A Novel Approach to Identify Sources of Errors in IMERG for GPM Ground Validation

JACKSON TAN

*Universities Space Research Association, and NASA Goddard Space Flight Center,
Greenbelt, Maryland*

WALTER A. PETERSEN

Earth Science Office, NASA Marshall Space Flight Center, Huntsville, Alabama

ALI TOKAY

*University of Maryland, Baltimore County, Baltimore, and NASA Goddard Space Flight Center,
Greenbelt, Maryland*

(Manuscript received 1 April 2016, in final form 15 July 2016)

ABSTRACT

The comparison of satellite and high-quality, ground-based estimates of precipitation is an important means to assess the confidence in satellite-based algorithms and to provide a benchmark for their continued development and future improvement. To these ends, it is beneficial to identify sources of estimation uncertainty, thereby facilitating a precise understanding of the origins of the problem. This is especially true for new datasets such as the Integrated Multisatellite Retrievals for GPM (IMERG) product, which provides global precipitation gridded at a high resolution using measurements from different sources and techniques. Here, IMERG is evaluated against a dense network of gauges in the mid-Atlantic region of the United States. A novel approach is presented, leveraging ancillary variables in IMERG to attribute the errors to the individual instruments or techniques within the algorithm. As a whole, IMERG exhibits some misses and false alarms for rain detection, while its rain-rate estimates tend to overestimate drizzle and underestimate heavy rain with considerable random error. Tracing the errors to their sources, the most reliable IMERG estimates come from passive microwave satellites, which in turn exhibit a hierarchy of performance. The morphing technique has comparable proficiency with the less skillful satellites, but infrared estimations perform poorly. The approach here demonstrated that, underlying the overall reasonable performance of IMERG, different sources have different reliability, thus enabling both IMERG users and developers to better recognize the uncertainty in the estimate. Future validation efforts are urged to adopt such a categorization to bridge between gridded rainfall and instantaneous satellite estimates.

1. Introduction

Satellite remote sensing of precipitation is critically important because of the absence of ground measurements in many parts of the world, including over oceans, mountainous

regions, and sparsely populated areas. Early efforts on space-based precipitation retrievals focused on estimating rainfall from infrared measurements of cloud tops from geosynchronous satellites, but accuracy was limited because of the indirect connection between surface rain rates and cloud-top temperatures. Much progress has been made in the last two decades with a contingent of low-Earth-orbiting passive microwave satellites and two NASA/JAXA spaceborne radars in the microwave band, the Tropical Rainfall Measuring Mission (TRMM) and the Global Precipitation Measurement (GPM) mission. Unlike infrared radiation, microwave radiation is able to penetrate clouds and interact more directly with precipitation; consequently, microwave retrieval techniques generally provide a superior

Supplemental information related to this paper is available at the Journals Online website: <http://dx.doi.org/10.1175/JHM-D-16-0079.s1>.

Corresponding author address: Jackson Tan, NASA Goddard Space Flight Center, Code 613, Building 33, Room C327, 8800 Greenbelt Rd., Greenbelt, MD 20771.
E-mail: jackson.tan@nasa.gov

DOI: 10.1175/JHM-D-16-0079.1

estimate of precipitation. Furthermore, when the microwave radiation is employed in an active sense, such as with spaceborne radars, a three-dimensional structure of rainfall is provided, thus facilitating a determination of precipitation structure (e.g., convective or stratiform) and phase (liquid, frozen, and melting), often resulting in improved retrievals of precipitation rate.

Numerous gridded rainfall products have been produced by merging microwave and infrared products from low-Earth-orbiting and geosynchronous satellites, including Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS; [Hong et al. 2004](#)), the Climate Prediction Center (CPC) morphing technique (CMORPH; [Joyce et al. 2004](#); [Joyce and Xie 2011](#)), Global Satellite Mapping of Precipitation (GSMaP; [Kubota et al. 2007](#); [Kachi et al. 2014](#)), the Naval Research Laboratory blended-satellite technique (NRL blended; [Turk and Miller 2005](#)), and TRMM Multisatellite Precipitation Analysis (TMPA; [Huffman et al. 2007](#)). These gridded precipitation datasets differ in their sources of data and in their algorithms, as well as in their resolutions and coverage. They have been extensively compared to ground-based gauges and radars over different regions ([AghaKouchak et al. 2012](#); [Bharti and Singh 2015](#); [Chen et al. 2013a,b](#); [Ebert et al. 2007](#); [Gottschalck et al. 2005](#); [Gourley et al. 2010](#); [Habib et al. 2009, 2012](#); [Hossain and Huffman 2008](#); [Krajewski et al. 2000](#); [Kubota et al. 2009](#); [Li et al. 2013](#); [Maggioni et al. 2014](#); [Mei et al. 2014](#); [Roca et al. 2010](#); [Sapiano and Arkin 2009](#); [Sapiano 2010](#); [Stampoulis and Anagnostou 2012](#); [Tang et al. 2015](#); [Tian and Peters-Lidard 2007](#); [Tian et al. 2007, 2009](#); [Tian and Peters-Lidard 2010](#); [Tian et al. 2010](#); [Villarini and Krajewski 2007](#); [Xue et al. 2013](#); [Yong et al. 2014, 2015](#)). Hence, their errors and biases are well studied.

One of the goals of the GPM mission is to “provide uniform global precipitation products from a heterogeneous constellation of microwave sensors within a consistent framework” ([Hou et al. 2014](#)). To this end, the Integrated Multisatellite Retrievals for GPM (IMERG) was developed ([Huffman et al. 2015](#)). IMERG integrates algorithms from TMPA, CMORPH, and PERSIANN to produce rainfall estimates at 0.1° resolution every half hour between the latitudes of 60°N and 60°S . Currently, IMERG extends from April 2014 to present (with three different latency periods to cater to a range of user requirements), but IMERG processing will eventually be extended backward to the TRMM era (1998 onward).

In this study, we evaluate the IMERG Final run against a dense gauge network at the pixel level. Initial comparison of this Day-1 IMERG with a gauge network in China suggests that IMERG is at least as good as TMPA in its rainfall estimation ([Tang et al. 2016a,b](#);

[Guo et al. 2016](#)), while comparison of monthly averages with TMPA reveals systematic differences as a function of surface type and season ([Liu 2016](#)). Our aim is not just to compare IMERG with ground measurements but also to identify the sources and/or the algorithm from which differences between the IMERG and ground measurements arise. For example, existing biases in one of the microwave sensors may be causing a persistent overestimation of rain rates in IMERG when estimates from this sensor are used. Not only will such knowledge allow IMERG developers to improve the algorithm, it will also benefit users by providing information on situations when IMERG is more reliable. To achieve this aim, we compare IMERG against a dense network of 36 GPM ground validation (GV) gauges—all located within one IMERG pixel—in the mid-Atlantic coastal region in the United States over a period of 18 months. We evaluate its ability in identifying both rain occurrences and estimated rain rates. These results are contrasted against the radar- and gauge-based Multi-Radar/Multi-Sensor (MRMS) system from NOAA/NSSL ([Zhang et al. 2011](#)). To track down the source of the error, we employ a novel approach, leveraging on the existence of ancillary variables in IMERG to trace the ground validation back to the source of the estimates. This proof-of-concept study demonstrates the value of such an approach, which will be extended in coverage for a comprehensive evaluation in a future study. This enables us to bridge our ground validation efforts between gridded rainfall and instantaneous estimates from individual satellites, thereby introducing the possibility of evaluating “level 2” retrievals through gridded products.

2. Data and methods

a. IMERG

IMERG is a level 3 gridded precipitation product that unifies measurements from a network of satellites in the GPM constellation ([Huffman et al. 2015](#)). IMERG uses the GPM *Core Observatory* satellite, which carries a dual-frequency precipitation radar and a 13-channel passive microwave imager, as an on-orbit reference standard to intercalibrate and merge the individual passive microwave (PMW) precipitation estimates from the GPM constellation. Lagrangian time interpolation is then applied to the constellation estimates using displacement vectors derived from infrared (IR) measurements on geosynchronous satellites to produce gridded estimates of rainfall at fine resolution quasi globally. This process is called morphing and was first introduced as a key component of the CMORPH precipitation estimation algorithm ([Joyce et al. 2004](#); [Joyce and Xie 2011](#)). This

gridded estimate is further supplemented, via a Kalman filter, with microwave-calibrated rainfall estimates calculated directly from IR measurements using an artificial neural network model based on the PERSIANN-CCS algorithm (Hong et al. 2004). The final satellite estimate is then calibrated using gauge data from the Global Precipitation Climatology Centre (GPCC) monthly precipitation dataset following the approach employed in TMPA (Huffman et al. 2007).

