

## Comparative Ground Validation of IMERG and TMPA at Variable Spatiotemporal Scales in the Tropical Andes

BASTIAN MANZ

*Department of Civil and Environmental Engineering, Imperial College London, London, United Kingdom*

SEBASTIÁN PÁEZ-BIMOS AND NATALIA HORNA

*Departamento de Ingeniería Civil y Ambiental, Facultad de Ingeniería Civil y Ambiental,  
Escuela Politécnica Nacional, Quito, Ecuador*

WOUTER BUYTAERT AND BORIS OCHOA-TOCACHI

*Department of Civil and Environmental Engineering, and Grantham Institute—Climate Change and the Environment,  
Imperial College London, London, United Kingdom*

WALDO LAVADO-CASIMIRO

*Servicio Nacional de Meteorología e Hidrología, Lima, Peru*

BRAM WILLEMS


*Centro de Competencias del Agua, Lima, Peru*

(Manuscript received 30 November 2016, in final form 25 June 2017)

### ABSTRACT

An initial ground validation of the Integrated Multisatellite Retrievals for GPM (IMERG) Day-1 product from March 2014 to August 2015 is presented for the tropical Andes. IMERG was evaluated along with the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) against 302 quality-controlled rain gauges across Ecuador and Peru. Detection, quantitative estimation statistics, and probability distribution functions are calculated at different spatial ( $0.1^\circ$ ,  $0.25^\circ$ ) and temporal (1 h, 3 h, daily) scales. Precipitation products are analyzed for hydrometeorologically distinct subregions. Results show that IMERG has a superior detection and quantitative rainfall intensity estimation ability than TMPA, particularly in the high Andes. Despite slightly weaker agreement of mean rainfall fields, IMERG shows better characterization of gauge observations when separating rainfall detection and rainfall rate estimation. At corresponding space–time scales, IMERG shows better estimation of gauge rainfall probability distributions than TMPA. However, IMERG shows no improvement in both rainfall detection and rainfall rate estimation along the dry Peruvian coastline, where major random and systematic errors persist. Further research is required to identify which rainfall intensities are missed or falsely detected and how errors can be attributed to specific satellite sensor retrievals. The satellite–gauge difference was associated with the point–area difference in spatial support between gauges and satellite precipitation products, particularly in areas with low and irregular gauge network coverage. Future satellite–gauge evaluations need to identify such locations and investigate more closely interpixel point–area differences before attributing uncertainties to satellite products.

---

 Denotes content that is immediately available upon publication as open access.

---

*Corresponding author:* Bastian Manz, [bastian.manz10@alumni.imperial.ac.uk](mailto:bastian.manz10@alumni.imperial.ac.uk)

---

*Publisher's Note:* This article was revised on 17 August 2018 to include the open access designation and CCBY license that should have been present when originally published.



This article is licensed under a [Creative Commons Attribution 4.0 license](http://creativecommons.org/licenses/by/4.0/) (<http://creativecommons.org/licenses/by/4.0/>).

DOI: 10.1175/JHM-D-16-0277.1

© 2017 American Meteorological Society

## 1. Introduction

The lack of reliable observations of hydrological variables in the tropics leads to a poor understanding of the hydrological cycle in those regions (Wohl et al. 2012). Especially in mountain regions, such as the tropical Andes, the complex topography results in highly variable spatiotemporal precipitation patterns that are not fully captured by the existing gauge monitoring networks (Ochoa-Tocachi et al. 2016; Rollenbeck and Bendix 2011; Buytaert et al. 2006). In the last decades satellite-based precipitation products (SPPs) have become an alternative source of precipitation estimation with widespread applications such as (distributed) hydrological modeling (Falck et al. 2015; Zulkaffi et al. 2014; Jiang et al. 2012; Li et al. 2009), geomorphology and landscape evolution (Nesbitt and Anders 2009; Bookhagen and Strecker 2008), streamflow forecasting (Nikolopoulos et al. 2013; Li et al. 2009), and early warning systems (Tian et al. 2010), as well as investigations into atmospheric processes and storm structures (Boers et al. 2015; Mohr et al. 2014; Boers et al. 2013; Rasmussen et al. 2013; Demaria et al. 2011).

Assessments of SPPs against rain gauge networks in the tropical Andes of Ecuador and Peru have shown a general dependence of SPP performance on rainfall intensity (Mantas et al. 2015), in addition to underestimation of the amplitude of the seasonal cycle as well as underestimation of extreme rainfall intensities (Derin et al. 2016). SPPs have also been shown to systematically overestimate low rainfall intensities (under  $5 \text{ mm h}^{-1}$ ), which occur frequently above 2000 m MSL (Derin et al. 2016) and locally represent an important contribution to the total rainfall volume (Padrón et al. 2015). Furthermore, there are regional differences in SPP performance depending on the principal sensor technology and the local rainfall properties as a result of interaction of the synoptic-scale climate processes with the complex topography (Derin et al. 2016; Satgé et al. 2016; Dinku et al. 2010). The Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN; Hsu et al. 1999) product showed large biases and low correlation with rain gauges (Ward et al. 2011; Derin et al. 2016) and was particularly deficient in regions with a significant contribution by deep convective systems to the total rainfall volume (Derin et al. 2016). Comparative studies have shown that the Tropical Rainfall Measuring Mission (TRMM) Multisatellite Precipitation Analysis (TMPA) outperformed other SPPs, including PERSIANN, the Climate Prediction Center (CPC) morphing technique (CMORPH; Joyce et al. 2004), and the Global Satellite Mapping of Precipitation (GSMaP; Kubota et al. 2007),

among others, in terms of correlation, rainfall intensity distribution, and bias (Satgé et al. 2016; Derin et al. 2016).

