Journal of Geophysical Research: Atmospheres

RESEARCH ARTICLE
10.1029/2018JD028584

Investigating the Potential of Using Satellite-Based Precipitation Radars as Reference for Evaluating Multisatellite Merged Products

S. Khan[1], V. Maggioni[1], and P.-E. Kirstetter[2,3,4]

[1]Department of Civil, Environmental, and Infrastructure Engineering, George Mason University, Fairfax, VA, USA, [2]Advanced Radar Research Center, University of Oklahoma, Norman, OK, USA, [3]NOAA/National Severe Storms Laboratory, Norman, OK, USA, [4]School of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, USA

Key Points:
- Evaluating Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG) against satellite- and ground-based references
- Dual-frequency Precipitation Radar (2ADPR) is a viable reference to evaluate random errors associated with the infrared component of IMERG
- Climatology influences performance of infrared precipitation estimates when compared to 2ADPR and Multi-Radar/Multi-Sensor reference

Correspondence to:
S. Khan,
skhan51@gmu.edu

Citation:
Khan, S., Maggioni, V., & Kirstetter, P.-E. (2018). Investigating the potential of using satellite-based precipitation radars as reference for evaluating multisatellite merged products. *Journal of Geophysical Research: Atmospheres*, *123*, 8646–8660. https://doi.org/10.1029/2018JD028584

Received 26 FEB 2018
Accepted 28 JUL 2018
Accepted article online 5 AUG 2018
Published online 27 AUG 2018

Abstract This work investigates the viability of using the Global Precipitation Mission (GPM) Dual-frequency Precipitation Radar (DPR) as a reference for evaluating multisatellite precipitation products in locations where a ground-based reference is not available. The Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG) version V05 products (early, late, and final) and the microwave-only (MW) and infrared-only (IR) components are evaluated against two reference data sets, derived from the Multi-Radar/Multi-Sensor System suite of products (MRMS) and the level-2 DPR (2ADPR). The analysis focuses on a 2-year period (2014–2015), excluding winters to evaluate liquid-phase precipitation only. Systematic and random errors are assessed between the satellite-based products and the MRMS-based reference. Systematic error for the IMERG early, late, and final products and the MW component are shown to be comparable to those of 2ADPR, while the IR component exhibits a larger bias. Random errors of the IR estimates are found to be 1 order of magnitude larger than the 2ADPR random error. Thus, the hypothesis of using 2ADPR as a benchmark for quantifying uncertainties associated with IMERG holds true for evaluating random errors associated with IR precipitation estimates. Errors in the IR component are then investigated as a function of climatology and seasonality across contiguous United States. The temperate, oceanic, and subtropical climates show the best performance in terms of probability of detection and success ratio, whereas arid-desert manifests the smallest root-mean-square errors. This work suggests that climatic zone-specific error characterization model is necessary to estimate uncertainties associated with the IMERG products.

1. Introduction

The accurate global mapping of precipitation plays a pivotal role across numerous applications, including natural hazard mitigation, terrestrial hydrology, climate change studies, agriculture practices, and vector-borne disease monitoring (Hong et al., 2007; Hossain & Anagnostou, 2004; Johnson et al., 1993; Krajewski et al., 2006; Mousam et al., 2016). Ungauged regions around the world (like mountainous areas and oceans) depend entirely on satellite measurements to retrieve precipitation information (Kidd et al., 2017), which is why over the past few decades satellite precipitation products have gained popularity. One example is the wide usage of the Tropical Rainfall Measuring Mission (TRMM)-based precipitation products in scientific investigations (Adler et al., 2009; Curtis et al., 2007; Houze et al., 2015) and decision-making activities (Kirschbaum et al., 2016).

The most recent effort in satellite precipitation is the Global Precipitation Measurement (GPM) mission (Hou et al., 2014). The GPM core satellite was launched on 24 February 2014 and carries an advanced radar (Dual-frequency Precipitation Radar (DPR)) and a radiometer system (GPM Microwave Imager) that provides a benchmark to unify precipitation measurements from a constellation of satellites. Active remote sensing such as DPR provides indirect but range-resolved observations of precipitation. Compared to passive remote sensing such as passive microwave and infrared sensors, it enables more accurate estimation of precipitation and improved estimates of the hydrometeor size and a better classification of the precipitation phase (liquid, frozen, and mixed-phase) when compared to its predecessor, the TRMM Precipitation Radar (PR; Kojima et al., 2012; Prakash et al., 2016).

One of the main GPM-based precipitation products is the high-resolution Integrated Multi-satellitE Retrievals for GPM (IMERG; Huffman et al., 2014). The IMERG algorithm fuses microwave and infrared retrievals from

several sensors providing indirect observations of precipitation and merges them with ground-based rain gauge observations. Thus, IMERG exploits information from different sources in attempt to provide the best possible high-resolution precipitation estimation.

The successful operational use of these precipitation estimates is still limited by the quantification of their associated uncertainties and errors (Falck et al., 2015; Kirstetter et al., 2014; Maggioni et al., 2012; Stephens & Kummerow, 2007). The estimation of these errors is particularly challenging because of the complex definition of satellite retrieval uncertainty, which is a combination of measurement, algorithm, and sampling errors, as well as the diverse nature of measurement sources. Over the past decade, a few efforts attempted to provide error estimates with satellite-based products, such as the errors that are routinely produced for the global merged GPCP (Global Precipitation Climatology Project) analysis (Adler et al., 2003; Huffman, 1997). These errors are provided for every 2.5° grid at monthly and daily scale, based on the method developed by Huffman (1997). Uncertainty estimates, in terms of random error only, are also available for the TRMM Multi-satellite Precipitation Analysis (TMPA) analyses (Huffman et al., 2007).

Systematic and random errors provide valuable information in terms of accuracy and precision of precipitation retrievals. The characterization of systematic and random errors is fundamental for bias adjustment, error modeling, and improvements of precipitation retrieval algorithms (Sorooshian et al., 2011). Several research groups have directed their efforts to assess satellite precipitation errors (Anagnostou et al., 2010; Sapiano & Arkin, 2009; Tian et al., 2007), and numerous validation studies have evaluated the dependence of systematic and random errors on spatiotemporal scale, seasonality, and rain rate magnitude using TRMM-based precipitation products (Adler, 2001; Adler et al., 2009; AghaKouchak et al., 2012; Huffman et al., 2007; Maggioni et al., 2016). Recently, attention has been directed toward the IMERG precipitation products. Skofronick-Jackson et al. (2016) reported high biases in the IMERG V03 products at high rainfall rates. A recent study conducted over the U.S. Mid-Atlantic region evaluated IMERG V03 microwave-only (MW) and infrared-only (IR) fields against ground gauges and showed that IMERG underestimates heavy rain and exhibits significant random errors (Tan et al., 2016). Another study by Tan et al. (2017) evaluated the performance of IMERG V03 final run and TMPA 3B42V7 over the southeastern United States as a function of spatial and temporal scales using the radar-based Multi-Radar/Multi-Sensor System (MRMS) product as reference. Both TMPA and IMERG accuracy improves when aggregated to progressively coarser spatial (e.g., 0.1° to 2.5°) and temporal scales (e.g., 0.5 hr to daily). A recent study by O and Kirstetter (2018) compared IMERG final run against a surface precipitation reference derived from MRMS and observed that IMERG final run underestimates the diurnal variations of summer precipitation over mountainous regions of western and eastern United States and overestimates variations over the Central United States.

The GPM Ground Validation group relies heavily on reference data from ground radar and gauge measurements to evaluate the IMERG products (https://gpm-gv.gsfc.nasa.gov/). Among those, the MRMS suite of products, based on gauge-adjusted radar-based Quantitative Precipitation Estimates, provides a basis for deriving a high-quality independent reference to evaluate IMERG across the contiguous United States (hereinafter CONUS; Kirstetter, Hong, Gourley, Chen, et al., 2012). However, ground-based benchmarks are limited to specific locations, making it impossible to globally evaluate the performance of satellite precipitation products in regions where no ground-based observations exist, that is, oceans and impervious terrain. A first attempt was presented by Tian and Peters-Lidard (2010), who mapped the standard deviation of six precipitation products as an indicator to quantify the random error associated with those products. Nevertheless, this technique is simply a measure of the agreement among different data sets, but does not provide any information on the actual performance of each product.