IMERG has a high resolution of 0.1° every half hour covering up to $\pm 60^\circ$ latitudes. It has three runs to accommodate the different user requirements for latency and accuracy. The Early run, available at a 6-h delay for real-time applications such as in the prediction of flash floods, is limited in rainfall morphing to propagation only forward in time. The Late run, with an 18-h delay for purposes such as crop forecasting, employs forward and backward morphing in time. Both the Early and Final runs have climatological gauge calibrations. The Final run is at a 4-month delay for research applications, with monthly gauge adjustments to control for bias. Note that these delays will eventually be reduced toward the targets of 4 h, 12 h, and 2 months, respectively. In addition, latency in data delivery means that the Final run may have more PMW estimates than the Late run, which in turn may have more PMW estimates than the Early run. This study focuses on the Final run of Day-1 IMERG, using the gauge-adjusted estimates from 18 months of data between April 2014 and September 2015. This “Day-1” designation indicates that this version of IMERG uses the “Version 3” algorithms implemented at the launch of the GPM *Core Observatory* in February 2014 and is thus subject to further improvements as measurements are collected. We choose to evaluate the gauge-adjusted values since they are the quantity expected to be most commonly used; however, our conclusions are unchanged when repeated with the unadjusted estimates (not shown). IMERG can be accessed at <http://pmm.nasa.gov/data-access>.

As IMERG is a multisatellite product, the constitution of the PMW sources changes with time. For the current version of IMERG, there are five PMW instruments contributing to IMERG: GMI, AMSR, MHS, SSMIS, and TMI (see Table 1 for expansion of the acronyms). Note that other sensors may be listed as part of the constellation (Huffman et al. 2015), but they are currently not used in IMERG because of the uncertain quality of the estimates or differences in time period. Of the five instruments in use, MHS is the only sounder while the rest are imagers, which are considered better in the low and midlatitudes. When there are more than one satellite overpasses at a time step, IMERG prioritizes imagers over sounders, followed by the overpass

TABLE 1. Sensors used in the current version of IMERG. Index refers to the value in HQprecipSource (NASA PPS 2015). See Huffman et al. (2015) for the complete list of sensors in the GPM constellation.

Index	Acronym	Satellite name	Sensor type
1	TMI	TRMM Microwave Imager	Imager
3	AMSR	Advanced Microwave Scanning Radiometer	Imager
5	SSMIS	Special Sensor Microwave Imager/Sounder	Imager
7	MHS	Microwave Humidity Sounder	Sounder
9	GMI	GPM Microwave Imager	Imager

with an observation time closest to the center of the half hour. In pixels without a direct estimate from a PMW satellite overpass, IMERG uses IR imagers on geosynchronous platforms to complete the rainfall record in two ways. First, it provides the motion vectors for morphing between PMW estimates. Second, it allows for a direct estimation of rain rates on the basis of the connection between IR brightness temperature and cloud-top height. Precipitation estimates from morphing are generally more accurate when propagated no more than 90 min from the PMW measurement time, beyond which the inclusion of direct IR estimates via Kalman filter is beneficial (Joyce and Xie 2011). The morphing and direct precipitation estimation procedures used with the IR information should not be confused. In section 4, we will categorize the IMERG estimates into each of the five PMW instruments, as well as three additional IR-based categories: derived purely from morphing (morph only), a mixture of morphing and direct IR (IR + morph), and direct IR only (IR only).

To perform the analysis described above, we make use of numerous variables in the IMERG files. The variable precipitationCal provides the final IMERG estimate of rain rates (mm h^{-1}). HQprecipSource identifies the PMW instrument providing the estimate [see NASA PPS (2016) for the key between the value and the instrument name]. Note that, at the moment, this variable does not distinguish the same instrument on different platforms. If there is no PMW observation, we then use IRkalmanFilterWeight to determine which of the three IR-based categories the estimate belongs to: a weight of 0% is morph only, a weight of 100% is IR only, and anything in between is IR + morph. Because of the poor performance of PMW estimates over ice- or snow-covered surfaces, the IMERG estimate may use the direct IR estimate instead. For such instances, the IRkalmanFilterWeight is 100% even though HQprecipSource indicates a PMW

FIG. 1. Region of analysis. (left),(center) The red box shows the U.S. Eastern Shore, a mid-Atlantic coastal region, where we examine the IMERG estimate in a pixel (indigo box). (right) Locations of the gauge platforms (green crosses), each of which has two gauges and all fall into a grid (38.05°–38.10°N, 75.60°–75.54°W; green box) that is a subset of the IMERG pixel. The period of analysis is from April 2014 to September 2015.

observation (D. Bolvin 2016, personal communication). We classify such instances as IR only; they occur less than 1% of the time for our study. Additionally, for a case study in section 4, we use HQprecipitation, which is the PMW estimate after regridding to the IMERG grids and cross calibration between all instruments but prior to gauge adjustment. Note that these variable names are valid for the IMERG files used at the time of writing and may change in the future.

b. Gauges

The NASA GPM GV program deploys a dense network of dual tipping-bucket rain gauges near the NASA Wallops Flight Facility along the Eastern Shore region of the United States (red box in Fig. 1; <http://gpm-gv.gsfc.nasa.gov/Gauge/index.html>). This mid-Atlantic region is a coastal environment with the absence of significant orography, and the main rainfall systems in the area are tropical cyclones (or remnants of tropical cyclones), airmass convection, nor'easters, and frontal rainfall (Tokay et al. 2014). The gauge platforms are located in the vicinity of Pocomoke City, Maryland (green crosses in Fig. 1), with a reasonably uniform distribution in a rectangular box (green box in Fig. 1) that is a subset of a single IMERG grid box (indigo box in Fig. 1). In section 3, we will investigate the impact of the uneven distribution of gauges within the IMERG pixel. In this study, we ignore gauges that either lie outside this IMERG grid box or moved during the entire period; this leaves us with 18 gauge platforms. Since each platform has two rain gauges, the gauge rain rates against which IMERG will be validated is an average of 36 gauges. These gauges are not part of the GPCC network that is used for calibration in IMERG, making them an independent dataset for verification purposes.

Each pair of gauges on a single platform is separated by approximately 1 m. Such a dual-gauge system allows for quality control and for the calculation of sampling uncertainty. Each gauge records the time of tipping, with each tip corresponding to 0.254 mm of rain, and these data are transmitted in real time to Wallops Flight Facility. We convert the tip times into rain rates by accumulating them over half-hour windows and converting them to millimeters per hour, such that we have a dataset of ground reference with a time resolution matching that of IMERG. As these gauges are constantly maintained and calibrated on a regular and continual schedule, the measured rain rates can be considered highly reliable. The variance reduction factor (Villarini et al. 2008) of the gauges is 1.7%. Furthermore, for quality control, we ignore points in which the half-hour accumulations between two gauges differ by 1) more than one tip and 2) more than half the mean of both gauges. As tipping-bucket gauges cannot give an accurate measure of snow water equivalent, we exclude every day with reported snowfall from the Global Historical Climatology Network–Daily from NOAA and the day following that (to account for melting hysteresis in the measurement). Therefore, this study is limited to liquid precipitation only. We also exclude the last 10 days of our record (beginning on 21 September 2015) because of issues with gauge data records. In total, 39 days of data are excluded.

c. MRMS/Q3 product

The MRMS (formerly National Mosaic and Multi-Sensor QPE) system is a gridded product generated by NOAA/NSSL and based primarily on the U.S. WSR-88D network (Zhang et al. 2011). Reflectivity data are mosaicked onto a common 3D grid with quality control for beam blockages and bright band. Physically based heuristic rules are applied to the reflectivity structure

and environmental field at each grid point to determine the precipitation regime (e.g., snow and convective rain), from which a reflectivity–precipitation relationship is used to estimate the surface precipitation rate. These precipitation rates are corrected for bias using gauge data from the Hydrometeorological Automated Data Systems and regional rain gauge networks. Note that the Wallops Flight Facility gauges (section 2b) are not part of the MRMS system. For each precipitation estimate in MRMS, a radar quality index (RQI) is produced (Zhang et al. 2012). This index reflects sampling and estimation uncertainty, such as beam issues relating to orography and the presence of solid precipitation.