The Global Precipitation Measurement (GPM) *Core Observatory*, launched on 28 February 2014, has been designed to address critical limitations of TRMM and to further improve the scientific contribution of its predecessor in understanding the global water and energy cycle (Hou et al. 2014). The TRMM Precipitation Radar (PR) detection limit of 17 dBZ ( $\sim 0.7 \text{ mm h}^{-1}$ ) and the resulting large fraction of missed rainfall was addressed using a dual-frequency radar (DPR). The PR Ku-band frequency of 13.6 GHz is supplemented by a Ka-band frequency at 35.5 GHz for better identification of different phases of precipitation particles and detection of light rainfall to a resolution of  $0.2 \text{ mm h}^{-1}$ . The conical-scanning GPM Microwave Imager (GMI) consists of more frequency channels (10–183 GHz) than the TRMM Microwave Imager (TMI) for light-intensity rainfall and snow detection (Hou et al. 2014). Furthermore, satellite revisit time has been reduced from 11–12 h (TRMM satellite) to less than 3 h (GPM), and spatial coverage of the *Core Observatory* has increased from latitudes  $35^{\circ}\text{S}$ – $35^{\circ}\text{N}$  to  $65^{\circ}\text{S}$ – $65^{\circ}\text{N}$ , respectively (Hou et al. 2014). In addition to the *Core Observatory*, the increasing number of passive microwave sensors within the GPM constellation allows for more frequent sampling and cross calibration of sensors, thus increasing the spatial and temporal resolution of the gridded Integrated Multisatellite Retrievals for GPM (IMERG) precipitation product to  $0.1^{\circ}$  and 30 min compared with  $0.25^{\circ}$  and 3 h in TMPA (Huffman et al. 2015a), respectively.

Initial comparative evaluations of IMERG Day-1 and TMPA against rainfall gauges under different climatic and topographic conditions have confirmed the expected improvements of GPM. Prakash et al. (2016) have demonstrated higher estimation accuracy of IMERG over TMPA for heavy monsoon-type rainfall at a daily scale, although agreement with gauges was found to be lowest in orography-dominated regions (Prakash et al. 2017). Across China and the Tibetan Plateau, IMERG yields better statistical and hydrological performance than TMPA at (sub) daily scales and exhibits better detection of rainfall intensities than its predecessor, although overestimation of rainfall in dry regions persists (Guo et al. 2016; Ma et al. 2016; Tang et al. 2016; Li et al. 2016; Chen and Li 2016). Sharifi et al. (2016) found IMERG to perform better statistically than TMPA across a range of topographic and climatic conditions, with rainfall detection being improved in areas with orographic precipitation. Last, IMERG was found to overestimate (underestimate) drizzle (heavy rainfall frequency) compared to radar quantitative precipitation estimation (QPE) over the United States (Tan et al. 2016).

Early experiences indicate the potential improvements of IMERG compared to TMPA; however, performances vary substantially across different topographic and climatic conditions, especially in orography-dominated regions. Furthermore, there is to date a lack of IMERG evaluations in tropical mountain regions that are characterized by complex and variable precipitation patterns with strong orographic effects. Therefore, the objective of this study is to evaluate the performance of IMERG in comparison with TMPA against rain gauges during the first 17 months of GPM observations (from March 2014 to August 2015) to characterize the impact of rainfall regimes, as well as climatic and topographic conditions on IMERG estimates at different spatiotemporal scales. Following a description of the hydrometeorology of the study area, the satellite and rain gauge datasets are outlined (section 2). Section 3 introduces the evaluation methodology with a focus on rainfall detection, QPE, rainfall probability distribution, and spatial correlation. Results are presented for climatically homogeneous regions in section 4 and discussed in section 5. Last, critical findings are summarized in section 6.

## 2. Data

### a. Study area: Precipitation patterns of the tropical Andes

The study area (Fig. 1) extends from 2°N to 18.5°S and from 68.5° to 82°W (approximately 1 500 000 km<sup>2</sup>), covering a climatically diverse region, from northern Ecuador to the central Andean plateau (Altiplano), that is dominated by the tropical Andes, which results in extreme east–west precipitation gradients. Spatiotemporal precipitation patterns are controlled by the bi-annual migration of the intertropical convergence zone (ITCZ), El Niño–Southern Oscillation (ENSO), and the cold von Humboldt current in the Pacific Ocean, as well as the Andes mountains and the Amazon basin (Boers et al. 2013; Lavado-Casimiro et al. 2012; Vuille et al. 2000). Easterly trade winds resulting from the southerly position of the ITCZ during the monsoon season transport moist air from the tropical Atlantic over the Amazon basin (Boers et al. 2013) and are blocked by the topographic barrier of the tropical Andes (Romatschke and Houze 2010). The deflection of air moisture to the southeast gives rise to the South American low-level jet (SALLJ) that transports air moisture along the eastern Andes into the La Plata basin (Boers et al. 2013). The strong topographic gradients and easterly trade winds along the eastern flanks of the Andes also result in pronounced orographic effects (Espinoza et al. 2015; Espinoza Villar et al. 2009; Bookhagen and Strecker 2008). These, in turn, result in

deep convection (Romatschke and Houze 2010) and thereby highly intermittent spatiotemporal precipitation patterns with steep precipitation gradients of up to 190 mm km<sup>-1</sup> (Espinoza et al. 2015).

Given its hydrometeorological complexity, the region was divided into six subregions using the classification of Zulkafli et al. (2014), which identifies areas with distinct precipitation regimes based on topography and climate (Fig. 2). In the current study, the Amazon sub-Andes have been defined as 500–1500 m MSL in order to permit for adequate number of gauges (minimum 10) in each subregion. As shown in Table 1, the subregions differ in their climatic controls, resulting in a range of distinct precipitation regimes. Over the period 1998–2014, the mean annual precipitation varies considerably from the Pacific coast north of 4.5°S (PCN; approximately 1450 mm yr<sup>-1</sup>), over the Pacific coast south of 4.5°S (PCS; 300 mm yr<sup>-1</sup>), the western Andean slopes (AW, 575 mm yr<sup>-1</sup>), the eastern Andean slopes (AE, 1150 mm yr<sup>-1</sup>), the sub-Andes of the upper Amazon basin (ASA; 3500 mm yr<sup>-1</sup>), and at the Amazon lowlands (AL; 2375 mm yr<sup>-1</sup>) (Manz et al. 2016).