Hence, there is no method currently available to globally evaluate random and systematic errors associated with multisatellite precipitation products at fine resolutions. This study proposes to assess the viability of using a DPR-based precipitation product as a benchmark for evaluating IMERG products (early, late, and final) and the microwave-only (MW) and infrared-only (IR) precipitation estimates. First, the hypothesis that DPR can be used as a reference for validating IMERG precipitation products is tested. Second, the errors in the IMERG products for which DPR is found to be a viable reference are further investigated as a function of climatology across CONUS. This work evaluates IMERG at its native spatial resolution of 0.1°, as opposed to past validation studies that mainly focused on aggregating high-resolution precipitation products to coarser spatial and temporal scales (e.g., Tan et al., 2017). The data sets used in this study are described in section 2. The

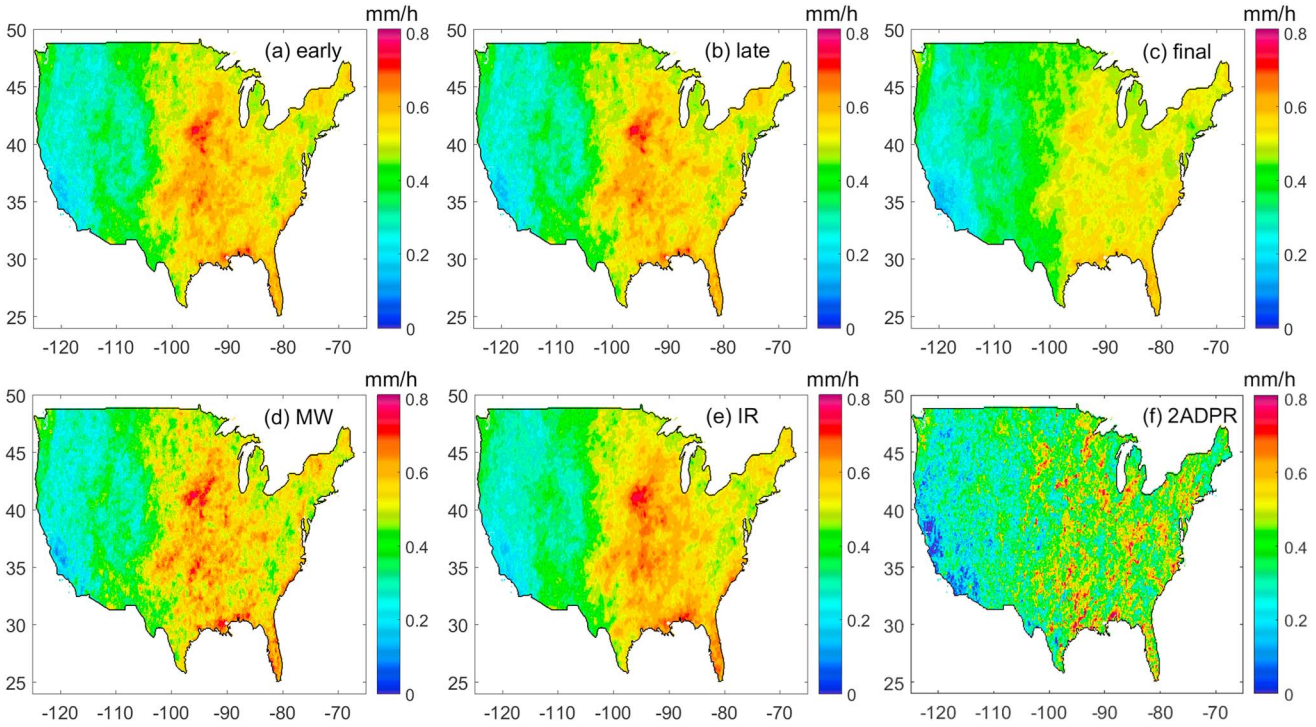

Figure 1. Average precipitation rate (mm/hr) over CONUS during April to October 2014–2015 for GPM-based products. (a) IMERG early. (b) IMERG late. (c) IMERG final. (d) Merged microwave component. (e) Infrared component. (f) 2ADPR.

methodology adopted to test the viability of using a DPR-based product as a reference for validating the IMERG is outlined in section 3. Results are presented and discussed in section 4, and conclusions are summarized in section 5.

2. Data Sets

CONUS is chosen as the study area for two reasons. First, the high-quality, high-resolution ground-based MRMS precipitation data set is available to evaluate the hypothesis that DPR can be used as an alternative to ground-based reference to validate GPM-based products. Second, CONUS presents large diversity in terms of terrain complexity and climatology (from tropical to continental to semiarid), characterized by various precipitation types (e.g., stratiform, convective, snow). A 2-year study period from April to October 2014–2015 is selected. Winter months are excluded from the analysis to avoid solid precipitation events and focus only on the evaluation of liquid-phase precipitation. The analysis is carried out at finest available spatial resolution of IMERG (0.1° × 0.1°) without any spatial aggregation and 1-hr temporal resolution. Figure 1 shows the mean precipitation over CONUS during the study period for all satellite products. Several products show high average precipitation over Midwest and Southeast. This can be attributed to the climatology of the area, characterized by moist continental midlatitude and tropical climates that are wet throughout the year. Satellite-based retrievals are better able to detect areas with strong precipitation signals. The performance of the satellite-based precipitation retrievals and sensors is largely dependent on the heterogeneity of the area, precipitation rates, seasonality, and climatology (Tian & Peters-Lidard, 2010). Details regarding the data sets used in this study are presented next.

2.1. MRMS

The MRMS algorithm incorporates data from all polarimetric WSR-88D radars and automated rain gauge networks (Zhang et al., 2011, 2016) to generate a number of hydrometeorological products, including quantitative precipitation estimates at fine resolution (0.01° and 2 min). The dual-polarized WSR-88D radars enable improved hydrometeor identification compared to nonpolarimetric methods (Chandrasekar et al., 2008; Melnikov et al., 2011). MRMS represents an independent reference for space-based precipitation products

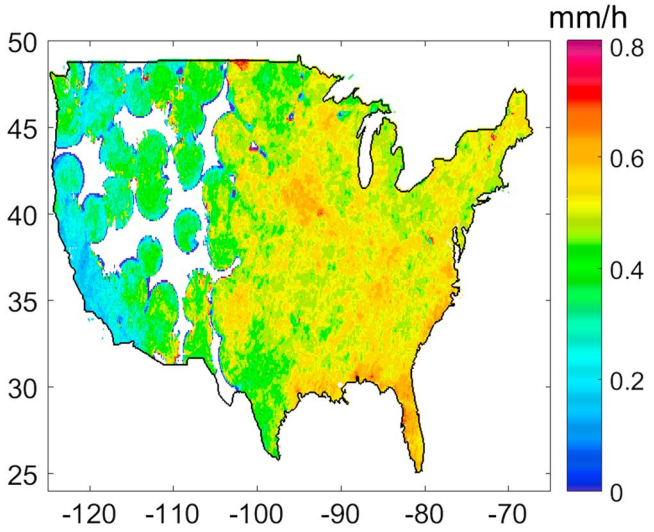

Figure 2. MRMS-based reference average precipitation rate (mm/hr) over CONUS (April to October 2014–2015). White pixels correspond to locations where the radar quality index is lower than 100.

and provides a consistent database in time and space over land, including various geographical (plains, mountains, etc.) and meteorological (subtropical to midlatitudes) conditions. This diversity enables to target the factors that often contribute to erroneous satellite retrievals. The ground-based precipitation reference is derived from a gauge-corrected, MRMS-gridded product with a temporal/spatial resolution of 1 hr/0.01°.

A reference precipitation is derived from MRMS precipitation products for comparison with IMERG. It includes gauge correction and quality filters, as described in Kirstetter, Hong, Gourley, Schwaller, et al. (2012) and Kirstetter et al. (2014). Only the best radar quality index values are selected for comparison with IMERG (Zhang et al., 2011). This selection assures that radar beam blockage is minimized and that the radar beam is below the bright band. The reference is regridded to match the IMERG spatial resolution and only pixels containing at least 90% of the highest-quality index values are considered in the comparison with the GPM-based products (Figure 2). The spatial precipitation pattern of IMERG final looks very similar to the MRMS-based reference's one, mainly because both products ingest gauge information in their algorithms.

2.2. Level-2 DPR Algorithms

The dual-frequency precipitation radar aboard the GPM core satellite operates at Ku band (13.6 GHz) and Ka band (35.5 GHz). The DPR scanning and sampling geometry lead to footprints spaced ~5 km, over 245-km (at Ku band) and 120-km-wide swaths (at Ka band). These observations provide a more direct measurement of the rain rate than the GPM Microwave Imager (GMI) and passive microwave sensors (PMW) in general, which form the basis of the IMERG products (described in the next section). Nevertheless, similarly to the TRMM PR, DPR observations are affected by precipitation phase, as well as the attenuation and extinction of the radar signal (Iguchi et al., 2000, 2009; Wolff & Fisher, 2008; Yang et al., 2006). These issues are addressed in the level-2 DPR algorithm (2ADPR). The DPR minimum theoretical detectable rain rate, which corresponds to its sensitivity, is ~18 dBZ (or ~0.5 mm/hr) at Ku band, and ~12 dBZ (or ~0.2 mm/hr) at Ka band. Recent studies have found that these values may be too conservative for certain meteorological analyses (Hamada & Takayabu, 2016), but in this work, we consider the threshold reported in the official *GPM/DPR Level-2 Algorithm Theoretical Basis Document* (Iguchi et al., 2010).

The level-2 DPR algorithms (2ADPR) produce radar-only derived meteorological quantities on an instantaneous field-of-view basis. 2ADPR algorithm version V05 is used to test the validity of using a satellite radar-based benchmark for validating the IMERG version V05 products and their MW and IR components. A study by Kirstetter et al. (2014) suggests that space-based active sensors are more accurate than passive sensors for precipitation estimation. Thus, testing 2ADPR as a potential reference to evaluate the IMERG products, which heavily rely on PMW observations, is a promising option. Specifically, the data archived in the *PreciEsurface* data field are employed in this study (Iguchi et al., 2010). In the GPM mission, the Level-2 PMW precipitation retrievals are trained with the 2BCMB combined algorithm before being used in the IMERG algorithm. When applied to the DPR radar reflectivity observations, the 2BCMB algorithm uses a different approach from the 2ADPR algorithm to interpret the reflectivity profiles and the resulting near-surface precipitation estimates display a different error structure (Petersen et al., 2016). Besides, several levels of processing from the 2BCMB estimates to the level-2 PMW (Bayesian retrieval) to IMERG (e.g., intercalibration across PMW, regridding, merging with IR) make the IMERG precipitation different from the 2BCMB precipitation. 2ADPR instantaneous orbital precipitation estimates are regridded to the IMERG resolution (0.1°) by averaging all the 2ADPR values falling within an IMERG specific pixel. The MRMS-based quality masks are consistently applied to the gridded 2ADPR data.