For the analysis herein, we use the “Level 3” MRMS dataset processed in support of the GPM mission (Kirstetter et al. 2012). This product gives the hourly accumulated rain rate over the conterminous United States (20°–55°N, 130°–60°W) with a high spatial resolution of 0.01°. The high spatial resolution allows us to calculate the averages over the gauges (30 pixels; green box in Fig. 1) and the IMERG grid box (100 pixels; indigo box in Fig. 1), thus providing a connection between the estimates over the two areas, an advantage we will exploit in section 3. We make extensive use of the RQI in the data analysis. For this Level 3 MRMS product, the RQI ranges from 0 (lowest confidence) to 100 (highest confidence). In general, the RQIs of the MRMS pixels over the gauges and IMERG grid box are 100, chiefly due to the flatness of the region and the proximity of the radar at Dover Air Force Base, Delaware (~80 km). Nevertheless, we mask MRMS pixels with RQI less than 100, thus keeping only perfect-RQI MRMS pixels in computing the areal averages. For most parts in this study, MRMS is used not as a ground reference but as a context against which the performance of IMERG is compared.

d. Approach

In our study, we adopt the proposal suggested by Tang et al. (2015) and evaluate the gridded products over the period of 18 months (from April 2014 to September 2015) from two perspectives. First, we compute the contingency tables, that is, the hits H , misses M , false alarms F , and correct negatives C between each pair of rain rates. A hit occurs when both the reference and the estimate are raining, a miss occurs when the reference is raining but the estimate is not, a false alarm occurs when the reference is not raining but the estimate is, and a correct negative occurs when both the reference and the estimate are not raining. We consider an instance to be raining at a time step when the rain rate at that instance is at least 0.1 mm h^{-1} . If it is below 0.1 mm h^{-1} , it is considered not raining. This threshold is an acknowledgment of measurement uncertainty, whereby values lower than

0.1 mm h^{-1} may be a result of noise. The choice of this threshold is subjective, though our conclusions remain robust when we repeat our analysis for thresholds of 0.2 and 0.5 mm h^{-1} (see supplemental material). These two thresholds correspond to the estimated minimum detectable rain rates of the Ka- and Ku-band radar against which GMI is calibrated (Hou et al. 2014), though it should be noted that subsequent processing to the final IMERG estimate may produce rain rates lower than these two thresholds. Note that 0.1 mm h^{-1} constitutes at least eight tips in the 36-gauge average.

From the hits, misses, false alarms, and correct negatives, we can calculate four scores that quantify the estimates performance in terms of identifying rain occurrences: probability of detection (POD), false alarm ratio (FAR), bias in detection (BID), and Heidke skill score (HSS). The POD is simply the fraction of actual rain occurrences that the estimate detected; a perfect score is 1. The FAR is the fraction of rain occurrences in the estimates that are actually wrong, that is, false alarms; a perfect score is 0, implying that the satellite algorithm never estimates a qualified precipitation rate when the ground reference indicated no precipitation. The BID quantifies the tendency for the estimate to overestimate ($\text{BID} > 1$) or underestimate ($\text{BID} < 1$) the number of rain occurrences; a perfect score is 1. The HSS is a generalized skill score that quantifies whether the estimate is worse ($\text{HSS} < 0$) or better ($\text{HSS} > 0$) than random chance relative to the reference; a perfect score is 1. The formulas for these scores are

$$\text{POD} = \frac{H}{H + M}, \quad (1)$$

$$\text{FAR} = \frac{F}{H + F}, \quad (2)$$

$$\text{BID} = \frac{H + F}{H + M}, \quad \text{and} \quad (3)$$

$$\text{HSS} = \frac{H + C - E}{N - E}, \quad (4)$$

where N is the sample size and E is the expected number of instances that can be correctly identified based on random chances alone, calculated with

$$E = \frac{1}{N} [(H + M)(H + F) + (C + M)(C + F)]. \quad (5)$$

After quantifying the rain occurrences, we will then focus on the hits and analyze the rain rates from this subset. To compare the rain rates, we will use three conventional measures, the normalized mean error (NME), normalized mean absolute error (NMAE), and Pearson correlation coefficient (CC), as well as the multiplicative error model introduced in Tian et al. (2013). The NME

and NMAE quantify the degree of systematic and random error respectively, normalized by the mean rain rate of the reference. They are defined here as

$$\text{NME} = \frac{\frac{1}{n} \sum_i (y_i - x_i)}{\frac{1}{n} \sum_i x_i} \quad \text{and} \quad (6)$$

$$\text{NMAE} = \frac{\frac{1}{n} \sum_i |y_i - x_i|}{\frac{1}{n} \sum_i x_i}, \quad (7)$$

where x_i and y_i are the rain rates for the reference and estimates, respectively, and n is the number of hits. Perfect NME and NMAE are 0. Note that root-mean-square error (RMSE) is similar to NMAE and thus not replicated here. The correlation coefficient is another commonly used measure that quantifies the covariability between the estimate and reference. The multiplicative error model, on the other hand, assumes a multiplicative error structure (or lognormal distribution in rain rates). As suggested by Tian et al. (2013, 2016), this model should in principle provide a fairer characterization of the actual errors, including a better separation of systematic and random errors and a variance that is constant with rain rate. The multiplicative error model assumes that x_i and y_i are related through

$$y_i = e^\alpha x_i^\beta e^{\varepsilon_i}, \quad (8)$$

where α and β represents the systematic errors and ε_i represents the bias-corrected random error characterized by a normal distribution of mean 0 and standard deviation σ . These parameters (α , β , and σ) can be obtained via an ordinary least squares fit after a logarithmic transformation [see Tian et al. (2013) for details]. A perfect α is 0, a perfect β is 1, and a perfect σ is 0. Variable α quantifies the “tilt” from the one-to-one line: for a perfect β , the deterministic part of Eq. (8) becomes $y = e^\alpha x$, with α controlling the gradient of the line. Variable β characterizes the nonlinear departure from the one-to-one line: for a perfect α , the deterministic part of Eq. (8) becomes $y = x^\beta$, with β being the exponent in the power-law relationship. In the logarithmic space, Eq. (8) is a straight line, with α being the y-axis intercept at $x = 1$ and β being the gradient of the line. Variable σ quantifies the stochastic component in Eq. (8), describing the spread of the points from the best fit curve of $y = e^\alpha x^\beta$.

3. Comparisons of raining events and rain rates

In this section, we compare three pairs of data, examining their performances as a function of 1) rain

TABLE 2. From top to bottom, MRMS rain rates averaged over the gauges box (green in Fig. 1; i.e., P_{MG}) and those averaged over the IMERG box (indigo in Fig. 1; i.e., P_{MI}), rain rates from the gauges (i.e., P_G) and MRMS rain rates averaged over the gauges (i.e., P_{MG}), and rain rates from the gauges (i.e., P_G) and from IMERG (i.e., P_I). For each pair of estimates, hits (top left), misses (top right), false alarms (bottom left), and correct negatives (bottom right) are shown.