### b. Rain gauge data

Subhourly precipitation records were obtained from 302 rainfall gauges (Fig. 1) from the Instituto Nacional de Meteorología e Hidrología (INAMHI; 55 gauges), Empresa Pública Metropolitana de Agua Potable y Saneamiento de Quito (EPMAPS; 22 gauges), Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI; 186 gauges), and Iniciativa Regional de Monitoreo Hidrológico de Ecosistemas Andinos (iMHEA; 41 gauges). All gauges are tipping-bucket rain gauges and adhere to World Meteorological Organization (WMO) instrumentation standards (WMO 2014). While minor differences between the providers in terms of preprocessing and temporal interpolation may occur, these are thought to be negligible for the purposes of this evaluation study as the original datasets were aggregated to hourly rainfall accumulations for the assessment period from 1 April 2015 to 31 August 2015. Hourly rain gauge data were quality-controlled using the protocol defined by Shen et al. (2010), consisting of a check for unsupported extremes as well as internal and spatial consistency checks. This resulted in 0.64% of hourly measurements being removed as a result of the extremes check and a further 0.01% as a result of the consistency checks. Finally, 3-hourly and daily average rainfall rates (mm h<sup>-1</sup>) were computed in order to evaluate the satellite precipitation data at different temporal scales.

### c. Satellite data: TMPA V7

TMPA version 7 (V7), also known as TRMM 3B42 based on its algorithm names (hereafter TMPA), is a

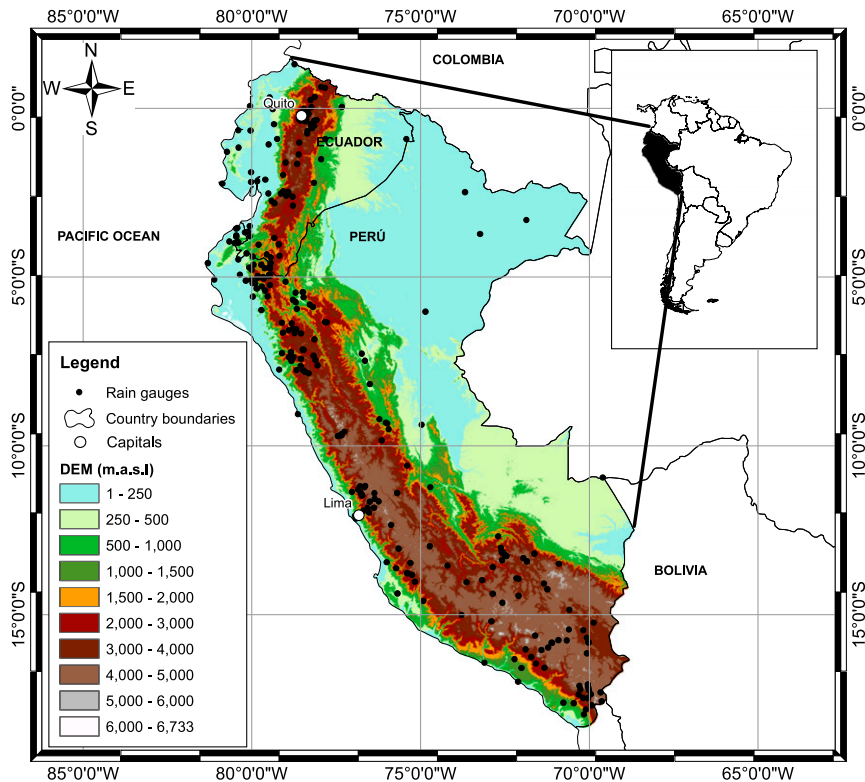


FIG. 1. Topographical map of the study area showing elevation and rain gauges as black dots.

precipitation dataset derived from multiple microwave (MW) and infrared (IR) sensors placed on low-Earth-orbit (LEO) satellites. Observations from MW and IR sensors are converted to precipitation estimates and intercalibrated and combined producing real-time (RT) estimates; finally, resulting estimates are adjusted with rain gauge data generating the TMPA “Research” version (Huffman et al. 2010, 2007). The TRMM satellite started its terminal phase in October 2014 with both the TRMM PR and the TMI having shut down on 8 April 2015; however, the demise of the TRMM satellite itself does not substantially affect the production the TMPA Research version over land (Huffman et al. 2015a). TMPA will be produced until approximately mid-2017 (Huffman et al. 2015a). Hence, TMPA was considered a suitable benchmark for comparative evaluation of IMERG. TMPA at its native resolution of  $0.25^{\circ}/3$  h was resampled to closed 3-h periods (e.g., 0000–0300 UTC) and for comparison to IMERG.

#### d. Satellite data: IMERG Day-1

The IMERG product provides high-resolution precipitation estimates by combining various passive microwave (PMW) and IR sensor measurements. This process is described in detail in Huffman et al. (2015b) and is briefly summarized as follows.

Analogous to the TMPA algorithm, the GMI is calibrated to the DPR and the resulting combined instrument (GCI) is used as a calibration standard for other PMW sensors in the GPM constellation. Precipitation estimates are derived from the PMW sensors using the 2014 Goddard Profiling Algorithm (GPROF2014). GPROF2014 relies on external radar and PMW observations for calibrating PMW measurements and is set to be replaced by GPROF2015 in due course, which uses GCI instead. In contrast to TRMM, both DPR and GCI are available in real time, allowing for the same calibrating sensors across all IMERG runs.