2.3. IMERG Products and Components

The IMERG algorithm fuses data from multiple sources, including several satellite microwave and infrared precipitation estimates, merged with gauge information (Huffman et al., 2014). PMW on low-Earth-orbit platforms provide relatively more accurate satellite precipitation estimates than IR, but with limited sampling.

This limitation is compensated by morphing the MW data using linear interpolation and tracking the motion of clouds through global infrared imagery. These data are then combined with MW-adjusted IR precipitation estimates in grid boxes that are relatively far in time from the nearest microwave overpass using a Kalman filter (Joyce & Xie, 2011). Additionally, the IMERG early and late products are adjusted to climatological sets of coefficients and IMERG final product to gauge observations.

The IMERG algorithm assembles and intercalibrates precipitation estimates from various satellite PMW sensors. Once gridded and combined into half-hourly fields, these estimates are provided to both the Climate Prediction Center (CPC) Morphing-Kalman Filter (CMORPH-KF) Lagrangian time interpolation and the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks–Cloud Classification System (PERSIANN-CCS; Hong et al., 2004) recalibration schemes. In parallel, CPC assembles and intercalibrates the IR fields and passes them to the PMM Precipitation Processing System (PPS) as inputs to PERSIANN-CCS. The PERSIANN-CCS estimates are computed and sent to the CMORPH-KF scheme, which uses the PMW and IR estimates to create half-hourly estimates.

At each observation time, the IMERG algorithm is run 3 times (twice in near real time and once after the monthly gauge analysis becomes available) to obtain (i) early multisatellite product, available ~4 hr (currently 5 hr) after the observation time for users primarily requiring fast answers for flood and landslide forecasts; (ii) the late multisatellite product, available ~12 hr (currently 16 hr) after the observation time for agricultural applications such as drought monitoring and crop yield forecasts and crop production; and (iii) the final satellite-gauge product, available ~2 months (currently 3.5 months) after the observation month for research applications. The IMERG final product is adjusted with monthly gauge observations, whereas the early and late products are calibrated with climatological sets of coefficients that vary by month and location. Therefore, final multisatellite estimates are considered more accurate and reliable. IMERG version V05 is employed in this study. The three products are provided at 0.1°/30-min spatial/temporal resolution. They were aggregated to the hourly resolution by simple averaging to match the resolution of the ground reference derived from MRMS.

The MW and IR components that contribute to the merged (early, late, final) products are also investigated in this work. The MW component is obtained from either calibrated conical-scan microwave radiometer retrieval or cross-track-scan microwave sounder retrieval, whichever is closer to the half-hour window. The GPM core satellite carries the GMI, which has 13 channels ranging from 10 to 183 GHz. There are typically 12–16 daily PMW overpasses over CONUS, whereas the IR component is assembled from a variety of sensors and is available all the time over CONUS. More details can be found in Huffman et al. (2014). The information about the data used in this study is archived in the following fields in the PPS files: *precipCal* for the (early, late, final) merged products, *HQprecipitation* for the MW component, and *IRprecipitation* for the IR component within each IMERG file (Huffman et al., 2014). Data fields provide global level-3 gridded precipitation estimates (mm/hr) at 0.1°/30-min spatial/temporal resolution. We understand that future versions of the IMERG products will provide improvements with respect to the current version V05. Nonetheless, the methodology developed in this study can be applied to any future version of IMERG.

3. Methodology

First, IMERG products, MRMS, and 2ADPR are matched in space and time during the evaluation period to a common 0.1°/1-h resolution. Second, the hypothesis that 2ADPR can be used as a reference for validating IMERG is tested for the IMERG products and the IR/MW components (section 3.1). Third, the effect of climatology is evaluated on the performance of those IMERG products and/or components for which 2ADPR can be used as a reference (section 3.2).

3.1. Assessing DPR as Benchmark

The systematic and random errors associated with all satellite-based products (i.e., IMERG products, IMERG components, and DPR) are investigated against the reference. In this study, we hypothesize that if the systematic and random errors between 2ADPR and the reference are smaller by at least 1 order of magnitude than the systematic and random errors in the IMERG satellite products, then 2ADPR can be used as reference to evaluate their performance. The rationale behind the choice of the 1-order-of-magnitude threshold is that a demonstrably superior reference is preferable to assess the data set being evaluated. However, this choice

is arbitrary and certain applications with less-than-perfect reference data sets may use a lower bar. While setting this bar depends on the end users, the overall methodology presented in this work will still apply.

The error is defined herein as the bias ratio. The conditional bias (CB) ratio between the satellite-based product R_{Sat} and the MRMS-based reference R_{Ref} is defined as

$$CB_j = \frac{\left(\sum_{n=1}^{N_j} R_{Sat|C_j}(n)\right)}{\left(\sum_{n=1}^{N_j} R_{Ref|C_j}(n)\right)} \tag{1}$$

where j ∈ [1 54] represents the rainfall rate bin number; R_{Sat} and R_{Ref} represent the satellite and reference rainfall rate, respectively, at each bin number j that satisfies the condition C_j; and N_j is the total number of bins that satisfies the condition C_j. The condition C_j applied to compute the bias is defined as $C_j : th_{j-1} \le R_{Ref} \le th_j$, where th_{j-1} and th_j represents the upper and lower limits of the j^{th} bin, respectively, with threshold, $th \in [0.2\ 10]$. Note that rainfall intervals are not uniform: denser for lower and more frequent rain rates (0.1 mm/hr between 0.2 and 5 mm/hr), sparser for higher and less frequent rain rates (1 mm/hr between 5 and 10 mm/hr).

The systematic error is assessed through the median conditional bias ratio (50th quantile), whereas the random error is represented by the spread of the conditional bias ratio distribution. Specifically, the interquantile difference (*IQD*) between the 90th (q_{90}) and 10th (q_{10}) quantiles of the conditional bias ratio distribution is adopted as an index for the random error:

$$IQD^j = \log_{10} q^j_{90} - \log_{10} q^j_{10} \tag{2}$$

The base-10 logarithmic operator in the *IQD* definition is used to highlight the order of magnitude difference among different *IQD* values. This procedure not only identifies the potential IMERG products and/or components for which 2ADPR is a viable reference (and could be used in place of the MRMS-derived reference where/when the latter is not available) but also verifies whether both systematic and random errors can be evaluated using 2ADPR.

3.2. IMERG Performance Metrics

Once the IMERG products and/or components for which 2ADPR is a viable reference are identified, they are evaluated against 2ADPR and against MRMS-based reference using both categorical and continuous statistics. The categorical statistics used are probability of detection (POD), success ratio (SR), critical success index (CSI), and the hit bias, which are defined as

$$\text{POD} : \frac{H}{H+M} \tag{3a}$$

$$\text{SR} : \frac{H}{H+F} \tag{3b}$$

$$\text{CSI} : \frac{H}{H+M+F} \tag{3c}$$

$$\text{Hit bias} : \frac{H+F}{H+M} \tag{3d}$$

Table 1
Contingency Table

		Satellite	
		$R_{Sat} \ge$ th	$R_{sat} <$ th
Reference	$R_{Ref} \ge$ th	H	M
	$R_{Ref} <$ th	F	Z

where H represents "hit" cases, that is, both the satellite (R_{Sat}) and the reference (R_{Ref}) are greater than or equal to the rain/no-rain threshold (th); F represents "false alarms," that is, R_{Sat} is greater than or equal to th, but R_{Ref} is less than th; M represents "misses," that is, R_{Ref} is greater than or equal to th but R_{Sat} is less than th; and Z represents "true negative," that is, R_{Sat} and R_{Ref} are both less than th. The contingency table parameters H, M, F, and Z are defined in Table 1. The ideal value for all the performance metrics (POD, SR, CSI, and hit bias) is 1.

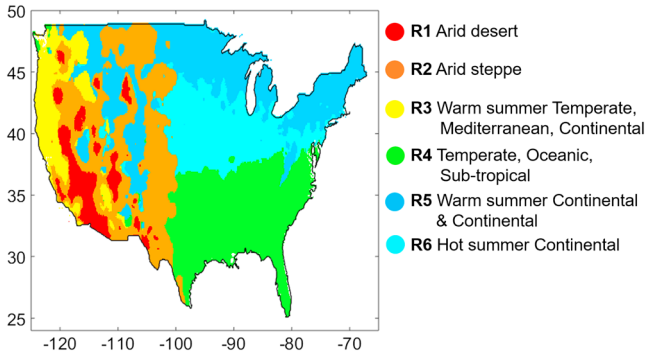

Figure 3. Koppen climatic mask for classification of CONUS into six regions. Refer to Table 2 for details regarding the classes.

Continuous statistics such as correlation, normalized standard deviation, and root-mean-square error are also evaluated. The ideal values for these metrics are 1 for correlation and standard deviation and 0 for root-mean-square error.

These statistics are evaluated as a function of the climatological region and seasonality, as the performance of satellite precipitation products has been proved to depend on several factors such as the prevailing climatology in the respective location, seasonality, and topography (Gebregiorgis & Hossain, 2013). Climate classification maps based on the Koppen-Geiger system are adopted over CONUS at the spatial resolution of 0.1° × 0.1° to investigate this dependence (Peel et al., 2007).