	$P_{MI} \geq 0.1 \text{ mm h}^{-1}$	$P_{MI} < 0.1 \text{ mm h}^{-1}$
$P_{MG} \geq 0.1 \text{ mm h}^{-1}$	844 (7.9%)	53 (0.5%)
$P_{MG} < 0.1 \text{ mm h}^{-1}$	87 (0.8%)	9669 (90.8%)
	$P_{MG} \geq 0.1 \text{ mm h}^{-1}$	$P_{MG} < 0.1 \text{ mm h}^{-1}$
$P_G \geq 0.1 \text{ mm h}^{-1}$	613 (5.8%)	111 (1.0%)
$P_G < 0.1 \text{ mm h}^{-1}$	284 (2.7%)	9645 (90.5%)
	$P_I \geq 0.1 \text{ mm h}^{-1}$	$P_I < 0.1 \text{ mm h}^{-1}$
$P_G \geq 0.1 \text{ mm h}^{-1}$	436 (3.6%)	357 (2.9%)
$P_G < 0.1 \text{ mm h}^{-1}$	321 (2.6%)	11150 (90.9%)

occurrences, that is, if they agree on whether it is raining ($P \geq 0.1 \text{ mm h}^{-1}$) or not ($P < 0.1 \text{ mm h}^{-1}$), and 2) rain rates of the hits, that is, when both are raining, the degree to which the rain rates are similar. For brevity, we shall refer to this latter aspect as simply the rain rates. As the current MRMS product used has a lowest temporal resolution of 1 h, in this section we will average both IMERG and the gauges to this common 1-h denominator.

Looking at Fig. 1, one immediate concern is the degree to which uneven coverage of the gauges within the IMERG grid box (i.e., the difference between the green and indigo boxes) will contribute to a difference between the estimates from IMERG P_I and the average of the gauges P_G . Investigation of this impact is relevant as many gauge networks may not perfectly fit in and/or cover a given IMERG grid cell. Hence, to investigate this area coverage effect, we compare MRMS averaged over the IMERG grid box P_{MI} against MRMS averaged over the box containing the gauges P_{MG} . For rain occurrences, the MRMS estimates agree between the two areas more than 98% of the time (Table 2). The miss rate is 0.5% and the false alarm rate is 0.8%. Focusing on the hits, the parameters in the multiplicative error model indicate that there is practically no systematic error between the two MRMS estimates, with an α of ~ 0 and a β of ~ 1 (Table 3). This is expected because both are calculated from the same dataset and there are no marked geographical features that would introduce systematic bias between the two areas. More importantly, the random error is reasonably low, as evidenced by both the low σ (Table 3) and tight scatter in the rain rates (Fig. 2a). Note that the multiplicative error model fit (straight line) in Fig. 2a is remarkably close to the one-to-one line. Hence, the difference in area alone does not result in an

TABLE 3. NME, NMAE, CC, and the parameters for the multiplicative error model (α , β , and σ) for rain rates of the hits between the three pairs of estimates in Table 2.

	NME	NMAE	CC	α	β	σ
P_{MI}, P_{MG}	0.041	0.230	0.963	0.012	0.963	0.403
P_{MG}, P_G	0.033	0.376	0.896	0.108	0.778	0.611
P_I, P_G	0.103	0.982	0.349	0.273	0.339	1.141

appreciable difference in the rain estimates, both in rain occurrences and in rain rates.

Next, we compare MRMS averaged over the box with the gauges (i.e., P_{MG} ; green box in Fig. 1) to the gauges (i.e., P_G). Since it is expected that the ground-based MRMS will perform better than the satellite-based IMERG, this comparison will serve as a yardstick for evaluating IMERG with the gauges. MRMS agrees with the gauges 96.3% of the time (Table 2), a reassuring indicator that MRMS can identify rain and nonrain occurrences accurately. However, MRMS has more than twice the number of false alarms than misses, a skewed identification in rain occurrence that surfaces clearly when contrasting its POD to its FAR (Table 4). That is, the difference of the FAR from the perfect value (which is 0) is greater than the difference of the POD from the perfect value (which is 1). This is also evident in a BID that is considerably greater than 1, meaning an overestimation by MRMS in the number of rain occurrences. For the rain rates, while α is close to 0, indicating negligible offset in its distribution, β is less than 1, which reveals a reduction in the range of values that MRMS measures (Table 3). The combination of $\alpha > 0$ and $\beta < 1$ manifests in the scatter diagram as a slight overestimation of lower rain rates and a slight underestimation of higher rain rates (Fig. 2b). Nevertheless, the scatter remains well constrained, especially at rain rates above 1 mm h^{-1} ; this is also captured in its low σ

(Table 3). Therefore, MRMS is reasonably reliable in capturing rain and nonrain occurrences as well as the actual rain rates. It is interesting to note that the difference in dataset (P_{MG} and P_G) produces larger errors than the difference in area (P_{MI} and P_{MG}), except in the magnitude of NME (Table 3), an anomaly we will examine in detail in section 5.

Having established the insignificant effect of the areal difference and a set of scores by MRMS, we now proceed to evaluate IMERG (i.e., P_I) against the gauges (i.e., P_G). IMERG has significantly lower hits and higher misses than MRMS (Table 2); indeed, its POD is only slightly above 0.5 (Table 4). Intriguingly though, the percentage of false alarms is similar, possibly reflecting the propensity of MRMS to overestimate the number of raining events, as observed previously. Note that, despite this, the FAR is higher in IMERG than in MRMS because the IMERG FAR is normalized against a smaller number of hits. Interestingly, IMERG has a comparable number of misses as false alarms, which results in a BID of closer to 1. This means that IMERG, unlike MRMS, does not overestimate the number of rain occurrences. Despite the lower performance of IMERG in its POD and FAR as compared to MRMS, it does have an HSS of 0.533. Since HSS is a measure of the ability to identify rain occurrence compared to random chance ($HSS = 0$), this HSS value means that IMERG does possess reasonable ability in identifying rain or no rain. Focusing on the rain rates, the two most salient aspects of the comparison are the considerable scatter between IMERG and the gauges as well as the high bias at low rain rates and low bias at high rain rates (Fig. 2c). The first observation, which indicates the significantly higher degree of random error, is also corroborated by the large σ and NMAE (Table 3). Likewise, the bias, which denotes the systematic error in IMERG, is succinctly captured in the larger NME, as well as in the relatively high

FIG. 2. Scatter diagrams in logarithmic scale showing the rain rates of the hits between the three pairs of estimates in Table 2. The straight line shows the fit using the multiplicative error model parameters α and β (Table 2). The number of cases n is indicated in each diagram.

TABLE 4. POD, FAR, BID, and HSS for the three pairs of estimates, calculated from the contingency table (Table 2).

	POD	FAR	BID	HSS
P_{MI}, P_{MG}	0.941	0.093	1.038	0.916
P_{MG}, P_G	0.847	0.317	1.239	0.737
P_I, P_G	0.550	0.424	0.955	0.533

value of α , indicating a positive offset, and a β of significantly less than 1, indicating a reduced range of values. Similarly, it has a considerably lower CC. Therefore, contrasted against MRMS, we conclude that IMERG has lower hits and higher misses and possesses considerable scatter in the rain rates with an overestimation of drizzle and underestimation of heavy rainfall.

4. Categorization by source

The provision of ancillary variables within the IMERG dataset enables us to categorize how the IMERG estimate is derived. Specifically, we can identify each IMERG estimate as being derived from a particular PMW instrument, purely from morphing, purely from direct IR estimation, or from a mixture of morphing and direct IR estimation (see section 2a for details). However, since such a classification restricts the comparison to the time resolution of IMERG (i.e., half hour), we will use only the gauges and not MRMS. It should be noted that a half-hour MRMS product is currently in development for the GPM GV efforts that will permit a future extended analysis of this nature. As with section 3, we will evaluate IMERG on its ability to identify rain occurrence and rain rates of the hits.

Comparing IMERG to the gauges in each category, more than 90% of the comparisons in all sources except for IR only are correct negatives (Fig. 3). For hits, a majority of the sources fall between 2.6% and 3.1%, though AMSR has a higher hit rate of 3.7%, while IR only and, surprisingly, GMI have lower hit rates of 0.9% and 1.6%, respectively. On the other hand, for misses, the percentages are consistently between 2.3% and

3.7%, except for the IR-only category. In fact, this category has an abnormally high miss rate (19.8%). Indeed, with an HSS of only 0.040 (not shown), the use of direct IR estimation appears to confer only marginal benefits to the identification of rain occurrences. Fortunately, only 0.9% of the observations here use direct IR estimation. However, it is worth noting that the IR-only category occurs predominantly during winter months because of the lower performance of PMW retrievals, leading to larger errors during morphing, which favor direct IR estimation in the Kalman filter weights (G. Huffman 2015, personal communication). As for false alarms, it seems that estimates from morphing are prone to a higher number of false alarms than estimates from PMW satellite overpasses, implying that the current implementation of the morphing algorithm in IMERG is biased toward higher values in rain rates around the threshold of 0.1 mm h^{-1} . Comparing misses to false alarms, all categories except for the mixture of morphing and IR have more misses than false alarms, suggesting that the tendency for IMERG to underestimate rain occurrences is common to all PMW instruments. Therefore, in terms of identifying rain occurrences, estimates from PMW satellites perform best, though slightly susceptible to underestimating the number of rain occurrences. This is followed by morphing, which has an increased chance of false alarms. IR-only estimates, on the other hand, are remarkably poor in identifying rain occurrences.