All calibrated PMW estimates within the GPM constellation are gridded to  $0.1^{\circ}/30$  min, prioritizing canonical-scan radiometers over cross-track scanners. Geosynchronous IR (geo-IR) measurements are converted to precipitation estimates using the PERSIANN–Cloud Classification System (CCS) recalibration scheme. Herein regional cloud patch groups are defined and precipitation is assigned to each of these based on an LEO PMW precipitation training set. Next, PMW and IR estimates are combined to create half-hour precipitation estimates using the CMORPH Kalman filter (KF) Lagrangian time interpolation scheme. In this approach, PMW estimates of instantaneous precipitation are propagated from observation to

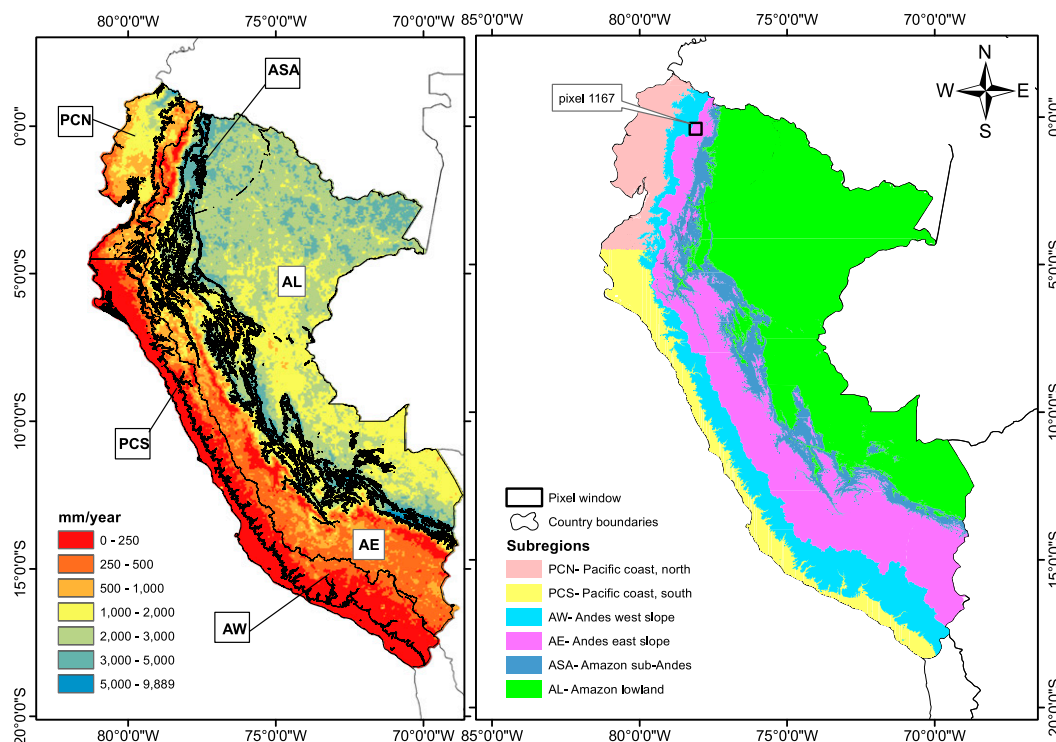


FIG. 2. (left) Mean annual rainfall climatology (1998–2014) derived from the TRMM PR (Manz et al. 2016). (right) Spatial definition of the climatic subregions. The rectangle in northern Ecuador outlines the location of pixel 1167, analyzed in detail in Fig. 11.

analysis time using cloud motion vectors, which were derived by correlation of spatially lagged consecutive geo-IR images.

Finally, a monthly satellite–gauge combination is performed akin to the approach adopted in TMPA, where multisatellite and gauge fields are combined using inverse error variance weighting. These bias-adjusted estimates are redistributed at the half-hourly scale for the “Final” product, which is distributed 2–4 months after measurement (Huffman et al. 2015a). In this evaluation study, the IMERG Day-1 Final run dataset (hereafter referred to as IMERG) is aggregated for the assessment period from its native resolution 0.1°/30 min to 0.25°/3 h and 0.25°/day for evaluation against TMPA.

### 3. Methods

#### a. Rainfall estimation: A problem of spatiotemporal scales

The selection of the spatiotemporal scale when evaluating precipitation products is of high importance given the nonlinear structure of precipitation in space and time. The spatiotemporal resolution of rain gauges (point-scale, short-term accumulation) profoundly differs from that of SPPs (gridcell average, temporal average of instantaneous measurements), which in turn vary internally (e.g., TMPA: 0.25°/3 h, IMERG: 0.1°/1 h), a problem that has received extensive attention (e.g., Wang and Wolff 2010; Villarini et al. 2008; Ciach and Krajewski 1999). Often all gauges within a satellite pixel are averaged

TABLE 1. Criteria used to define subregions. Parameter  $\mu$  represents the mean annual precipitation as estimated from merged TRMM PR and gauge measurements (Manz et al. 2016), and  $N$  indicates the number of rain gauges inside each subregion.

No.	Subregion	Elevation (m MSL)	Climate driver	Rainfall regime	$\mu$	$N$
PCN	Pacific Coast (N)	0–1500	ITCZ	Wet (Dec–May), dry (Jun–Nov)	1450	34
PCS	Pacific Coast (S)	0–1500	Von Humboldt, ITCZ	Wet (Dec–May), dry (Jun–Nov)	300	40
AW	Andes (W)	>1500	Elevation, ITCZ	Wet (Dec–May), dry (Jun–Nov)	575	123
AE	Andes (E)	>1500	Elevation, ITCZ, orography	Weak seasonality (drier JJA)	1150	77
ASA	Amazon sub-Andes	500–1500	Orography, ITCZ, SAMS, SALLJ	Weak seasonality (drier JJA)	3500	13
AL	Amazon lowlands	0–500	ITCZ, SAMS, trade winds	Weak seasonality (drier JJA)	2375	15

