V06 IMERG Release Notes

George J. Huffman, David T. Bolvin, Eric J. Nelkin, Erich F. Stocker, Jackson Tan

2 May 2019

The algorithm for the Integrated Multi-satellitE Retrievals for GPM (IMERG) has now been upgraded to Version 06. The transition to V05 for the IMERG Final Run began 13 March 2019 at PPS and the new data started flowing down to the GES DISC as well. However, on 15 March 2019 an error was discovered in processing the initial batches of V06 IMERG Final Run months. A design choice in the code ended up retaining microwave precipitation estimates in the latitude band 60°N-S when there is snow/ice on the surface, rather than masking out the estimates due to low performance in such cases. [Masking for microwave retrievals over surface snow and ice outside that band is correct.] All then-existent V06A Monthly and Half-hourly IMERG products (June 2014-December 2015) were retracted, and corrected datasets released. The GPM-era Final Run begins on 1 June 2014 to allow a spin-up of calibration files and extends to the (delayed) present. Retrospective processing for the TRMM era Final Run began on 15 April 2019. Due to IR data issues at present, the actual start of the record is June 2000 to allow spin-up of the calibration files. Initial Processing of the Early and Late Runs will begin on 1 May, with Retrospective Processing commencing when the Final Run is fully retrospectively processed. Version 06A supersedes all prior IMERG versions, and users are urged to shift to the new datasets as soon as practical.

Changes from Version 05 to 06

- The dimensions of the (native) HDF5 data fields have changed from (lon, lat) to (time, lon, lat), with the corresponding lengths of (1, 3600, 1800). This is done to conform to GES DISC archiving conventions. In most cases this should be transparent to application programs, but some users have reported needing to re-work code that makes individual calls to the HDF API.
- Up through V05, the "displacement vectors" used in the quasi-Lagrangian time interpolation scheme were computed from the IR data. In V06, these vectors are now computed from Modern Era Retrospective Reanalysis 2 (MERRA-2) and Goddard Earth Observing System model (GEOS) Forward Processing (FP) data, as ingested by PPS. See *https://pmm.nasa.gov/sites/default/files/document_files/MorphingInV06IMERG.pdf* for details.
- GPROF V05 continues to be used to compute precipitation estimates for all microwave sensors as input, except for SAPHIR, for which the Precipitation Retrieval and Profiling Scheme (PRPS) is used.
- GPROF-TMI V05 estimates are computed for the GPM era and incorporated in the V06 IMERG datasets (unlike V05).
- We continue to provide estimates from all constellation members in the microwave-only precipitation field (HQprecipitation) and the complete precipitation fields (precipitationCal, precipitationUncal) over the fully global domain (90ºN-S). Note: microwave estimates over snowy/icy surface types are not masked out in HQprecipitation, but are for precipitationCal and precipitationUncal (as was done in V05). So, in the latter two, IR fill-in occurs over 60°N-S and there are holes at higher latitudes. This was true in V05, except V06 morphs the microwave estimates at the high latitudes, so the holes are much smaller than the overpass-only scheme used at high latitudes in V05.
- SAPHIR estimates are incorporated into IMERG for the first time in V06, and are computed with PRPS. However, communication issues with Megha Tropiques that began on 15 December 2018 are preventing any use of SAPHIR data from that date to the present.

- The Version 05 GPROF estimates for AMSU, ATMS, and MHS do not provide estimates for the 5, 8, and 5 footprints (respectively) at each swath edge due to algorithm issues revealed in early testing. The same is true for 5 footprints at each swath edge for the PRPS estimates computed for SAPHIR.
- At the start of the GPM era (March 2014), we continue to use TRMM-based calibrations through May 2014, unlike V05, in which GPM-based calibrations started immediately. This is done to allow the GPM-based calibrations to spin up before they are applied, since "full" calibrations are available from TRMM. Due to code development considerations, GPROF-GMI V05 estimates are not incorporated in the V06 IMERG datasets while TRMM-based calibrations continue in this first 2.5 months of the GPM era. We expect to use GMI in this first 2.5 months of the GPM era in V07 IMERG.
- We continue the practice of thresholding input precipitation rates to adjust fractional coverage; GPROF estimates are currently thresholded at 0.03 mm/h, and PRPS-SAPHIR has a threshold of zero.
- Prior internal storage limitations were removed, eliminating discretization of the output precipitation rates and increasing the maximum permitted rate from 50 to 200 mm/h.
- We continue the practice started with V04 of calibrating 2BCMB to GPCP V2.3 over ocean (at middle and high latitudes) and land (at low and middle) to compensate for low (high) 2BCMB bias over non-tropical oceans (land).
- We now estimate the t=0 (i.e., actual microwave overpass data, contained in the HQprecipitation field) correlations in the half-hourly Quality Index; in V05 they were identically 1. See "IMERG Quality Index" document (*https://pmm.nasa.gov/sites/default/files/document_files/IMERG_QI.pdf*) for a summary.
- "Rippling" in the animation of IMERG data in the GPM era has been reduced somewhat from V05 to V06, but continues to be visible. This indicates that the different sensors continue to have different depictions of the same precipitation features due to each sensor type's unique combination of resolution and channel selection. We see similar fluctuations in early animations of the TRMM era as well.