Now, looking at the rain rates, the performance of each source varies widely. First of all, GMI has insufficient sample sizes for a robust conclusion, but in terms of the scatter it seems to have the best estimates, with almost all points close to the one-to-one line (Fig. 4). TMI and AMSR have relatively low random error as evident from their low σ values (Table 5) as well as their comparatively tight scatter. However, it is apparent that AMSR has a systematically high bias, with an error model fit that is consistently above the one-to-one line up to 10 mm h^{-1} . This is also well reflected in its anomalously high NME, as well as a high α while its β is close to 1. Together with its high hit percentage, this suggests that AMSR has an

FIG. 3. Hits, misses, false alarms, and correct negatives for each source. The numbers at the right indicate the total number of events.

FIG. 4. As in Fig. 2, but for each source.

underlying issue with its calibration. On the other hand, TMI seems to have a systematic bias that is characteristic of IMERG as a whole: too heavy at low rain rates and too light at high rain rates. This behavior is seen in the other PMW platforms—SSMIS and MHS—and the two categories with morphing. These four categories also have the highest random errors of all sources, with σ values exceeding 1 and NMAEs above 0.9. For SSMIS and MHS, this large scatter may be a symptom of their low resolutions (mean footprint sizes of 14 and 17 km, in contrast to the ~ 6 km of other instruments). Furthermore, since morphing interpolates between PMW estimates and

hence inherits errors from those estimates, it is no surprise that morphing possess similar systematic and random errors as SSMIS and MHS, especially since they are the most numerous of all satellite overpasses (77% of the time). In fact, the similarity in their numbers in Table 5 raises the possibility that the comparatively high errors in the morphing estimates are not a consequence of the morphing technique itself; that is, an improvement to the PMW estimates may lead to a corresponding improvement in the estimates from morphing. In summary of the rain rates, GMI appears to perform best. This is followed by AMSR, which has a low random error but a persistent

TABLE 5. Number of samples (i.e., n), NME, NMAE, CC, and the parameters for the multiplicative error model (α , β , and σ) for each source.

	n	NME	NMAE	CC	α	β	σ
IR only	2	-0.620	0.620	—	-0.511	0.000	0.000
IR + morph	165	0.007	1.085	0.293	0.316	0.246	1.210
Morph only	331	0.124	0.967	0.323	0.410	0.373	1.079
TMI	12	0.198	0.990	0.337	0.816	0.148	0.806
AMSR	29	1.492	1.771	0.498	0.854	0.680	0.939
SSMIS	70	0.150	1.193	0.330	0.481	0.300	1.177
MHS	93	-0.019	0.988	0.104	0.368	0.195	1.086
GMI	7	-0.110	0.438	0.761	-0.425	0.826	0.974

high bias that is suggestive of calibration issues. TMI has similarly low random error but a systematic error of an overestimation of drizzle and an underestimation of heavy rainfall. All the other sources share this systematic error and suffer high random errors.

The preceding results show that, by separating the errors into different sources, we are better able to pick apart the random and systematic errors that contribute to those seen in section 3. Here, we present an alternative perspective of the errors by showing, in a violin diagram, the distributions of the differences between the rain rates ($P_I - P_G$) for cases where at least one of the pair is $\geq 0.1 \text{ mm h}^{-1}$ (Fig. 5). It is important to note that this threshold criterion is different from that of Fig. 4; here we consider hits, misses, and false alarms, whereas there we considered only the hits. The approach of taking the difference assumes additive errors, which will not be representative of the actual error structure of rainfall. Nonetheless, bearing this in mind, the violin diagram can provide additional insights: the directions of bias are conspicuous and outliers are emphasized. From Fig. 5, the previously observed high bias in AMSR is evident, with the distribution and box-and-whisker both skewed toward the positive values. In addition, we can see that the IR-only category has a strong tendency to contribute to a low-biased estimation, which stems from the large number of misses (Fig. 3). Another striking observation in Fig. 5 is the vulnerability of SSMIS to outliers. There are at least two points in the violin diagram of SSMIS that stand out, especially when compared to MHS, which has comparable values in σ and NMAE. In particular, one of these outliers involves a remarkable underestimation of rain rates by nearly 50 mm h^{-1} (this is the rightmost point in Fig. 4); we shall further investigate this particular case.

This event of an underestimation of about 50 mm h^{-1} by SSMIS occurred at 1030–1100 UTC 16 July 2014, with the overpass occurring at 1032 UTC. During the summer season of this region, the dominant rainfall system types are often highly cellular short lines of

FIG. 5. Violin diagrams for IMERG minus the gauges for cases where at least one of the pair is $\geq 0.1 \text{ mm h}^{-1}$ for each platform, showing the distribution (shaded gray areas), median (red lines), interquartile range (boxes), 10th and 90th percentiles (whiskers), and outliers beyond the whiskers (blue crosses). Numbers at the top indicate the number of samples in each category.

airmass convection (Tokay et al. 2014), which can lead to strong but brief downpours. To better understand the precise rain system during this overpass, we compare the large-scale rain fields between IMERG and MRMS (Fig. 6). For IMERG, we use the half-hour data; for MRMS, we use the full hour. MRMS shows a small but intense convective cell passing over the gauges, with rain rates exceeding 60 mm h^{-1} . This is consistent with the accumulation from the gauges, which recorded a mean rain rate of 57.5 mm h^{-1} for this half hour, with one gauge reaching 83.8 mm h^{-1} . Events of such magnitude may result in localized urban flooding, so it is important for hazard monitoring that IMERG can identify such rain systems. Vertical cross section from MRMS reflectivity mosaic reveals a shallow system, with 20 dBZ echo tops reaching only 7 km above mean sea level (<http://mrms.ou.edu/>; Fig. 1 in the supplemental material). This cell is significantly weaker in the IMERG rain field by an order of magnitude (Fig. 6). The failure to pick up this cell can be traced to HQprecipitation, a variable in the IMERG files that gives the PMW estimate after regridding to the 0.1° grids and cross calibration between all instruments but prior to gauge adjustment. This error, in turn, originated from the Goddard profiling algorithm (GPROF; Kummerow et al. 2001, 2015), an instantaneous “Level 2” rainfall estimate from the satellite platform that constitutes the input into IMERG. To eliminate the possibility that this error emanated from GPROF itself, we confirmed that

FIG. 6. Case study of an event at 1030–1100 UTC 16 Jul 2016, with an SSMIS overpass at 1032 UTC, showing the large-scale rain field between MRMS, IMERG, microwave-based estimate in IMERG (HQprecipitation), and the instantaneous satellite estimate (GPROF). The gray box marks the IMERG grid box in our study (indigo box in Fig. 1). Note that the color scale is logarithmic.

the GSMaP (Kubota et al. 2007; Kachi et al. 2014), a similar rainfall product based on a different algorithm but that also uses SSMIS for this event, failed to capture the cell as well (Fig. 2 in the supplemental material). Moreover, on top of underestimating the rain rate during this half hour, the fact that IMERG employs morphing between PMW estimates meant that this error is propagated forward and backward with the cell motion (Fig. 3 in the supplemental material). Hence, by categorizing the IMERG estimates according to the source, we are able to easily trace the cause of an error down to a specific instrument and platform, and thereby narrow the processes that are likely responsible for the significant underestimation.