Figure 3 shows the geographical distribution of the six climatological zones into which CONUS has been classified. Although 29 Koppen climate classes exist over CONUS, some regions have been aggregated to ensure that each zone has a significant sample size to compute the performance metrics (Table 2). Details regarding the sample size of the data sets employed to compute the performance metrics are presented in Table 2. All five major climate classes defined by Peel et al. (2007) exist over CONUS, that is, tropical moist climates, dry climates, moist midlatitude climates with mild winters, moist midlatitudes climates with cold winters, and polar climates. As shown in Figure 3 and Table 2 , the northeastern region is characterized by moist continental midlatitude climates with cold winters and polar climates, the Midwest area by moist continental midlatitude climates that are wet throughout the year, whereas the southeast is characterized by humid subtropical and marine/oceanic climates. The northwest Pacific region is the wettest in the continental United States with year-round foggy and light drizzle conditions, whereas the southwestern part experiences the hottest weather and mostly convective rainfall events. The middle/south Pacific region is characterized by dry climates, Mediterranean, and moist continental midlatitude climates with cold winters primarily due to the mountainous terrain across this region.

4. Results and Discussion

4.1. DPR as Benchmark

The conditional bias ratio of each satellite-based product (IMERG early, late, final, IR, MW, and 2ADPR) with respect to the MRMS-based reference is investigated first. The 50th (dashed line), 25th–75th (dark gray-shaded), and 10th–90th (ivory-shaded) percentiles of the bias ratio distribution are shown in Figure 4. An underestimation with respect to the reference is observed for all products, with IR showing the largest underestimation. Moreover, the median of the conditional bias is larger at small rain rates and it stabilizes as the rain rates reach ~2 mm/hr to a value of 0.3 for IR, 0.6 for 2ADPR, and 0.7 for the other products. The larger bias at smaller rain rates can be attributed to the fact that small satellite estimated rain rates, close to sensitivity limits of the sensors, are normalized by low reference values (denominator of equation (1)), which leads to a large relative error. The spread of the conditional bias does not significantly change as a function of rain rate in logarithmic space. The IR component shows the widest spread among all satellite products and 2ADPR the narrowest, whereas the MW component and the IMERG products present very similar distributions, with the final IMERG product showing a slightly narrower spread than the near-real time and MW ones.

Table 2
Koppen Climate Classification Codes as Defined by Peel et al. (2007) and Their Description for CONUS With Corresponding Sample Size for Each Region (Matching Data Sets for Rain Rates ≥0.2 mm/hr)

	Koppen codes	Climate class	Sample size		
			IR/2ADPR/ref.	IR/2ADPR	IR/ref.
R1	BWh + BWk	Arid-desert	1,512	2,063	2,443
R2	BSh + BSk	Arid-steppe	12,218	15,661	20,227
R3	Csa + Csb + Dsa + Dsb + Dsc	Temperate, Mediterranean, continental with warm summers	1,368	2,129	2,486
R4	Cfa + Cfb + Af + Am	Temperate, oceanic, subtropical	34,959	42,986	52,938
R5	Dwa + Dwb + Dfb + Dfc	Warm summer continental, continental	20,418	26,232	32,525
R6	Dfa	Hot summer continental	22,506	29,274	34,850

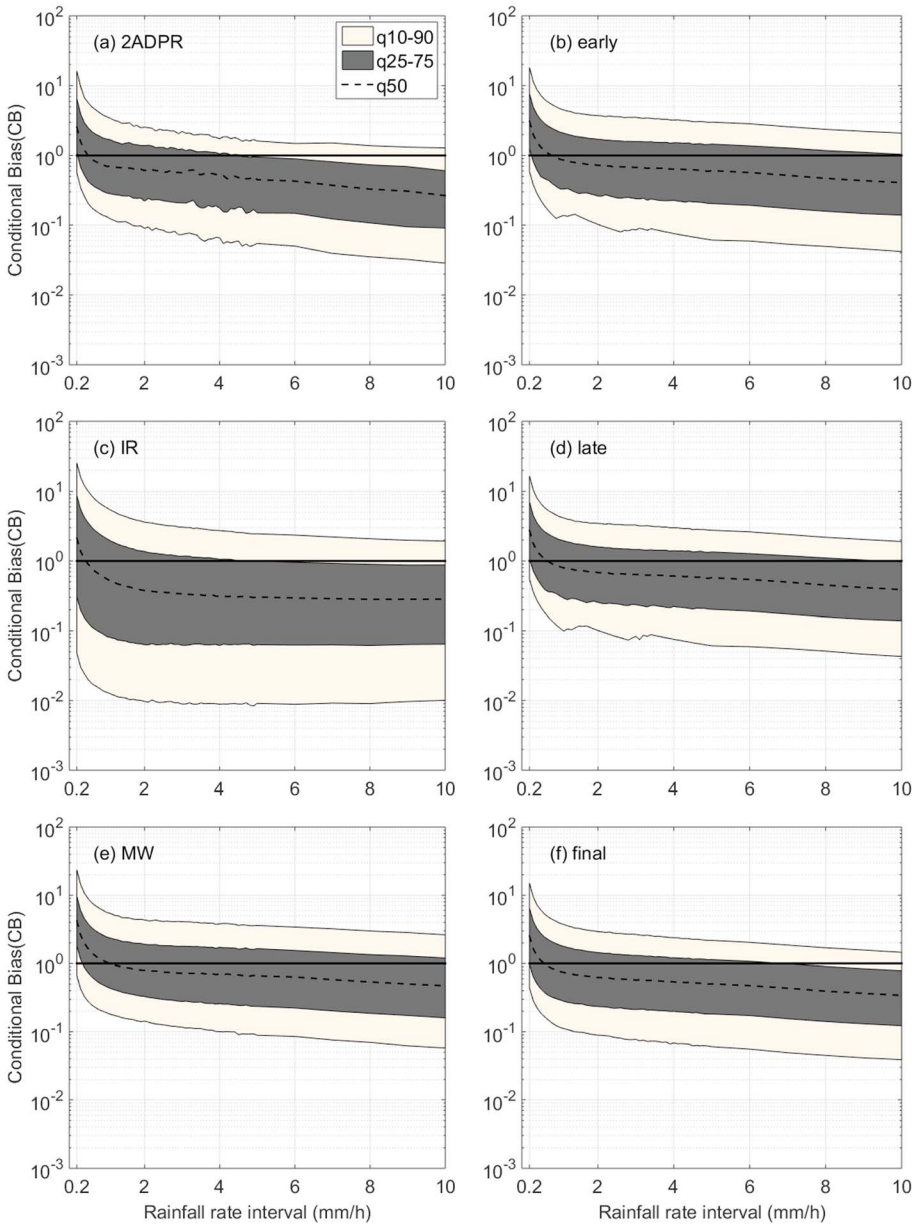

Figure 4. Conditional bias (CB) quantiles as a function of rainfall rate for (a) 2ADPR, (b) IMERG early, (c) IR component, (d) IMERG late, (e) MW component, and (f) IMERG final with respect to the MRMS-based reference. Note that there is a total of 54 rain rate interval bins and intermediate rainfall intervals are not labeled on the *x* axis. The dashed line represents the median (50% quantile), the dark gray-shaded region represents the area between the 25% and 75% percentiles, and the ivory-shaded region represents the area between the 10% and 90% percentiles of the bias ratio distribution. Solid black lines represent a perfect bias of 1.

In order to further investigate these differences, the systematic error (median conditional bias) and the random error (*IQD* between the 90th and the 10th quantiles) are shown in Figure 5 as a function of rainfall rate for all satellite-based products. The systematic errors for the IMERG products, MW, and 2ADPR are relatively close with each other, whereas IR presents a larger systematic error than any other satellite product. However, the difference between the 2ADPR systematic bias and any other satellite-based product/component is smaller than 1 order of magnitude. Therefore, the hypothesis that 2ADPR can be used as a reference to evaluate the systematic error of the GPM-based products and components is rejected.

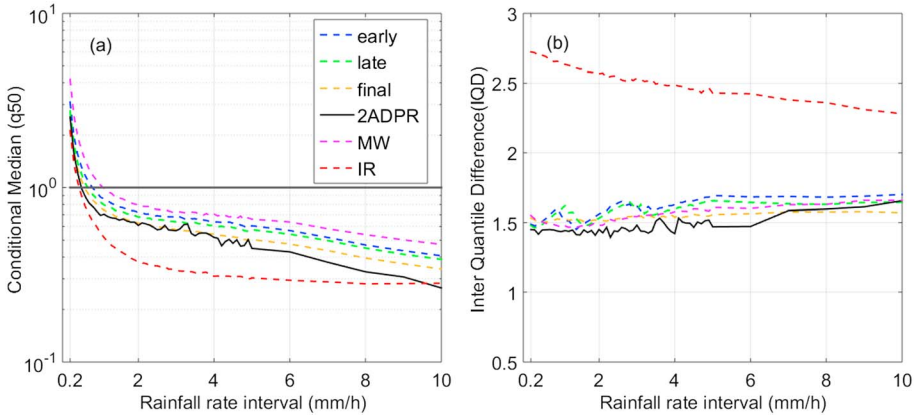

Figure 5. (a) Conditional median and (b) interquantile difference as a function of rainfall rate for all satellite-based products with respect to the MRMS-based reference.

In terms of random error, 2ADPR exhibits the smallest *IQD* among all satellite-based products, followed by the IMERG final product, first, and the IMERG early and late products and MW estimates, next. IR exhibits the largest random error against the MRMS-based reference with the *IQD* deviating from that of DPR by 1 order of magnitude (~1.25) for rain rates <5 mm/hr and slightly less than 1 order of magnitude for rain rates 5 to 10 mm/hr. Therefore, the hypothesis that DPR can be used as reference to validate IMERG products and MW precipitation estimates is rejected, but it is valid for evaluating random errors in the IR estimates.