and the gauge average is compared to the SPP pixel. With increasing spatial and temporal aggregation these differences become less relevant; however, spatiotemporal aggregation implies the averaging of zero rainfall areas or periods with variable positive rainfall intensities. While rainfall occurrence is often represented by a binary (Bernoulli) distribution, positively skewed rainfall intensities are frequently modeled as a Gamma distribution, despite being subject to “heavy tails” (Tarnavsky et al. 2012; Wilson and Toumi 2005). Averaging zeros and nonzeros combines these into a single statistical distribution with much less variation than the intensity alone. Given these statistical properties, the performance of satellite rainfall products can be expected to improve with increasing spatiotemporal aggregation without any improvements in estimation skill. By corollary, it is important to understand how precipitation estimation deteriorates with increasing resolution. However, this general behavior is further complicated in regions where low and irregular gauge network density is combined with high precipitation variability at the subgrid scale, such as tropical mountain regions. For instance, in the current study, the number of gauges varies substantially across the region (Fig. 3) with only 9.6% and 24% of all satellite pixels containing more than one gauge at 0.1° and 0.25°, respectively. Hence, in order to comprehensively evaluate TMPA and IMERG against rain gauge observations despite the differences in spatiotemporal scale (resolution), three separate spatiotemporal scales were selected for the assessment and higher-resolution products were aggregated to the respective scales:

- 1) 0.1°/1 h: This is the highest available spatiotemporal resolution given the IMERG spatial resolution (0.1°) and the gauge time step (1 h). IMERG and spatial gauge averages were compared at this scale.
- 2) 0.25°/3 h: This is the TMPA V7 native resolution. Gauges and IMERG were aggregated and all three products were compared at this resolution.
- 3) 0.25°/1 day: This is a common resolution used of hydrological simulation based on SPPs. All three products were space–time averaged to this scale.

At all scales rainfall products were evaluated as rainfall rates ( $\text{mm h}^{-1}$ ). Furthermore, especially at finer spatial scales, there are insufficient gauges available to overcome the statistical impacts of the point–area difference. Therefore, this study further investigates the impact of scale in relation to the number of rain gauges available (section 3c).

### b. Evaluation metrics

In the main ground validation, all gauges within the respective satellite pixel are averaged to evaluate the IMERG and TMPA pixels with ground observations of

the same spatial support. The evaluation is split into four categories:

- 1) rainfall detection—empirical rainfall occurrence frequency and detection indicator scores,
- 2) quantitative errors in the estimation of rainfall intensities,
- 3) comparison of the cumulative probability distributions of the rainfall intensities, and
- 4) comparison of spatially averaged subregion time series across the assessment period.

In order to perform ground validation of satellite precipitation data, it is necessary to assume a minimum threshold in the intensities recorded by the rain gauges (Wang et al. 2008). Because of the high frequency of low rainfall intensities in some parts of the tropical Andes (e.g., Padrón et al. 2015), a threshold corresponding to one single record of the highest-resolution gauges available in the network ( $0.1 \text{ mm h}^{-1}$ ) has been found to be suitable to assess the entire range of rainfall intensities without eliminating the lowest intensities or introducing assumptions as to their distribution. The empirical rainfall occurrence frequency (ROF) is expressed as

$$\text{ROF} = \frac{\sum_{i=1}^{n_{\text{tot}}} (p_{r,i} \geq 0.1)}{N_{\text{tot}}}, \quad (1)$$

where  $i = 1, \dots, n_{\text{tot}}$  is the number of time steps (total of  $N_{\text{tot}}$  time steps),  $p$  is the precipitation intensity at a particular grid cell, and  $r$  is the aggregated spatial resolution (0.1° or 0.25°).

Rainfall detection is assessed by three categorical error scores: the accuracy index (ACC; Duan et al. 2015), the probability of detection (POD), and the false alarm ratio (FAR):

$$\text{ACC} = \frac{H + C}{N_{\text{sync}}}, \quad (2)$$

$$\text{POD} = \frac{H}{H + M}, \quad (3)$$

$$\text{FAR} = \frac{F}{F + H}, \quad (4)$$

where  $H$  is the number of rainfall time steps correctly detected (hits),  $C$  is the number of time steps correctly identified as nonraining (correct zeros),  $M$  is the number of raining time steps missed by the SPP (misses),  $F$  is the number of time steps falsely identified as raining (false alarms), and  $N_{\text{sync}}$  is the total number of synchronous measurements. ACC represents the fraction of time steps correctly classified (score ranges 0–1, perfect score of 1), POD represents the fraction of rain occurrences