5. Discussion

In section 4, we traced a case of severe underestimation in IMERG to the failure of SSMIS in capturing a shallow but intense cell. This case study demonstrates how errors in rainfall retrieval from satellite orbits permeate into the gridded product. In addition, these errors will linger in the IMERG field before and after the overpass because of the use of morphing. Therefore, the framework of classifying the errors to different sources allows us to trace the errors back to the platforms and/or algorithm. Our approach illustrates a viable strategy to bridge between gridded products (denoted Level 3 in the GPM project) and individual orbit products (denoted Level 2 in the GPM project). This provides a more precise diagnosis in the errors of satellite-based precipitation products, thus making ground validation efforts more effective at contributing to the improved

development of the products. We will follow up on this study and employ this bridging approach to larger regions using MRMS once the half-hour data are available.

In analyzing the rain rates, we presented measures that assume an additive error structure (NME, NMAE, and CC) and measures that assume a multiplicative error structure (α , β , and σ). Theoretically, the latter assumption is more valid, a point we shall attempt to demonstrate here. Comparing between P_{MI} versus P_{MG} and P_{MG} versus P_G , the line fits (which is based on the parameters from the multiplicative error model; Table 2) and the scatter points (Fig. 2) both point to a larger systematic error for the latter, that is, between MRMS and the gauges. Yet, its NME is lower. This is likely a result of treating rainfall values linearly as opposed to logarithmically, thus overemphasizing outliers. Moreover, the NME is a single number that yields less information than the multiplicative error model; for example, the fact that both P_{MG} versus P_G and P_I versus P_G overestimate low rain rates and underestimate high rain rates is in no way apparent in their NME values. Hence, NME is limited in its representation of systematic error. In addition, the multiplicative error model explicitly distinguishes between systematic and random errors, as opposed to measures such as CC, which is a function of both types of errors (Tian et al. 2016). Potential limitations of the multiplicative error model include the difficulty in intuitively comprehending what α and β represent, the viability of using ordinary least squares in fitting the log-transformed rain rates (Maggioni et al. 2014), and possible difficulty in capturing the rain rates at the extreme ends of the rain-rate spectrum, but these can be overcome by

wider adoption of and further examination on this approach. Therefore, our results agree with [Tian et al. \(2016\)](#) that the multiplicative error model is more advantageous than NME, NMAE, or any other additive or linear measures in characterizing rainfall errors.

Regardless of the framework used, it is, however, clear that, for the case of IMERG, there are different systematic and random errors associated with each source. While the random errors, that is, widespread in the scatter, may give the impression that IMERG is very inaccurate ([Figs. 2, 4](#)), the fact that these errors are random means that they will diminish when averaging upscale. Indeed, the rationale for providing the estimates at such a high resolution is to give users the flexibility to average over domains that suit their purposes, for example, daily precipitation of a 1° climate model or hourly accumulations over a large river basin. As such, for these applications, the random errors are likely to be lower. On the other hand, systematic errors are more pressing as they cannot be removed by statistical methods available to the user. The gauge adjustment, intended to rein in systematic errors, may not be sufficiently resolved to address biases at the pixel level, as it occurs at a monthly time scale over 1° grids. Given the varying performance of each platform, it may be worth investigating whether this gauge adjustment can be applied independently to each source. Alternatively, improvements to cross calibration when ingesting estimates from various platforms may increase the efficacy of gauge adjustment by reducing compensating errors.

6. Conclusions

IMERG is a next-generation product from the GPM family that builds on the TRMM multisatellite product era and is expected to be widely used for a range of applications from understanding rainfall patterns to hydrological modeling to hazard monitoring. In this study, we provided an evaluation of Day-1 IMERG at the pixel level against a well-maintained network of dual tipping-bucket gauges. Existing studies of this nature on other gridded products focus on a direct comparison between the final estimates and the ground reference (e.g., [Tian and Peters-Lidard 2010](#); [Tang et al. 2015](#)). Moving beyond such direct comparisons, we introduced a novel approach in bridging the ground validation process to the instantaneous rainfall estimates of the GPM constellation satellites via the gridded product.

Compared to MRMS, we found that IMERG has fewer hits and more misses and possesses considerable scatter in the rain rates with an overestimation of drizzle and underestimation of heavy rainfall. When divided according to their sources, we discovered that estimates

from PMW platforms are in general superior in identifying rain occurrences, though their rain rates show a range of systematic and random errors. Errors in the estimates from morphing are comparable to those from PMW as a whole, but errors from IR-only estimates are poor. This demonstrates the benefits of harnessing the sources of the IMERG estimates to expose the causes of the errors. We further illustrated the utility of categorizing the comparisons to the sources by focusing on a particular case in which we traced a severe underestimation to the failure of SSMIS in capturing a shallow but intense cell. Therefore, we advocate the use of ancillary information when performing ground validation so as to illuminate the precise nature of the errors. Furthermore, this method of tracing back to the passive microwave platforms offers a convenient approach to evaluate instantaneous satellite estimates from the entire GPM constellation.

For users of IMERG, the identification of the source of the estimates provides an indication of how reliable the estimate is. For example, based on our results, estimates from GMI are generally reliable, whereas estimates from AMSR may exhibit a positive bias, while estimates from SSMIS have larger random errors. Depending on the purpose IMERG is used for, our results may allow users to make better informed decisions. For developers of IMERG, the hierarchy of performance between different PMW instruments raises the possibility of a ranking order for the PMW sources. Currently, when there are two or more overpasses over a grid box at a time step, IMERG prioritizes imagers over sounders, followed by the estimate closest to the center of the half hour. In light of our results, the feasibility of using the relative performance of each instrument should be considered.

In our analysis, we adopted the approach proposed in [Tang et al. \(2015\)](#) of investigating first whether IMERG is able to correctly identify rain occurrences, and then how well IMERG is able to estimate the correct rain rate of the hits. This gives a better understanding of the nature of the errors, that is, if it is a detection error or a magnitude error. Furthermore, in evaluating the estimation of rain rates, we presented measures that assume additive error structures and multiplicative error structures ([Tian et al. 2013](#)). We showed that the latter is better able to and can more completely characterize the errors in the comparison. Hence, we urge for a wider adoption of the two-step ground validation approach and the use of multiplicative error model.

One shortcoming of our study is the use of only one IMERG grid box because of the location of the gauge network. This is a major factor for the low sample size in some of our analysis, thus limiting the robustness of our

conclusions and preventing further categorization such as by season. Furthermore, our results are only applicable to coastal areas with similar climate. A follow-on study will employ MRMS over a wider region to increase the sample size and examine it over different climatic conditions and surface types as well as investigate the impact of gauge adjustment, thus verifying or refining the conclusions herein. This will provide a comprehensive understanding of the uncertainties in IMERG and the various instruments in orbit, allowing precise refinement of the IMERG estimates and ultimately contributing to the advancement of precipitation science and the benefit for society.

Acknowledgments. We thank George Huffman and David Bolvin for informative discussions on IMERG and Yudong Tian for instructive consultation on the multiplicative error model. The gauge data are maintained by the NASA Wallops GPM GV Team, and we acknowledge David Wolff for his assistance with the data. The MRMS data were processed for the GPM GV Program by Pierre-Emmanuel Kirstetter, and we appreciate the further assistance provided by Jianxin Wang. We also thank two anonymous reviewers for their comments and suggestions. J.T. is supported by an appointment to the NASA Postdoctoral Program at Goddard Space Flight Center, administered by Universities Space Research Association through a contract with NASA. W.A.P. and A.T. acknowledge support from the GPM Mission (Project Scientist, Gail S. Jackson, and GV Systems Manager, Mathew Schwaller) and also PMM Science Team funding provided by Dr. Ramesh Kakar. The IMERG data were provided by the NASA Goddard Space Flight Center's PMM and PPS teams, which develop and compute the IMERG as a contribution to GPM, and archived at the NASA GES DISC. All codes used in this analysis are freely available at https://github.com/JacksonTanBS/2016_Tan-et-al._JHM.