In order to better understand the prevalence of IR measurements in IMERG algorithm, the contribution of the IR and MW components to the early, late, and final products are investigated. Specifically, the percentage weights that the IR data get in the Kalman filter step described in "Kalman Smoother framework" (Joyce & Xie, 2011), stored in the "*IRkalmanFilterWeight*" data field of the level-3 data set (0.1°/half-hourly spatial/temporal resolution) are considered to quantify the contribution of the IR component to the merged products (Huffman et al., 2014). The IR weights over CONUS during the study period reveal that ~50% of the weights are nonzero and that the nonzero weights are distributed between 15% and 100% with a slight skew to the left (Figure 6). This indicates that the IR component has nonnegligible contribution to the IMERG products, and therefore, error in the IR fields will influence the merged products.

4.2. IR Component Performance Analysis

As IR estimates show large random errors with respect to the ground-based reference and 2ADPR can be used as a reference to assess these errors, an in-depth analysis of the IR product is presented here. IR precipitation estimates are based on indirect estimation of surface precipitation via measurement of cloud top temperature and albedo. The algorithms used to translate that information into precipitation estimates lead to uncertainties associated with any IR retrieval (Kirstetter et al., 2017), thus supporting the need to evaluate errors in IR estimates.

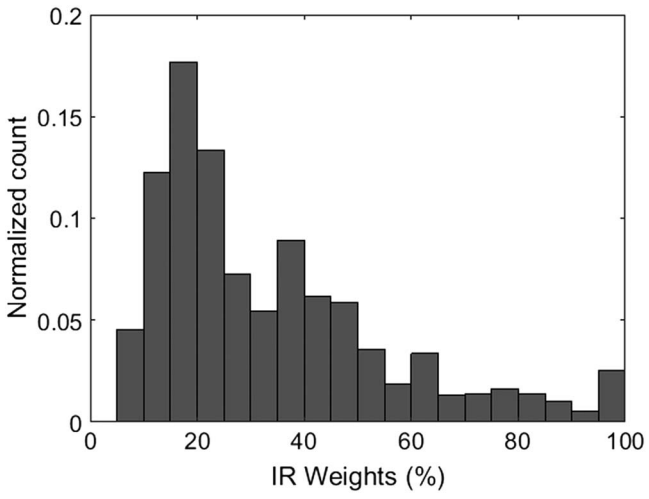

Figure 6. Probability distribution of nonzero IR weights over CONUS during April to October 2014–2015.

As climatology is shown to influence satellite precipitation observations (e.g., Khan et al., 2016), IR error metrics are computed with respect to 2ADPR- and MRMS-based reference for each of the six Koppen regions identified in section 3.2 over CONUS. The same error metrics are computed with respect to the MRMS-based reference. Note that these metrics are computed only for locations/time steps where/when IR estimates, MRMS-based reference, and 2ADPR are jointly available. A rain/no rain threshold value is set at 0.2 mm/hr, which is the theoretical minimum detectable rain rate of DPR. Then, categorical statistical scores and continuous metrics are computed for IR estimates against the 2ADPR- and MRMS-based references for each region at the pixel level (0.1°) and presented in the performance and Taylor diagrams.

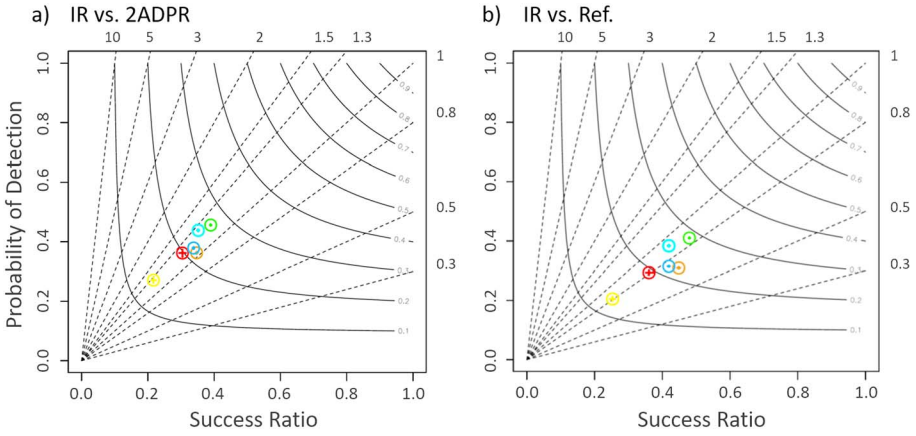

Figure 7. Performance diagrams of (a) IR estimates versus 2ADPR and (b) IR estimates versus MRMS-based reference over six Koppen climatic regions (colors correspond to those in Figure 3). Dotted lines correspond to hit bias and solid curves to critical success index (CSI) values. Circles represent probability of detection (POD) and success ratio (SR) for different regions. "Cross" indicates variance, with a perfect forecast lying in the top right region of the diagram.

The categorical statistics are summarized in the performance diagram (Roebber, 2009), which concisely presents probability of detection (POD), success ratio (SR), critical success index (CSI), and the hit bias. All the statistics presented in the performance diagram clearly show the dependence of the error metrics on climatology (Figure 7a). The best performance in terms of POD and SR (POD = 0.46 and SR = 0.4) is observed in R4 (temperate, oceanic, and subtropical). This region does not experience frequent temperature extremes and is usually characterized by cool temperatures and year-round precipitation. Cumulonimbus clouds are typically responsible for bringing heavy downpours and hail in this climatic region. As these clouds have ice crystals at their top surface, they reflect more sunlight (albedo~1) from the cloud top, which results in better detection of rainfall by the IR sensors (Adler & Negri, 1988).

R2 (arid-steppe) has slightly higher SR as compared to R1 (arid-desert), whereas R3 (temperate Mediterranean and continental with warm summers) exhibit low SR (~0.22) as compared to other regions. This is mainly attributed to the complex orography at high altitudes, where no-rain clouds are falsely identified as rainy clouds by the IR algorithms (Maggioni et al., 2016; Tian et al., 2007) and to snow cover on the mountains, which may affect the rain estimate (Maggioni et al., 2016). Moreover, the current IR algorithm is challenged in identifying warm precipitating clouds (Joyce et al., 2004; Karbalaee et al., 2017). For most climate regions, biases are greater than 1. R4 presents a hit bias comparable to R6 (~1.2), but for other Koppen climatic regions it is slightly less. This corroborates with what shown in past literature, which suggests that IR-dominant algorithms, such as PERSIANN-CCS, exhibit a bias to extreme precipitation events (Chen et al., 2014).

The Taylor diagram (Taylor, 2001) summarizes continuous statistics, such as correlation (COR), standard deviation (SD), and the centered root-mean-square error (RMSE) on a single plot by expressing RMSE as a function of COR and SD. We considered a standardized version of the Taylor diagram in Figure 8 where the SD and RMSE are normalized by the standard deviation of the reference estimates.

An average correlation of 0.12 (with a standard deviation of 0.05) is observed for the six Koppen regions, which suggests that the overall correlation between IR and 2ADPR estimates is low and comparable in all climatic regions (Figure 8a). The best correlation (0.2) is observed in R4 (temperate, oceanic, and subtropical) and the worst ($<$0.1) in R3 (temperate Mediterranean and continental with warm summers). Although correlations are low (0.1), R1 manifests the smallest relative RMSE (1.15). R2 and R4 have comparable SD and RMSE. RMSE and SD values are relatively high (~2) for R3 and R6. In R3, temperatures from April to October are mostly hot, the coastal areas in this region are relatively humid, and rainfall occurs mostly in winters. Hence, the low correlation value and poor rain detection for this region may be due to the study period. The temperate, oceanic, subtropics, and continental climates appear to be less influenced by seasonal variations. Among different subclasses of continental climates, R5 performs better than R6 in terms of correlation (COR). R5 and R6 show similar behavior in terms of SR with slightly better POD values for R6 and smaller RMSE

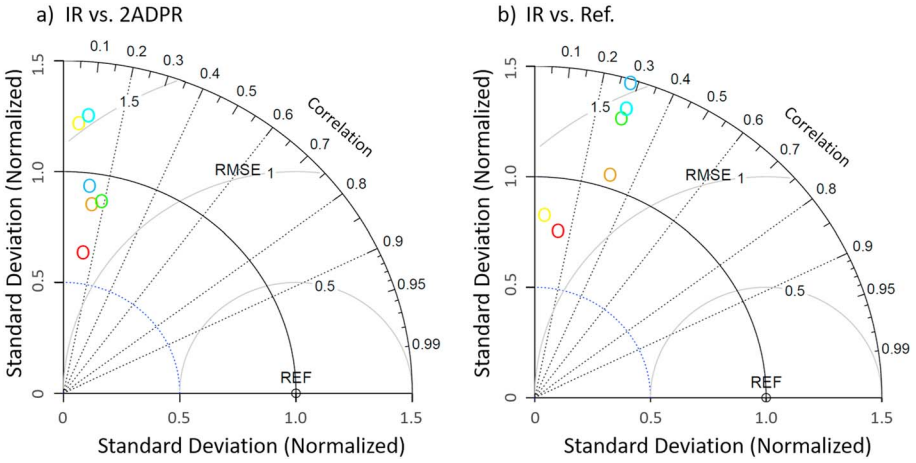

Figure 8. Same as in Figure 7 but for the Taylor diagrams, which depict standard deviation (SD dotted blue curves), RMSE (solid gray curves), and correlation (COR radial dotted black lines). All metrics are normalized with respect to the reference data. REF indicates the 2ADPR/MRMS-based reference (ref.) metrics (with itself), that is, SD and COR of 1 and RMSE of 0.

for R5. Although the correlations are small on an absolute scale, R4 has relatively better correlation than other regions and the percent increase could be considered meaningful.