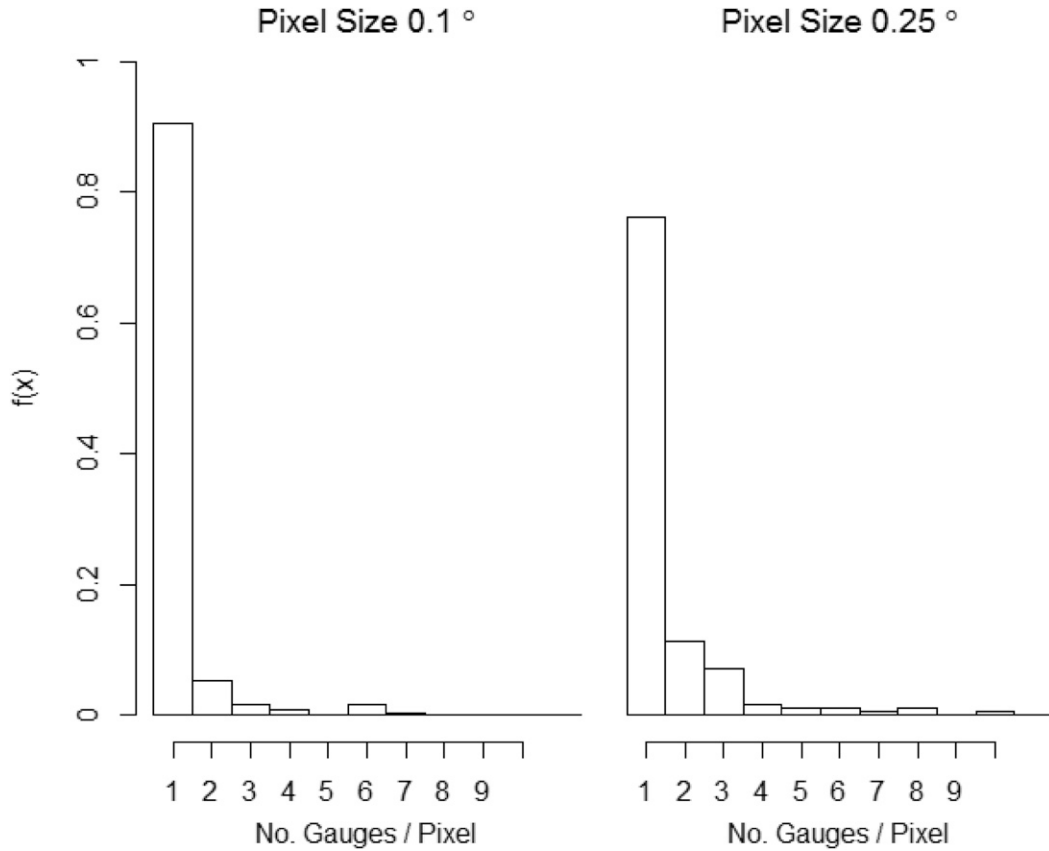


FIG. 3. Number of gauges by satellite pixel at 0.1° and 0.25°, expressed as the empirical probability density function  $f(x)$ . The histogram only shows pixels with at least one gauge, which are used throughout the analysis of this study. Ungauged pixels are not considered here.

correctly detected (score ranges 0–1, perfect score of 1), and FAR represents the fraction of detected rainfall occurrences that were false alarms (score ranges 0–1, perfect score of 0). To evaluate the impact on the rainfall detection scores introduced by the limited satellite rainfall detection ability as well as due to averaging of zero and nonzero rainfall intensities across the spatial support of the SPPs, detection statistics were computed for precipitation thresholds from 0 to 10 mm h<sup>-1</sup>. Hereby, rainfall rates of both the SPP and the gauge average ground reference rainfall below the threshold are treated as zero rain. The threshold is then iteratively raised for both the SPP and the ground reference rainfall and the detection scores are recomputed accordingly.

In order not to double count rainfall detection errors, quantitative rainfall rate estimation errors are only computed for time steps where rainfall is accurately detected, as previously proposed by Tang et al. (2015) and Tan et al. (2016). Time steps where rainfall is falsely detected or missed are omitted prior to computing the following statistical metrics: Pearson correlation coefficient (CC), root-mean-square error

(RMSE), relative RMSE (rRMSE), and percentage bias (PBIAS):

$$CC = \frac{\sum_{i=1}^n (S_i - \bar{S})(G_i - \bar{G})}{\sqrt{\sum_{i=1}^n (S_i - \bar{S})^2} \sqrt{\sum_{i=1}^n (G_i - \bar{G})^2}}, \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - G_i)^2}, \quad (6)$$

$$rRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (S_i - G_i)^2}}{\bar{G}}, \quad (7)$$

$$PBIAS = \frac{\sum_{i=1}^n (S_i - G_i)}{\sum_{i=1}^n G_i} \times 100\%, \quad (8)$$

where  $S_i$  and  $G_i$  are, respectively, the satellite and gauge rainfall intensity at time step  $i$ ;  $\bar{S}$  and  $\bar{G}$  are their

corresponding arithmetic means over the assessment period; and  $n$  is the total number of time steps. While RMSE expresses random error in absolute terms (i.e.,  $\text{mm h}^{-1}$ ) and, therefore, will likely result in elevated errors in wet regions, rRMSE expresses random error relative to the mean precipitation rate.

Detection and quantitative estimation errors evaluate whether SPPs and gauges agree at the same time step. While SPP estimation accuracy for individual time steps might be low, SPPs may still be able to characterize the rainfall intensity distribution over the assessment period. For this purpose, the empirical cumulative distribution function [CDF;  $F(x)$ ] is determined for each SPP and compared to the respective gauge CDF. The CDF can be represented as discrete percentiles, and the ratio of the percentiles across the entire CDF shows how well the SPPs estimate the gauge rainfall distribution:

$$\text{CDF ratio} = \frac{p_{S,j}}{p_{G,j}}, \quad (9)$$

where  $p$  is the rainfall intensity corresponding to the percentile  $j$ .

### c. Impact of spatial scale

A fundamental difference between IMERG and TMPA is the improved native resolution ( $0.1^\circ/30$  min compared to  $0.25^\circ/3$  h), resulting from an increased number of available satellites, especially PMW sensors, in the GPM Constellation and the approach used for combining them. While the higher-resolution results in a finer definition of precipitation fields and intensity gradients, a high density of rain gauges is required to provide areal mean ground observations to evaluate individual grid cells. As shown in Fig. 3, the gauge network in Ecuador and Peru does not provide high-density coverage: over 90% of gauged IMERG pixels only contain a single reference gauge. This implies that the majority of IMERG pixels that are evaluated in this study may be subject to substantial point-area difference effects (Ciach and Krajewski 1999). The impact of the scale difference between SPPs and corresponding ground-based estimates was investigated using the following analysis: 1) changes in systematic error (PBIAS) using a spatial bootstrap subsampling approach and 2) evaluation of the subgrid scale variability of a single satellite pixel as well as 3) spatial correlation structure of the different rainfall products across the study domain.