REFERENCES

- AghaKouchak, A., A. Mehran, H. Norouzi, and A. Behrangi, 2012: Systematic and random error components in satellite precipitation data sets. *Geophys. Res. Lett.*, **39**, L09406, doi:[10.1029/2012GL051592](https://doi.org/10.1029/2012GL051592).
- Bharti, V., and C. Singh, 2015: Evaluation of error in TRMM 3B42V7 precipitation estimates over the Himalayan region. *J. Geophys. Res. Atmos.*, **120**, 12 458–12 473, doi:[10.1002/2015JD023779](https://doi.org/10.1002/2015JD023779).
- Chen, S., and Coauthors, 2013a: Similarity and difference of the two successive V6 and V7 TRMM Multisatellite Precipitation Analysis performance over China. *J. Geophys. Res. Atmos.*, **118**, 13 060–13 074, doi:[10.1002/2013JD019964](https://doi.org/10.1002/2013JD019964).
- , and Coauthors, 2013b: Evaluation of the successive V6 and V7 TRMM Multisatellite Precipitation Analysis over the continental United States. *Water Resour. Res.*, **49**, 8174–8186, doi:[10.1002/2012WR012795](https://doi.org/10.1002/2012WR012795).
- Ebert, E. E., J. E. Janowiak, and C. Kidd, 2007: Comparison of near-real-time precipitation estimates from satellite observations and numerical models. *Bull. Amer. Meteor. Soc.*, **88**, 47–64, doi:[10.1175/BAMS-88-1-47](https://doi.org/10.1175/BAMS-88-1-47).
- Gottschalk, J., J. Meng, M. Rodell, and P. Houser, 2005: Analysis of multiple precipitation products and preliminary assessment of their impact on Global Land Data Assimilation System land surface states. *J. Hydrometeorol.*, **6**, 573–598, doi:[10.1175/JHM437.1](https://doi.org/10.1175/JHM437.1).
- Gourley, J. J., Y. Hong, Z. L. Flamig, L. Li, and J. Wang, 2010: Intercomparison of rainfall estimates from radar, satellite, gauge, and combinations for a season of record rainfall. *J. Appl. Meteor. Climatol.*, **49**, 437–452, doi:[10.1175/2009JAMC2302.1](https://doi.org/10.1175/2009JAMC2302.1).
- Guo, H., S. Chen, A. Bao, A. Behrangi, Y. Hong, F. Ndayisaba, J. Hu, and P. M. Stepanian, 2016: Early assessment of integrated multi-satellite retrievals for global precipitation measurement over China. *Atmos. Res.*, **176–177**, 121–133, doi:[10.1016/j.atmosres.2016.02.020](https://doi.org/10.1016/j.atmosres.2016.02.020).
- Habib, E., A. Henschke, and R. F. Adler, 2009: Evaluation of TMPA satellite-based research and real-time rainfall estimates during six tropical-related heavy rainfall events over Louisiana, USA. *Atmos. Res.*, **94**, 373–388, doi:[10.1016/j.atmosres.2009.06.015](https://doi.org/10.1016/j.atmosres.2009.06.015).
- , A. T. Haile, Y. Tian, and R. J. Joyce, 2012: Evaluation of the high-resolution CMORPH satellite rainfall product using dense rain gauge observations and radar-based estimates. *J. Hydrometeorol.*, **13**, 1784–1798, doi:[10.1175/JHM-D-12-017.1](https://doi.org/10.1175/JHM-D-12-017.1).
- Hong, Y., K.-L. Hsu, S. Sorooshian, and X. Gao, 2004: Precipitation Estimation from Remotely Sensed Imagery Using an Artificial Neural Network Cloud Classification System. *J. Appl. Meteor.*, **43**, 1834–1853, doi:[10.1175/JAM2173.1](https://doi.org/10.1175/JAM2173.1).
- Hossain, F., and G. J. Huffman, 2008: Investigating error metrics for satellite rainfall data at hydrologically relevant scales. *J. Hydrometeorol.*, **9**, 563–575, doi:[10.1175/2007JHM925.1](https://doi.org/10.1175/2007JHM925.1).
- Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement Mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, doi:[10.1175/BAMS-D-13-00164.1](https://doi.org/10.1175/BAMS-D-13-00164.1).
- Huffman, G. J., and Coauthors, 2007: The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, **8**, 38–55, doi:[10.1175/JHM560.1](https://doi.org/10.1175/JHM560.1).
- , D. T. Bolvin, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie, 2015: NASA Global Precipitation Measurement Integrated Multi-satellite Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 4.5, 30 pp. [Available online at http://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf.]
- Joyce, R. J., and P. Xie, 2011: Kalman filter-based CMORPH. *J. Hydrometeorol.*, **12**, 1547–1563, doi:[10.1175/JHM-D-11-022.1](https://doi.org/10.1175/JHM-D-11-022.1).
- , J. E. Janowiak, P. A. Arkin, and P. Xie, 2004: CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, **5**, 487–503, doi:[10.1175/1525-7541\(2004\)005<0487:CAMTPG>2.0.CO;2](https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2).
- Kachi, M., T. Kubota, K. Aonashi, T. Ushio, S. Shige, M. Yamamoto, T. Mega, and R. Oki, 2014: Recent improvements in the Global Satellite Mapping of Precipitation (GSMaP). *2014 Int. IEEE Geoscience and Remote Sensing Symposium*, Quebec City, QC, Canada, IEEE, 3762–3765, doi:[10.1109/IGARSS.2014.6947302](https://doi.org/10.1109/IGARSS.2014.6947302).
- Kirstetter, P.-E., and Coauthors, 2012: Toward a framework for systematic error modeling of spaceborne precipitation radar