We further analyze IR estimates against the MRMS-based reference to ascertain the climatological dependence of IR estimates (Figure 7b) and check the consistency when using 2ADPR as a reference. Results look similar to what is observed using 2ADPR as reference, with R4 and R3 manifesting the best and the worst performances, respectively. All other regions show the same relative behavior with marginal variations in POD, SR, BIAS, and CSI values. Statistics shown in the Taylor diagram also present very similar patterns as the ones observed using 2ADPR as reference, with R3 exhibiting low correlation and R6 exhibiting larger SD and RMSE.

The absolute differences in the IR errors with respect to 2ADPR and the MRMS-based references (ref.) are comparable for most metrics except for the hit bias. The absolute difference in POD, SR, CSI, and COR falls within the range of 0.01 to 0.19, whereas in terms of SD and RMSE it ranges from 0.03 to 0.54 mm/hr in all climatic regions. The hit bias shows the largest differences, which corroborates that 2ADPR can be used as a benchmark for quantifying uncertainties associated with IMERG when evaluating random errors associated with IR precipitation estimates (section 4.1).

The influence of seasonality on the performance of IR against 2ADPR- and MRMS-based reference precipitation estimates for the six Koppen climatic regions is assessed during three seasons (Figure 9): spring (April,

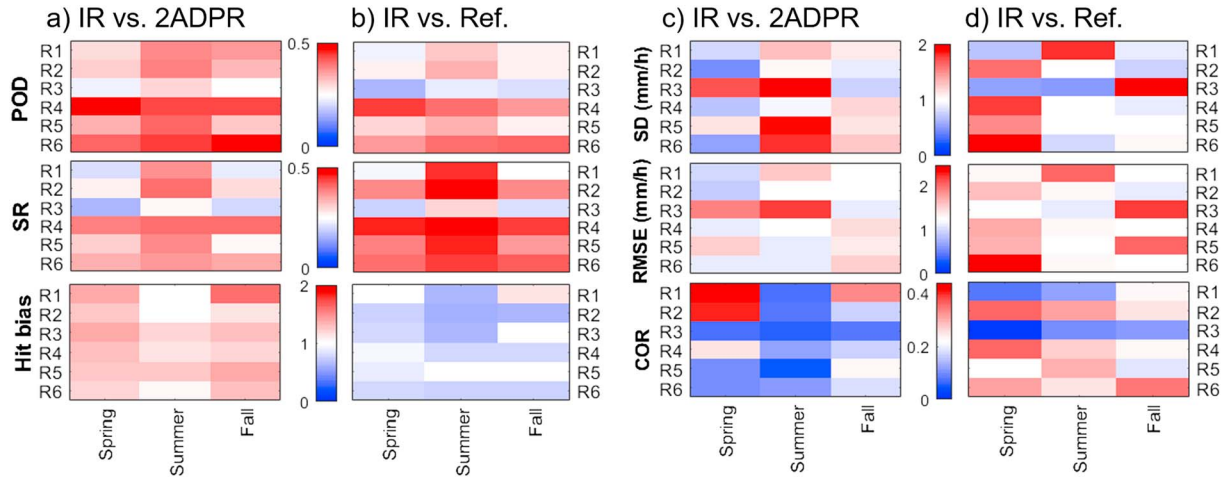

Figure 9. Performance of IR against (a, c) 2ADPR- and (b, d) MRMS-based reference (ref.) across the six regions during spring, summer, and fall of 2014 and 2015 in terms of categorical (a and b) and continuous (c and d) statistics.

May), summer (June, July, August), and fall (September, October), respectively. The best POD and SR appear in the same season in both comparisons, which is promising and supports the use of 2ADPR as a reference. Overall, the performance of IR versus the MRMS-based reference is comparable to that of IR versus 2ADPR, which is in line with the analyses over the entire evaluation period (Figures 7 and 8). Figures 9a and 9b demonstrate that the best performances in terms of POD, SR, and hit bias for almost all the regions (R1–R6) are during the summer season, with the exceptions of R4 and R6, which exhibit the highest POD in spring and fall, respectively. Similarly, R1, R2, R3, and R6 showed the worst performance in spring season and R5 performed worst in fall, whereas R4 has more variability with worst POD in summer and worst SR and hit bias in spring.

For each metric, the maximum seasonal change in each region is defined as the difference between maximum and minimum metric values among the three seasons. The range of maximum seasonal change for each metric is then computed across all the regions. This allows ranking the regions based on the maximum seasonal change for each metric. The maximum seasonal change ranged between 0.07 and 0.09 for POD, between 0.02 and 0.16 for SR, and between 0.15 and 0.6 for the hit bias. Based on the maximum seasonal change, the regions are ranked as R1, R5, R2, R3, R6, and R4 for each of POD, SR, and hit bias.

Figures 9c and 9d show the seasonal and climatological dependence of the continuous statistics. All regions (except R4) exhibit maximum standard deviation in summer with the maximum seasonal change in SD in different regions ranging over between 0.5 and 1.3 mm/hr. RMSE values showed more seasonal variability than SD. The maximum RMSE occurs in summer in R1 and R3, in fall in R2, R4 and R6, and in spring in R5. The maximum seasonal change in RMSE values ranged between 0.35 and 1.05 mm/hr. In terms of correlations, R1, R2, and R4 had maximum correlations in spring, whereas R3, R5, and R6 in fall when 2ADPR is used as the reference. The maximum seasonal change in correlations ranged between 0.08 and 0.3, with R1, R2, and R4 exhibiting the best correlations in spring and R3, R5, and R6 in fall.

Overall, R1, R2, and R5 depict relatively higher seasonal variations than the other three regions. Precipitation in R1 (arid-desert) and R2 (arid-steppe) is infrequent and characterized by low intensities, whereas continental climates with warm summers (R5) exhibit large variations in temperature and precipitation, which possibly resulted in their pronounced interseasonal fluctuations. This analysis highlights that the uncertainties associated with IR precipitation estimates depend on both seasonality and climatology.

5. Conclusions

This study assesses the potential of using the GPM 2ADPR estimates as a reference to evaluate the IMERG products for liquid-phase precipitation. Several satellite-based precipitation estimates, including version V05 of the IMERG early, late, and final products, their MW and IR components, and 2ADPR are evaluated against a ground radar reference derived from MRMS over CONUS. This work hypothesizes that if the error in 2ADPR (with respect to the ground reference) is smaller by at least 1 order of magnitude than the error in the satellite product (with respect to the ground reference), then 2ADPR can be used as reference to evaluate the performance of the satellite product. Both systematic and random errors are investigated.

Results show that the MW component and IMERG early, late, and final products have similar error distributions, while IR estimates have higher systematic error (defined as the median conditional bias ratio) and larger random errors (defined as the spread of the conditional bias ratio). Under the hypothesis of this study, 2ADPR is shown to be a viable reference for evaluating the random errors of the IR component of the IMERG product, as the spread of the conditional bias ratio for IR is larger than the 2ADPR one by 1 order of magnitude with respect to the reference. However, the end-users may deem a smaller error bar to be sufficient for their own goals, in which case 2ADPR could still be used as reference for evaluating PMW-influenced IMERG estimates. The proposed methodology can be generalized to other thresholds.

Uncertainties associated with IR estimates are further investigated by computing performance metrics as a function of climatology and seasonality against 2ADPR and the ground reference. Significant variations in the performance of IR estimates are observed in different climatic regions and seasons, which confirms the dependence of IR errors on climatology and seasonality. All climatic regions show low correlation values ($\leq$0.2) and an overall underestimation of the 2ADPR rainfall. In terms of contingency metrics, the temperate, oceanic, and subtropical climates perform best, but hit biases are close to 1 for most of the climate regions. In

terms of RMSE arid-desert manifests the smallest error (RMSE~0.6). At seasonal scale, arid-desert, arid-steppe, and continental climates with warm summer manifest largest variability in terms of both categorical and continuous statistics. This analysis on the performance of the IR estimates in different climatic regions calls for climatology-based improvements in the IR algorithms (Karbalaee et al., 2017).

This study focuses on testing the proposed methodology over CONUS only. A complete evaluation of error uncertainties using 2ADPR as a reference should be conducted globally and will be considered in future extension of the proposed approach. The application of this methodology to a longer time series and to the entire IMERG coverage could provide valuable information on random errors of IR estimates at a global scale and under different geographical locations and climatic conditions (e.g., complex terrains, precipitation types). Furthermore, the rainfall rate range (0.2–10 mm/hr) investigated here represents the most probable range of precipitation worldwide, but it does not include extreme rainfall events (>10 mm/hr). Nevertheless, the methodology developed in this study for the IMERG products and components could be applied to any other satellite-based precipitation data set. Also, the availability of high-resolution, high-quality 2ADPR precipitation estimates, although available only for the GPM overpasses, makes 2ADPR an ideal reference to evaluate random errors associated with IR estimates over the entire IMERG coverage.

The IMERG time series will eventually encompass both the TRMM and the GPM eras (1998 to present) and the coverage will be extended to fully global. This methodology serves as a stepping stone for validating IR component of IMERG estimates across the globe and assessing their uncertainties.