Specific Comments About the TRMM Era

Significant effort has gone into ensuring that there is reasonable homogeneity in the IMERG record given the constantly evolving satellite constellation. This includes ensuring that the Level 1 (brightness temperature) and Level 2 (precipitation) data are as consistent as possible with the GMI across the constellation of passive microwave radiometers during the GPM era, and correspondingly TMI in the TRMM era. Then IMERG enforces calibration to the perceived "best" core satellite estimate, which is the combined radar-radiometer algorithm (CORRA), as well as selective climatological calibration to a recognized standard (the Global Precipitation Climatology monthly satellite-gauge combined product) where the combined product is known to be low (high-latitude oceans) or high (tropical and mid-latitude land). Despite this intercalibration, users should note that variations remain across sensors and between the TRMM and GPM eras. Specifically:

- Version 06 is the first time that CORRA and IMERG have been computed for both the TRMM and GPM era. As such, we expect to learn a great deal about how the detailed behavior of the IMERG estimates differs between the two eras.
- The TRMM CORRA is based on fewer channels of microwave and radar data than for the GPM CORRA, so one would expect the underlying calibrations to shift across the transition boundary (May/June 2014). This could affect the histogram of precipitation rates and/or the mean rates.
- The TRMM CORRA only covers the latitude band 35°N-S, while it is 65°N-S for GPM. Thus, calibrations to TRMM CORRA outside 35°N-S are necessarily approximate, based on the

(monthly climatological) "shape" of the GPM CORRA calibration field, scaled to match the precipitation volume for each month of TRMM-era calibration for northern and southern hemisphere ocean and land separately in the region 33°N-S. See the IMERG Technical Document (listed below in "Key IMERG Documents") for more details.

- The early TRMM era contains estimates from SSMI and AMSU-B sensors, which are less capable than the subsequent SSMIS and MHS sensors that were phased in starting in 2005. This implies that a shift in estimation skill is likely.
- The early TRMM era lacked microwave observations for several hours around the 0/12 and 04/16 UTC orbital times. As a result, there is more use of IR-based estimates in the first few years. IR estimates are generally of lower quality and potentially introduce a systematic regional offset in the timing of precipitation systems. [IR Tb's tend to lag precipitation occurrence, but IMERG uses the PERSIANN-CCS algorithm, which considers other factors, such as spatial texture in the IR Tb's.] Furthermore, the Japanese GMS observations only provide hourly data for certain periods of the day, and in the missing half-hours the adjacent METEOSAT and GOES-W IR values are used to the extent possible, but in some cases the necessary zenith-angle parallax corrections apparently yield very different IR values, leading to dithering between two precipitation estimate scenes in successive half hours. Before 2005 there are large gaps in the available microwave data, which makes the behavior of the IR estimates more important. In addition, there is apparently a somewhat complicated interaction between the morphed microwave and available IR data.
- There is an extended data dropout in the presently available IR dataset for Japanese sector from late on 17 November 2005 to the middle of 22 March 2006 due to issues in NOAA coping with the format of the then-newly introduced MTSat-1. Consequently, during that period a small sector over Japan lacks all data, and the adjacent IR are entirely based on high-zenith-angle data from the METEOSAT to the west and GOES-W to the east.

Additional Access Information

The data archive sites are now populating the various repositories of original and value-added data with the new V06. See *https://pmm.nasa.gov/data-access/downloads/gpm* to find all of the various formats and their locations. Recall that access to the various systems (PPS, PPS near-real time, and GES DISC) requires separate simple, free, and automatic registrations to satisfy NASA data system requirements.

IMERG Early, Late, and Final Run datasets are computed about 4 hours, 14 hours, and 3.5 months after observation time, respectively. Early and Late are computed an hour at a time, while Final is computed a month at a time. The native HDF5 half-hourly products for Early, Late, and Final runs have the prefixes "*3B-HHR-E*". "*3B-HHR-L*", and "*3B-HHR*", respectively. The monthly products have the prefix "*3B-MO*", and are only computed as a native product for the Final Run. The complete file naming convention can be found at

https://pmm.nasa.gov/sites/default/files/document_files/FileNamingConventionForPrecipitationProductsForGPMMissionV1.4.pdf.

The version number for the revised release is Version 06A. The field named *precipitationCal* contains the "complete" IMERG precipitation estimate.