In the bootstrap analysis, pixels containing more than one gauge are subsampled by removing a single gauge at a time and spatially averaging the remaining  $n - 1$  gauges across the pixel. PBIAS is then computed for the pixel using Eq. (8) based on the gauge average and the

SPP estimates. The removed gauge is then replaced and the process is repeated  $n$  times, resulting in a PBIAS score associated with each removed gauge. The mean of the  $n$  PBIAS scores is then computed to obtain a pixel-wide gauge average for each pixel. Comparing the PBIAS scores of the bootstrap approach against those of the gauge-averaging approach allows for interpreting the impact of removing individual gauges on the pixel-wide gauge average. This offers insight into the dependence of systematic errors in the SG evaluation on the gauge density.

The impact of spatial scale was further analyzed by focusing specifically on the individual  $0.25^\circ$  pixel with the highest number of gauges, hereafter referred to as pixel 1167 (see Fig. 2 for location). This pixel covers the mid-Andean water divide and includes both the west and east Andean slopes. For pixel 1167, scatterplots of synchronous measurements of 1) individual gauges within that pixel, 2) the spatial gauge average across the pixel, and 3) the SPP estimate (IMERG or TMPA) are presented for the different space–time scales to demonstrate how the internal variability at the subgrid scale affects the match of SPP and gauge observations.

Last, the ability of SPPs to capture the geographical structure of precipitation fields is evaluated by comparing the spatial correlation structure between satellite and pixel-average gauge rainfall. The CC is calculated for any pixel pairs across the entire assessment period. The CC results are categorized by the distance between the pixel pairs, averaged across bins of 27 km (which approximately corresponds to  $0.25^\circ$  across the study region) and presented as a spatial correlogram.

## 4. Results

### a. Mean rainfall fields

The spatial mean precipitation across the assessment period (Fig. 4) shows that IMERG contains an improved definition of the orographically enhanced precipitation hot spots along the eastern Andean flanks and Amazonian sub-Andes in Peru between  $7^\circ$  and  $14^\circ\text{S}$  compared to TMPA. Furthermore, because of the improved spatial resolution of IMERG ( $0.1^\circ$ ), the steep precipitation gradient from the drier Andean highlands to the orographically enhanced sub-Andes is better defined, while it was previously “smoothed out” over a larger spatial distance by TMPA. Isolated intra-Andean valleys, for instance, those between  $9^\circ$  and  $12^\circ\text{S}$  with elevated precipitation levels, are represented with a far smaller spatial footprint by IMERG than TMPA. In terms of the mean satellite–gauge (S–G) difference at the aggregated



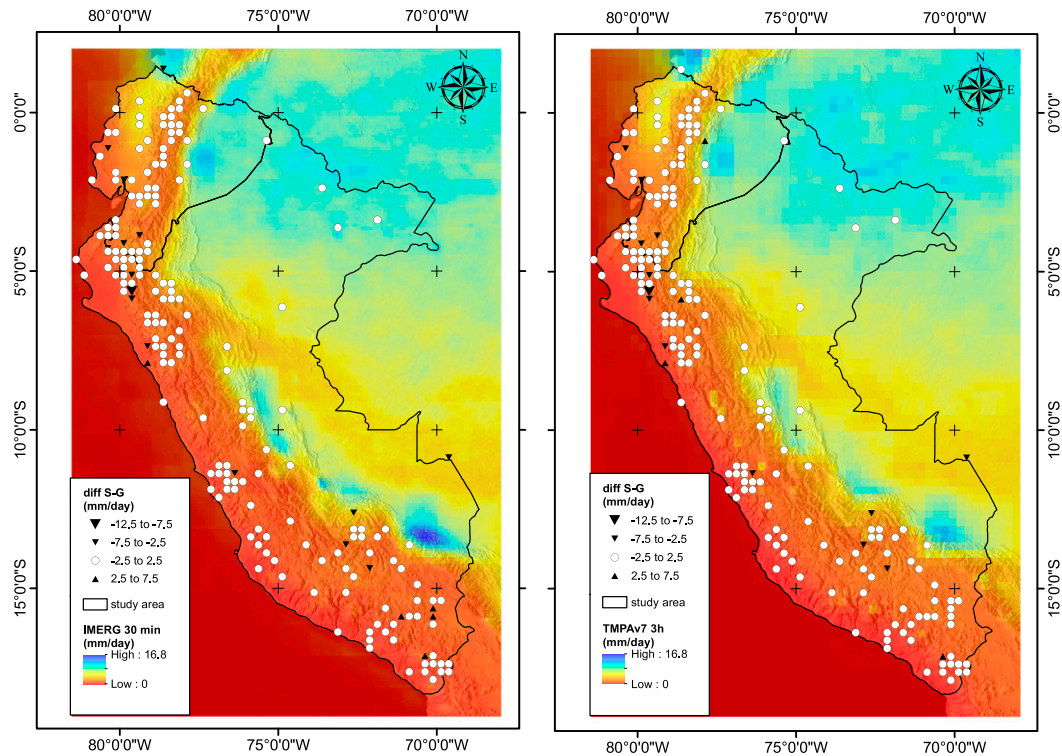


FIG. 4. Mean precipitation rate ( $\text{mm day}^{-1}$ ) of (left) IMERG and (right) TMPA across the assessment period (from April 2014 to August 2015). S-G differences between the SPPs and gauge averages are superimposed at the aggregated  $0.25^\circ$ /daily scale.

$0.25^\circ$ /daily scale (Fig. 4), IMERG and TMPA show a very similar pattern with good S-G agreement (within  $2.5 \text{ mm day}^{-1}$ ) at most locations. Both SPPs show a cluster of pixels where the gauge mean is underestimated in the Piura region (northern Peru/southern Ecuador), while for IMERG a smaller cluster of pixels overestimate gauges in the Altiplano region (southeastern Peru). As a result, across the study region TMPA correlates better with mean pixel gauge averages ( $\rho_{\text{TMPA}} = 0.80$ ) than IMERG ( $\rho_{\text{IMERG}} = 0.76$ ).