Finally, it should be highlighted that because of the different native spatiotemporal resolutions of different data sets, some discrepancies could have been introduced between IMERG, 2ADPR, and the MRMS-based reference, even if all three data sets were perfect. For example, the near-instantaneous overpasses of DPR would not yield the same precipitation rate as a ground-based radar, even if both were error free, because of uncertainties in the temporal and spatial averaging (the number of 2ADPR sample points fitting in the $0.1° \times 0.1°$ grid cell is not uniform). Furthermore, the MRMS-based reference has temporal/spatial resolution of 1 hr/0.01°, and is resampled to align with the IMERG and 2ADPR resolution. In light of these issues, future studies and different applications could apply the methodology developed here by relaxing our hypothesis and setting a bar lower than 1 order of magnitude. Nevertheless, the findings of this study clearly support that such an approach would still need to differentiate between IR-based and PMW-based IMERG observations and treat each separately. An alternative approach to address these deficiencies in the analyzed data sets could be the Triple Collocation Analysis (Alemohammad et al., 2015; McColl et al., 2014; Roebeling et al., 2012), which compares data sets with an unknown truth and is therefore extremely useful when/where no reference is available. However, there are no sufficient data samples over CONUS during the 2-year period at fine spatial/temporal resolution (0.1°/1 hr). When the IMERG products will be reprocessed for the TRMM era, Triple Collocation Analysis could be a useful tool to investigate uncertainties of each product.

Acknowledgments
The authors would like to thank the NASA Precipitation Processing System (PPS) for providing the IMERG and DPR data. All the data used are listed in the references or archived in https://pmm.nasa.gov/data-access/downloads/gpm repository. Pierre Kirstetter and Viviana Maggioni acknowledge support through the NASA Precipitation Measurement Missions (PMM) award NNX16AE39G. The authors would like to thank the reviewers for their constructive comments and suggestions that helped in improving the quality of this article.

References

Adler, R. F. (2001). TRMM: Status of precipitation estimates, science highlights, and 3-hour global, tropical precipitation estimates (invited presentation), in: 11th Conference on Satellite Meteorology and Oceanography.

Adler, R. F., Huffman, G. J., Chang, A., Ferraro, R., Xie, P.-P., Janowiak, J., et al. (2003). The version-2 Global Precipitation Climatology Project (GPCP) monthly precipitation analysis (1979–present). *Journal of Hydrometeorology*, *4*(6), 1147–1167. https://doi.org/10.1175/1525-7541(2003)004<1147:TVGPCP>2.0.CO;2

Adler, R. F., & Negri, A. J. (1988). A satellite infrared technique to estimate tropical convective and stratiform rainfall. *Journal of Applied Meteorology*, *27*(1), 30–51. https://doi.org/10.1175/1520-0450(1988)027<0030:ASITTE>2.0.CO;2

Adler, R. F., Wang, J.-J., Gu, G., & Huffman, G. J. (2009). A ten-year tropical rainfall climatology based on a composite of TRMM products. *Journal of the Meteorological Society of Japan Series II*, *87*, 281–293.

AghaKouchak, A., Mehran, A., Norouzi, H., & Behrangi, A. (2012). Systematic and random error components in satellite precipitation data sets. *Geophysical Research Letters*, *39*, L09406. https://doi.org/10.1029/2012GL051592

Alemohammad, S. H., McColl, K. A., Konings, A. G., Entekhabi, D., & Stoffelen, A. (2015). Characterization of precipitation product errors across the United States using multiplicative triple collocation. *Hydrology and Earth System Sciences*, *19*(8), 3489–3503. https://doi.org/10.5194/hess-19-3489-2015

Anagnostou, E. N., Maggioni, V., Nikolopoulos, E. I., Meskele, T., Hossain, F., & Papadopoulos, A. (2010). Benchmarking high-resolution global satellite rainfall products to radar and rain-gauge rainfall estimates. *IEEE Transactions on Geoscience and Remote Sensing*, *48*(4), 1667–1683. https://doi.org/10.1109/TGRS.2009.2034736

Chandrasekar, V., Bringi, V. N., Rutledge, S. A., Hou, A., Smith, E., Jackson, G. S., et al. (2008). Potential role of dual-polarization radar in the validation of satellite precipitation measurements: Rationale and opportunities. *Bulletin of the American Meteorological Society*, *89*(8), 1127–1146. https://doi.org/10.1175/2008BAMS2177.1

Chen, S., Liu, H., You, Y., Mullens, E., Hu, J., Yuan, Y., et al. (2014). Evaluation of high-resolution precipitation estimates from satellites during July 2012 Beijing flood event using dense rain gauge observations. *PLoS One*, *9*, e89681.

Curtis, S., Salahuddin, A., Adler, R. F., Huffman, G. J., Gu, G., & Hong, Y. (2007). Precipitation extremes estimated by GPCP and TRMM: ENSO relationships. *Journal of Hydrometeorology*, *8*(4), 678–689. https://doi.org/10.1175/JHM601.1

Falck, A. S., Maggioni, V., Tomasella, J., Vila, D. A., & Diniz, F. L. (2015). Propagation of satellite precipitation uncertainties through a distributed hydrologic model: A case study in the Tocantins–Araguaia basin in Brazil. *Journal of Hydrology*, *527*, 943–957. https://doi.org/10.1016/j.jhydrol.2015.05.042

Gebregiorgis, A. S., & Hossain, F. (2013). Understanding the dependence of satellite rainfall uncertainty on topography and climate for hydrologic model simulation. *IEEE Transactions on Geoscience and Remote Sensing*, *51*(1), 704–718. https://doi.org/10.1109/TGRS.2012.2196282

Hamada, A., & Takayabu, Y. N. (2016). Improvements in detection of light precipitation with the global precipitation measurement dual-frequency precipitation radar (GPM DPR). *Journal of Atmospheric and Oceanic Technology*, *33*(4), 653–667. https://doi.org/10.1175/JTECH-D-15-0097.1

Hong, Y., Adler, R. F., Negri, A., & Huffman, G. J. (2007). Flood and landslide applications of near real-time satellite rainfall products. *Natural Hazards*, *43*(2), 285–294. https://doi.org/10.1007/s11069-006-9106-x

Hong, Y., Hsu, K.-L., Sorooshian, S., & Gao, X. (2004). Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *Journal of Applied Meteorology*, *43*(12), 1834–1853. https://doi.org/10.1175/JAM2173.1

Hossain, F., & Anagnostou, E. N. (2004). Assessment of current passive-microwave- and infrared-based satellite rainfall remote sensing for flood prediction. *Journal of Geophysical Research*, *109*, D07102. https://doi.org/10.1029/2003JD003986

Hou, A. Y., Kakar, R. K., Neeck, S., Azarbarzin, A. A., Kummerow, C. D., Kojima, M., et al. (2014). The global precipitation measurement mission. *Bulletin of the American Meteorological Society*, *95*(5), 701–722. https://doi.org/10.1175/BAMS-D-13-00164.1

Houze, R. A., Rasmussen, K. L., Zuluaga, M. D., & Brodzik, S. R. (2015). The variable nature of convection in the tropics and subtropics: A legacy of 16 years of the tropical rainfall measuring mission satellite. *Reviews of Geophysics*, *53*, 994–1021. https://doi.org/10.1002/2015RG000488

Huffman, G. J. (1997). Estimates of root-mean-square random error for finite samples of estimated precipitation. *Journal of Applied Meteorology*, *36*, 1191–1201.

Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Xie, P., & Yoo, S.-H. (2014). NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG). Algorithm Theor. Basis Doc. ATBD NASAGSFC Greenbelt MD USA.

Huffman, G. J., Bolvin, D. T., Nelkin, E. J., Wolff, D. B., Adler, R. F., Gu, G., et al. (2007). The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *Journal of Hydrometeorology*, *8*(1), 38–55. https://doi.org/10.1175/JHM560.1

Iguchi, T., Kozu, T., Kwiatkowski, J., Meneghini, R., Awaka, J., & Okamoto, K. (2009). Uncertainties in the rain profiling algorithm for the TRMM precipitation radar. *Journal of the Meteorological Society of Japan Series II*, *87*, 1–30.

Iguchi, T., Kozu, T., Meneghini, R., Awaka, J., & Okamoto, K. (2000). Rain-profiling algorithm for the TRMM precipitation radar. *Journal of Applied Meteorology*, *39*(12), 2038–2052. https://doi.org/10.1175/1520-0450(2001)040<2038:RPAFTT>2.0.CO;2

Iguchi, T., Seto, S., Meneghini, R., Yoshida, N., Awaka, J., & Kubota, T. (2010). GPM/DPR Level-2 Algorithm Theoretical Basis Document. NASA Goddard Space Flight Cent. Greenbelt MD USA Tech Rep.

Johnson, G. E., Achutuni, V. R., Thiruvengadachari, S., & Kogan, F. (1993). The role of NOAA satellite data in drought early warning and monitoring: Selected case studies. In *Drought Assessment, Management, and Planning: Theory and Case Studies, Natural Resource Management and Policy* (pp. 31–47). Boston, MA: Springer. https://doi.org/10.1007/978-1-4615-3224-8_3

Joyce, R. J., Janowiak, J. E., Arkin, P. A., & Xie, P. (2004). CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *Journal of Hydrometeorology*, *5*(3), 487–503. https://doi.org/10.1175/1525-7541(2004)005<0487:CAMTPG>2.0.CO;2

Joyce, R. J., & Xie, P. (2011). Kalman filter-based CMORPH. *Journal of Hydrometeorology*, *12*(6), 1547–1563. https://doi.org/10.1175/JHM-D-11-022.1

Karbalaee, N., Hsu, K., Sorooshian, S., & Braithwaite, D. (2017). Bias adjustment of infrared-based rainfall estimation using passive microwave satellite rainfall data. *Journal of Geophysical Research: Atmospheres*, *122*, 3859–3876. https://doi.org/10.1002/2016JD026037

Khan, S., Maggioni, V., & Porcacchia, L. (2016). Uncertainties associated with the IMERG Multi-Satellite precipitation product, in: Geoscience and Remote Sensing Symposium (IGARSS), 2016 IEEE International. IEEE (pp. 2127–2130).