Additional Notes

Recall that the Final Run uses calibrations based on accumulations of match-ups that include the entire current month and data from previous months, while Early and Late Runs necessarily use

calibrations based on trailing accumulations of match-ups, since these cannot be computed into the future. In addition, the Early Run only has forward propagation of the microwave data (unlike both the Late and Final Runs), and neither has calibration to the monthly gauge data as in the Final Run. Both the Early and Late Runs used the latest "seed" calibration files for the Kalman coefficients and the 2BCMB-GMI and HQ-IR calibrations from the Final Run of V06A on 30 April 2019 to begin Initial Processing on 1 May 2019. Accordingly, users should expect the estimates in May 2019 to be less accurate than following months of data that will have calibrations fully populated with V06A. By contrast, the Final Run does not require a seed file, since the calibrators are spun up to start on 1 June 2019. Validation results will be posted as they are developed.

Complete records of all Runs are posted (honoring the necessary latency), labeled as V06A. The current retrospective computation starts in June 2000 because the necessary IR data are not presently available until February 2000, and then the calibrators must be spun up. It is planned that all three Runs will be retrospectively processed back to the start of TRMM (January 1998) when the requisite data become available.

The Quality Index (QI) document (see below) describes the QI variables for both half-hourly and monthly data. This includes advice about what might be considered "good" ranges of QI. We encourage users to report their experience with QI (and any other variable) to aid in developing the next versions of IMERG.

Key IMERG Documents

Algorithm Theoretical Basis Document
https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V06.pdf pdf

Morphing Vector Document
https://pmm.nasa.gov/sites/default/files/document_files/MorphingInV06IMERG.pdf

Quality Index
https://pmm.nasa.gov/sites/default/files/document_files/IMERG_QI.pdf

Technical Documentation
https://pmm.nasa.gov/sites/default/files/document_files/IMERG_doc.pdf

First Validation Results for the Final Run

Initial Comparison to MRMS-Based Surface Reference (Jackson Tan)

One key validation is to compare IMERG precipitation estimates to the Multi-Radar Multi-Sensor (MRMS) suite of precipitation observations over CONUS processed in support of the GPM mission for ground validation. MRMS is produced by NOAA/NSSL, based primarily on the WSR-88D network over conterminous US with bias correction using gauge data from the Hydrometeorological Automated Data System and regional rain gauge networks. Only the MRMS pixels with the best quality are selected, requiring

1) a perfect radar quality index, which filters out sampling and estimation uncertainty such as beam issues due to orography and bright band;
2) gauge correction factors within the range [0.5, 2], which removes pixels in which the radar and gauge estimates differ by more than a factor of 2; and
3) only liquid-phase precipitation.

The first evaluation here is conducted with the IMERG Final Run "complete" precipitation fields with and without the last calibration (precipitationCal [bottom] and precipitationUncal [top], respectively) for both Version 05 and 06 at the native IMERG resolution (0.1°, 30 min) from June 2014 to May 2015. The rain/no-rain threshold is chosen to be 0.2 mm/h. The Final Uncal is likely representative of the Late Cal, while the Final Cal should be the best answer.

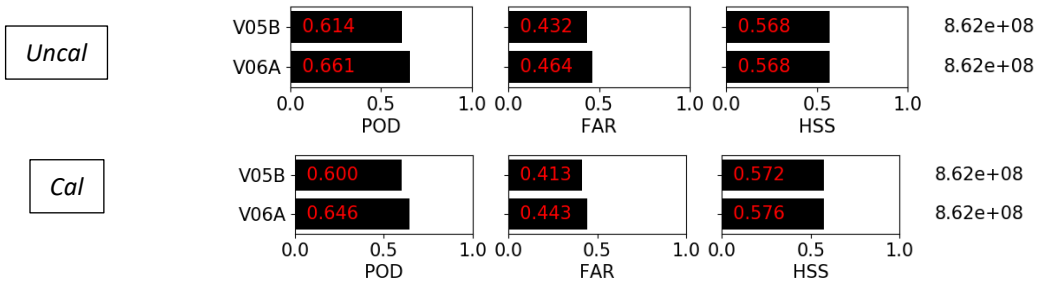

Fig. 1. Probability of detection (POD), false alarm ratio (FAR), and Heidke Skill Score (HSS) for Uncal (top) and Cal (bottom). The numbers on the right indicate the sample sizes. The precipitation detection skill is similar between V05 and V06 in terms of HSS, reflecting an increase in both POD and FAR in V06.

