMI	Special Sensor Microwave Imager
SSMIS	Special Sensor Microwave Imager/Sounder
SYNOP	Synoptic Weather Report
TCI	TRMM Combined Instrument
TMI	TRMM Microwave Imager
TOVS	Television-Infrared Observation Satellite (TIROS) Operational Vertical Sounder
TRMM	Tropical Rainfall Measuring Mission
USWRP	U.S. Weather Research Program
UTC	Universal Coordinated Time