- with NOAA/NSSL ground radar-based National Mosaic QPE. *J. Hydrometeorol.*, **13**, 1285–1300, doi:[10.1175/JHM-D-11-0139.1](https://doi.org/10.1175/JHM-D-11-0139.1).
- Krajewski, W. F., G. J. Ciach, J. R. McCollum, and C. Bacotiu, 2000: Initial validation of the Global Precipitation Climatology Project monthly rainfall over the United States. *J. Appl. Meteor.*, **39**, 1071–1086, doi:[10.1175/1520-0450\(2000\)039<1071:IVOTGP>2.0.CO;2](https://doi.org/10.1175/1520-0450(2000)039<1071:IVOTGP>2.0.CO;2).
- Kubota, T., and Coauthors, 2007: Global precipitation map using satellite-borne microwave radiometers by the GSMaP project: Production and validation. *IEEE Trans. Geosci. Remote Sens.*, **45**, 2259–2275, doi:[10.1109/TGRS.2007.895337](https://doi.org/10.1109/TGRS.2007.895337).
- , T. Ushio, S. Shige, S. Kida, M. Kachi, and K. Okamoto, 2009: Verification of high-resolution satellite-based rainfall estimates around Japan using a gauge-calibrated ground-radar dataset. *J. Meteor. Soc. Japan*, **87A**, 203–222, doi:[10.2151/jmsj.87A.203](https://doi.org/10.2151/jmsj.87A.203).
- Kummerow, C., and Coauthors, 2001: The evolution of the Goddard profiling algorithm (GPROF) for rainfall estimation from passive microwave sensors. *J. Appl. Meteor.*, **40**, 1801–1820, doi:[10.1175/1520-0450\(2001\)040<1801:TEOTGP>2.0.CO;2](https://doi.org/10.1175/1520-0450(2001)040<1801:TEOTGP>2.0.CO;2).
- , D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic, 2015: The evolution of the Goddard profiling algorithm to a fully parametric scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280, doi:[10.1175/JTECH-D-15-0039.1](https://doi.org/10.1175/JTECH-D-15-0039.1).
- Li, Z., D. Yang, and Y. Hong, 2013: Multi-scale evaluation of high-resolution multi-sensor blended global precipitation products over the Yangtze River. *J. Hydrol.*, **500**, 157–169, doi:[10.1016/j.jhydrol.2013.07.023](https://doi.org/10.1016/j.jhydrol.2013.07.023).
- Liu, Z., 2016: Comparison of Integrated Multisatellite Retrievals for GPM (IMERG) and TRMM Multisatellite Precipitation Analysis (TMPA) monthly precipitation products: Initial results. *J. Hydrometeorol.*, **17**, 777–790, doi:[10.1175/JHM-D-15-0068.1](https://doi.org/10.1175/JHM-D-15-0068.1).
- Maggioni, V., M. R. P. Sapiano, R. F. Adler, Y. Tian, and G. J. Huffman, 2014: An error model for uncertainty quantification in high-time-resolution precipitation products. *J. Hydrometeorol.*, **15**, 1274–1292, doi:[10.1175/JHM-D-13-0112.1](https://doi.org/10.1175/JHM-D-13-0112.1).
- Mei, Y., E. N. Anagnostou, E. I. Nikolopoulos, and M. Borga, 2014: Error analysis of satellite precipitation products in mountainous basins. *J. Hydrometeorol.*, **15**, 1778–1793, doi:[10.1175/JHM-D-13-0194.1](https://doi.org/10.1175/JHM-D-13-0194.1).
- NASA PPS, 2016: Precipitation Processing System Global Precipitation Measurement: File specification for GPM products, version 4.01 TKIO 3.70.8, 1381 pp. [Available online at <ftp://gpmweb2.pps.eosdis.nasa.gov/pub/GPMfilespec/filespec.GPM.V1.pdf>.]
- Roca, R., P. Chambon, I. Jobard, P.-E. Kirstetter, M. Gosset, and J. C. Bergès, 2010: Comparing satellite and surface rainfall products over West Africa at meteorologically relevant scales during the AMMA campaign using error estimates. *J. Appl. Meteor. Climatol.*, **49**, 715–731, doi:[10.1175/2009JAMC2318.1](https://doi.org/10.1175/2009JAMC2318.1).
- Sapiano, M. R. P., 2010: An evaluation of high resolution precipitation products at low resolution. *Int. J. Climatol.*, **30**, 1416–1422, doi:[10.1002/joc.1961](https://doi.org/10.1002/joc.1961).
- , and P. A. Arkin, 2009: An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *J. Hydrometeorol.*, **10**, 149–166, doi:[10.1175/2008JHM1052.1](https://doi.org/10.1175/2008JHM1052.1).
- Stampoulis, D., and E. N. Anagnostou, 2012: Evaluation of global satellite rainfall products over continental Europe. *J. Hydrometeorol.*, **13**, 588–603, doi:[10.1175/JHM-D-11-086.1](https://doi.org/10.1175/JHM-D-11-086.1).
- Tang, G., Y. Ma, D. Long, L. Zhong, and Y. Hong, 2016a: Evaluation of GPM Day-1 IMERG and TMPA version-7 legacy products over mainland China at multiple spatiotemporal scales. *J. Hydrol.*, **533**, 152–167, doi:[10.1016/j.jhydrol.2015.12.008](https://doi.org/10.1016/j.jhydrol.2015.12.008).
- , Z. Zeng, D. Long, X. Guo, B. Yong, W. Zhang, and Y. Hong, 2016b: Statistical and hydrological comparisons between TRMM and GPM level-3 products over a midlatitude basin: Is day-1 IMERG a good successor for TMPA 3B42V7? *J. Hydrometeorol.*, **17**, 121–137, doi:[10.1175/JHM-D-15-0059.1](https://doi.org/10.1175/JHM-D-15-0059.1).
- Tang, L., Y. Tian, F. Yan, and E. Habib, 2015: An improved procedure for the validation of satellite-based precipitation estimates. *Atmos. Res.*, **163**, 61–73, doi:[10.1016/j.atmosres.2014.12.016](https://doi.org/10.1016/j.atmosres.2014.12.016).
- Tian, Y., and C. D. Peters-Lidard, 2007: Systematic anomalies over inland water bodies in satellite-based precipitation estimates. *Geophys. Res. Lett.*, **34**, L14403, doi:[10.1029/2007GL030787](https://doi.org/10.1029/2007GL030787).
- , and —, 2010: A global map of uncertainties in satellite-based precipitation measurements. *Geophys. Res. Lett.*, **37**, L24407, doi:[10.1029/2010GL046008](https://doi.org/10.1029/2010GL046008).
- , —, B. J. Choudhury, and M. Garcia, 2007: Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *J. Hydrometeorol.*, **8**, 1165–1183, doi:[10.1175/2007JHM859.1](https://doi.org/10.1175/2007JHM859.1).
- , and Coauthors, 2009: Component analysis of errors in satellite-based precipitation estimates. *J. Geophys. Res.*, **114**, D24101, doi:[10.1029/2009JD011949](https://doi.org/10.1029/2009JD011949).
- , C. D. Peters-Lidard, R. F. Adler, T. Kubota, and T. Ushio, 2010: Evaluation of GSMaP precipitation estimates over the contiguous United States. *J. Hydrometeorol.*, **11**, 566–574, doi:[10.1175/2009JHM1190.1](https://doi.org/10.1175/2009JHM1190.1).
- , G. J. Huffman, R. F. Adler, L. Tang, M. Sapiano, V. Maggioni, and H. Wu, 2013: Modeling errors in daily precipitation measurements: Additive or multiplicative? *Geophys. Res. Lett.*, **40**, 2060–2065, doi:[10.1002/grl.50320](https://doi.org/10.1002/grl.50320).
- , G. S. Nearing, C. D. Peters-Lidard, K. W. Harrison, and L. Tang, 2016: Performance metrics, error modeling, and uncertainty quantification. *Mon. Wea. Rev.*, **144**, 607–613, doi:[10.1175/MWR-D-15-0087.1](https://doi.org/10.1175/MWR-D-15-0087.1).
- Tokay, A., R. J. Roche, and P. G. Bashor, 2014: An experimental study of spatial variability of rainfall. *J. Hydrometeorol.*, **15**, 801–812, doi:[10.1175/JHM-D-13-031.1](https://doi.org/10.1175/JHM-D-13-031.1).
- Turk, F. J., and S. D. Miller, 2005: Toward improved characterization of remotely sensed precipitation regimes with MODIS/AMSR-E blended data techniques. *IEEE Trans. Geosci. Remote Sens.*, **43**, 1059–1069, doi:[10.1109/TGRS.2004.841627](https://doi.org/10.1109/TGRS.2004.841627).
- Villarini, G., and W. F. Krajewski, 2007: Evaluation of the research version TMPA three-hourly $0.25^\circ \times 0.25^\circ$ rainfall estimates over Oklahoma. *Geophys. Res. Lett.*, **34**, L05402, doi:[10.1029/2006GL029147](https://doi.org/10.1029/2006GL029147).
- , P. V. Mandapaka, W. F. Krajewski, and R. J. Moore, 2008: Rainfall and sampling uncertainties: A rain gauge perspective. *J. Geophys. Res.*, **113**, D11102, doi:[10.1029/2007JD009214](https://doi.org/10.1029/2007JD009214).
- Xue, X., Y. Hong, A. S. Limaye, J. J. Gourley, G. J. Huffman, S. I. Khan, C. Dorji, and S. Chen, 2013: Statistical and hydrological evaluation of TRMM-based Multi-satellite Precipitation Analysis over the Wangchu basin of Bhutan: Are the latest satellite precipitation products 3B42V7 ready for use in ungauged basins? *J. Hydrol.*, **499**, 91–99, doi:[10.1016/j.jhydrol.2013.06.042](https://doi.org/10.1016/j.jhydrol.2013.06.042).
- Yong, B., and Coauthors, 2014: Intercomparison of the version-6 and version-7 TMPA precipitation products over high and low latitudes basins with independent gauge networks: Is the

- newer version better in both real-time and post-real-time analysis for water resources and hydrologic extremes? *J. Hydrol.*, **508**, 77–87, doi:[10.1016/j.jhydrol.2013.10.050](https://doi.org/10.1016/j.jhydrol.2013.10.050).
- , D. Liu, J. J. Gourley, Y. Tian, G. J. Huffman, L. Ren, and Y. Hong, 2015: Global view of real-time TRMM Multisatellite Precipitation Analysis: Implications for its successor Global Precipitation Measurement Mission. *Bull. Amer. Meteor. Soc.*, **96**, 283–296, doi:[10.1175/BAMS-D-14-00017.1](https://doi.org/10.1175/BAMS-D-14-00017.1).
- Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) System: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, doi:[10.1175/2011BAMS-D-11-00047.1](https://doi.org/10.1175/2011BAMS-D-11-00047.1).
- , Y. Qi, K. Howard, C. Langston, and B. Kaney, 2012: Radar quality index (RQI)—A combined measure of beam blockage and VPR effects in a national network. *IAHS Publ.*, **351**, 388–393.