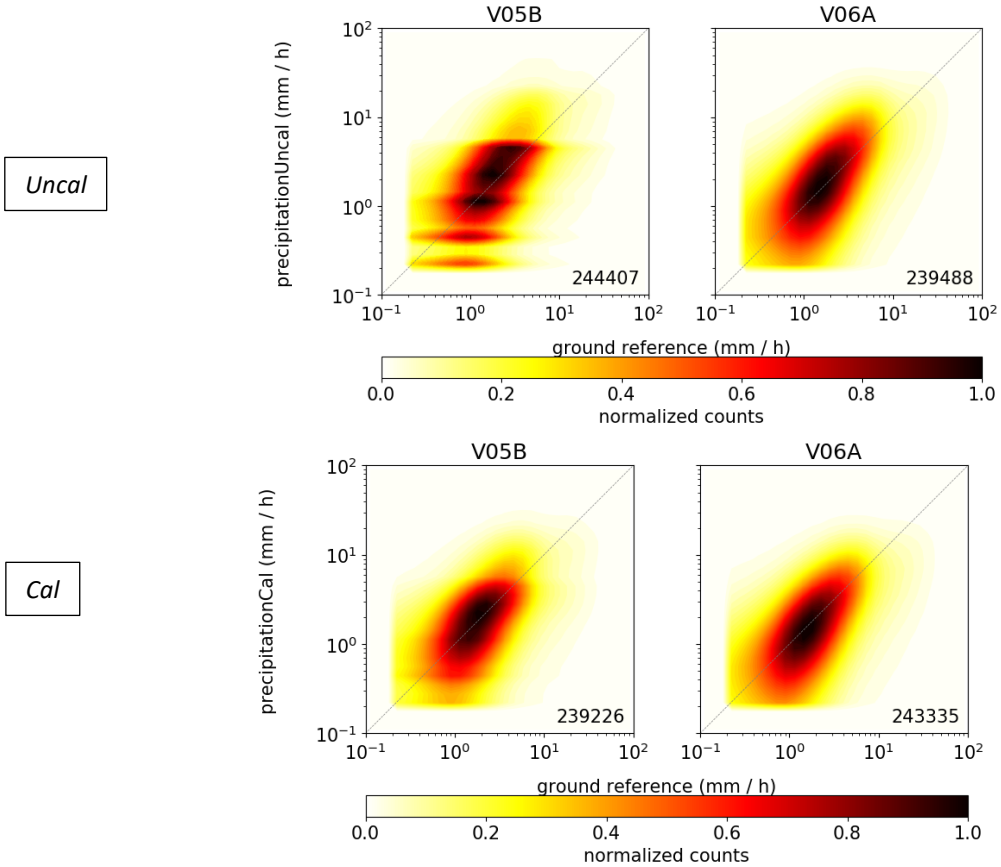

Fig. 2. Joint distributions of the rain rates between IMERG and the MRMS-based ground reference, limited to cases where both meet the 0.2 mm/h threshold) for Uncal (top) and Cal (bottom). The number in the lower-right corner of each plot shows the normalization factor for the color scale. There is a distinct improvement from V05 to V06, with (i) the elimination of

horizontal bands in the uncalibrated estimates due to the elimination of discretized values (see above), and (ii) a better representation of higher precipitation rates that can primarily be traced to estimates from the improved morphing scheme.

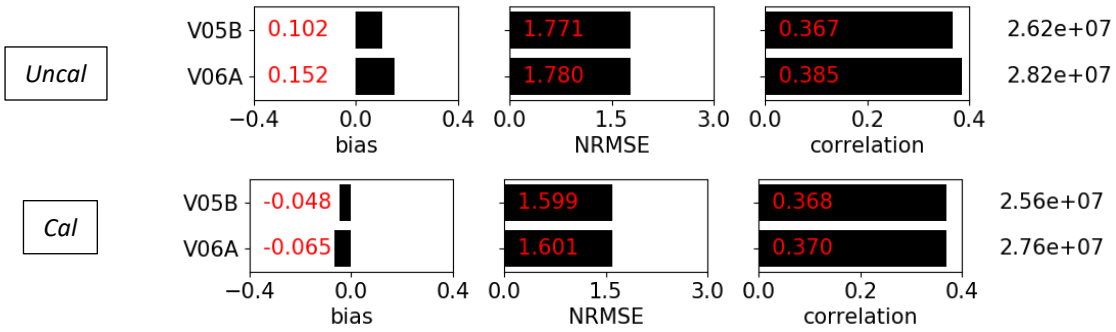

Fig. 3. Bias, normalized root mean square error (NRMSE), and Pearson correlation coefficient for cases where both IMERG and the MRMS-based ground reference meet the 0.2 mm/h threshold) for Uncal (top) and Cal (bottom). The numbers on the right indicate the sample sizes (number of hits). Between V05 and V06, there is a slight increase in the bias of the uncalibrated estimates, a marginal increase in NRMSE, and a modest increase in the correlation.