Kidd, C., Becker, A., Huffman, G. J., Muller, C. L., Joe, P., Skofronick-Jackson, G., & Kirschbaum, D. B. (2017). So, how much of the Earth's surface is covered by rain gauges? *Bulletin of the American Meteorological Society*, *98*(1), 69–78. https://doi.org/10.1175/BAMS-D-14-00283.1

Kirschbaum, D. B., Huffman, G. J., Adler, R. F., Braun, S., Garrett, K., Jones, E., et al. (2016). NASA's remotely-sensed precipitation: A reservoir for applications users. *Bulletin of the American Meteorological Society*, *98*(6), 1169–1184.

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Cao, Q., Schwaller, M., & Petersen, W. (2014). A research framework to bridge from the global precipitation measurement mission core satellite to the constellation sensors using ground radar-based National Mosaic QPE. In L. Venkataraman (Ed.), *Remote Sensing of the Terrestrial Water Cycle*, *AGU books Geophysical Monograph Series, Chapman monograph on remote sensing* (pp. 61–79). Hoboken, NJ: John Wiley & Sons Inc.

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Chen, S., Flamig, Z., Zhang, J., et al. (2012). Toward a framework for systematic error modeling of spaceborne precipitation radar with NOAA/NSSL ground radar-based National Mosaic QPE. *Journal of Hydrometeorology*, *13*(4), 1285–1300. https://doi.org/10.1175/JHM-D-11-0139.1

Kirstetter, P.-E., Hong, Y., Gourley, J. J., Schwaller, M., Petersen, W., & Zhang, J. (2012). Comparison of TRMM 2A25 products, version 6 and version 7, with NOAA/NSSL ground radar-based National Mosaic QPE. *Journal of Hydrometeorology*, *14*(2), 661–669. https://doi.org/10.1175/JHM-D-12-030.1

Kirstetter, P.-E., Karbalaee, N., Hsu, K., & Hong, Y. (2017). Probabilistic precipitation rate estimates with space-based infrared sensors. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3243

Kojima, M., Miura, T., Furukawa, K., Hyakusoku, Y., Ishikiri, T., Kai, H., et al. (2012). Dual-frequency precipitation radar (DPR) development on the global precipitation measurement (GPM) core observatory. Presented at the Earth observing missions and sensors: Development, implementation, and characterization II. *International Society for Optics and Photonics*, 85281A. https://doi.org/10.1117/12.976823

Krajewski, W. F., Anderson, M. C., Eichinger, W. E., Entekhabi, D., Hornbuckle, B. K., Houser, P. R., et al. (2006). A remote sensing observatory for hydrologic sciences: A genesis for scaling to continental hydrology. *Water Resources Research*, *42*, W07301. https://doi.org/10.1029/2005WR004435

Maggioni, V., Anagnostou, E. N., & Reichle, R. H. (2012). The impact of land model structural, parameter, and forcing errors on the characterization of soil moisture uncertainty. *Hydrology and Earth System Sciences Discussions*, *9*(2), 2283–2319. https://doi.org/10.5194/hessd-9-2283-2012

Maggioni, V., Sapiano, M. R., & Adler, R. F. (2016). Estimating uncertainties in high-resolution satellite precipitation products: Systematic or random error? *Journal of Hydrometeorology*, *17*(4), 1119–1129. https://doi.org/10.1175/JHM-D-15-0094.1

McColl, K. A., Vogelzang, J., Konings, A. G., Entekhabi, D., Piles, M., & Stoffelen, A. (2014). Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target. *Geophysical Research Letters*, *41*, 6229–6236. https://doi.org/10.1002/2014GL061322

Melnikov, V. M., Doviak, R. J., Zrnić, D. S., & Stensrud, D. J. (2011). Mapping Bragg scatter with a polarimetric WSR-88D. *Journal of Atmospheric and Oceanic Technology*, *28*(10), 1273–1285. https://doi.org/10.1175/JTECH-D-10-05048.1

Mousam, A., Maggioni, V., Delamater, P. L., & Quispe, A. M. (2016). Using remote sensing and modeling techniques to investigate the annual parasite incidence of malaria in Loreto, Peru. *Advances in Water Resources*, *108*, 423–438. https://doi.org/10.1016/j.advwatres.2016.11.009

O, S., & Kirstetter, P.-E. (2018). Evaluation of diurnal variation of GPM IMERG-derived summer precipitation over the contiguous US using MRMS data. *Quarterly Journal of the Royal Meteorological Society*. https://doi.org/10.1002/qj.3218

Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*, *4*(2), 439–473. https://doi.org/10.5194/hessd-4-439-2007

Petersen, W., Kirstetter, P., Wolff, D., Kidd, C., Tokay, A., Chandrasekar, V., et al. (2016). GPM level 1 science requirements: Science and performance viewed from the ground.

Prakash, S., Mitra, A. K., Pai, D. S., & AghaKouchak, A. (2016). From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, *88*, 1–7. https://doi.org/10.1016/j.advwatres.2015.11.008

Roebber, P. J. (2009). Visualizing multiple measures of forecast quality. *Weather and Forecasting*, *24*(2), 601–608. https://doi.org/10.1175/2008WAF2222159.1

Roebeling, R. A., Wolters, E. L. A., Meirink, J. F., & Leijnse, H. (2012). Triple collocation of summer precipitation retrievals from SEVIRI over Europe with gridded rain gauge and weather radar data. *Journal of Hydrometeorology*, *13*(5), 1552–1566. https://doi.org/10.1175/JHM-D-11-089.1

Sapiano, M. R. P., & Arkin, P. A. (2009). An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. *Journal of Hydrometeorology*, *10*(1), 149–166. https://doi.org/10.1175/2008JHM1052.1

Skofronick-Jackson, G., Petersen, W. A., Berg, W., Kidd, C., Stocker, F. E., Kirschbaum, D. B., et al. (2016). The Global Precipitation Measurement (GPM) mission for science and society. *Bulletin of the American Meteorological Society*. https://doi.org/10.1175/BAMS-D-15-00306.1

Sorooshian, S., AghaKouchak, A., Arkin, P., Eylander, J., Foufoula-Georgiou, E., Harmon, R., et al. (2011). Advanced concepts on remote sensing of precipitation at multiple scales. *Bulletin of the American Meteorological Society*, *92*(10), 1353–1357. https://doi.org/10.1175/2011BAMS3158.1

Stephens, G. L., & Kummerow, C. D. (2007). The remote sensing of clouds and precipitation from space: A review. *Journal of the Atmospheric Sciences*, *64*(11), 3742–3765. https://doi.org/10.1175/2006JAS2375.1

Tan, J., Petersen, W. A., Kirstetter, P.-E., & Tian, Y. (2017). Performance of IMERG as a function of spatiotemporal scale. *Journal of Hydrometeorology*, *18*(2), 307–319. https://doi.org/10.1175/JHM-D-16-0174.1

Tan, J., Petersen, W. A., & Tokay, A. (2016). A novel approach to identify sources of errors in IMERG for GPM ground validation. *Journal of Hydrometeorology*, *17*(9), 2477–2491. https://doi.org/10.1175/JHM-D-16-0079.1

Taylor, K. E. (2001). Summarizing multiple aspects of model performance in a single diagram. *Journal of Geophysical Research*, *106*, 7183–7192. https://doi.org/10.1029/2000JD900719

Tian, Y., & Peters-Lidard, C. D. (2010). A global map of uncertainties in satellite-based precipitation measurements. *Geophysical Research Letters*, *37*, L24407. https://doi.org/10.1029/2010GL046008

Tian, Y., Peters-Lidard, C. D., Choudhury, B. J., & Garcia, M. (2007). Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. *Journal of Hydrometeorology*, *8*(6), 1165–1183. https://doi.org/10.1175/2007JHM859.1

Wolff, D. B., & Fisher, B. L. (2008). Comparisons of instantaneous TRMM ground validation and satellite rain-rate estimates at different spatial scales. *Journal of Applied Meteorology and Climatology*, *47*(8), 2215–2237. https://doi.org/10.1175/2008JAMC1875.1

Yang, S., Olson, W. S., Wang, J.-J., Bell, T. L., Smith, E. A., & Kummerow, C. D. (2006). Precipitation and latent heating distributions from satellite passive microwave radiometry. Part II: Evaluation of estimates using independent data. *Journal of Applied Meteorology and Climatology*, *45*, 721–739.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., et al. (2016). Multi-Radar Multi-Sensor (MRMS) quantitative precipitation estimation: Initial operating capabilities. *Bulletin of the American Meteorological Society*, *97*(4), 621–638. https://doi.org/10.1175/BAMS-D-14-00174.1

Zhang, J., Howard, K., Langston, C., Vasiloff, S., Kaney, B., Arthur, A., et al. (2011). National Mosaic and multi-sensor QPE (NMQ) system: Description, results, and future plans. *Bulletin of the American Meteorological Society*, *92*(10), 1321–1338. https://doi.org/10.1175/2011BAMS-D-11-00047.1