*b. Subregional temporal patterns*

As a second step, the spatial mean precipitation time series of gauge observations, IMERG, and TMPA (at 10-day accumulations) are compared for each subregion (Fig. 5). The results show substantial regional variations. Subregions PCN, AW, AE, and AL show good agreement between gauge observations and IMERG and TMPA. Overestimation of individual rainfall peaks by TMPA, especially in the AE slopes, is notably reduced by IMERG. Underestimation of individual rainfall peaks by both SPPs is observed particularly during the austral winter season (June–August). In the dry PCS region, both IMERG and TMPA systematically overestimate the mean as well as rainfall peaks proportionally

to the rainfall magnitude, although overestimation of peaks is stronger by IMERG than TMPA. Subregion ASA differs in that the mean is underestimated substantially by both SPPs. Here rainfall peaks are generally underestimated, although IMERG returns higher estimates than TMPA with occasional overestimation by IMERG. However, this observation should be considered in the context of the low and irregular (and thereby potentially unrepresentative) gauge coverage in ASA.

*c. Rainfall detection and occurrence frequency*

As shown in Fig. 6, at its native resolution IMERG underestimates both the median and the variability of the ROF observed by the gauges in all climatic subregions except for PCS and AL. With increasing spatio-temporal aggregation, IMERG estimates are elevated, thus leading to better estimation of subregions previously underestimated (PCN, AW and AE), but extensive overestimation in PCS and AL. In contrast to IMERG, TMPA shows systematically lower ROF, resulting in substantial underestimation in most regions, but better estimation of gauge ROF in PCS and AL. At the daily time step, IMERG and TMPA return comparable ROF results, although gauge ROFs are still overestimated

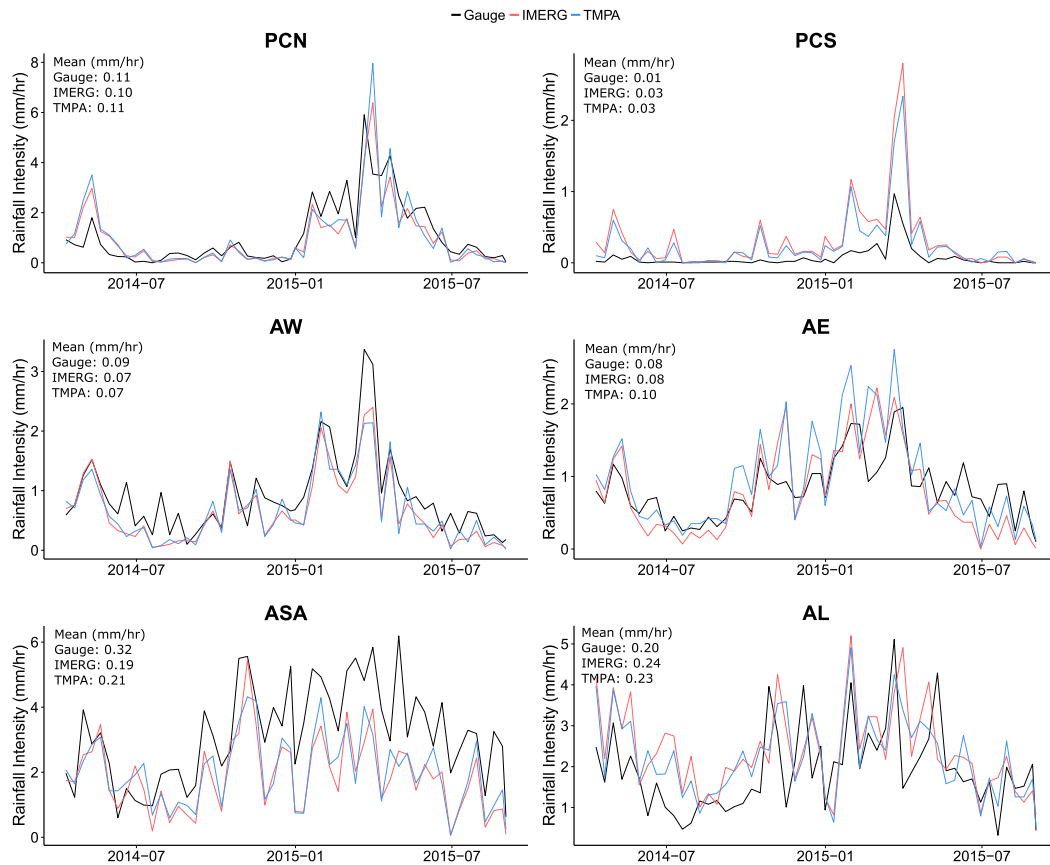


FIG. 5. Comparison of the regional mean precipitation time series at 10-day accumulations over the assessment period for gauges (black), IMERG (red), and TMPA (blue).

for PCS, while the median and variability are underestimated in ASA by both products.

Detection scores (Fig. 7) show superior performance of IMERG compared to TMPA in terms of higher POD and ACC and lower FAR for all scales and subregions. At  $0.25^\circ/3$  h IMERG performs substantially better than TMPA, while performance improves with increasing levels of spatiotemporal aggregation for both SPPs thereafter. In general, the detection ability of both SPPs weakens when the rainfall detection threshold is increased, resulting in decreasing POD and increasing FAR. This behavior can be explained in part by the reduction in sample size with increasing rainfall detection threshold: as all rainfall intensities below the threshold are set to zero, the total number of time steps with nonzero rainfall intensity is reduced. Hence, the number of rainfall events at higher thresholds decreases and failure to detect these is given proportionally higher weighting as per Eq. (3), resulting in a decreasing POD. The opposite applies to FAR: with increasing rainfall detection threshold, cases where both SPP and ground reference recorded a low rainfall intensity will no longer

be treated as a hit, but as correct zeros, such that with a constant rate of false alarms, the FAR will increase because of the reduction in hits. Similarly, with increasing rainfall thresholds, the number of zero rain events increases, leading to elevated ACC scores, as correct zeros become proportionally more frequent. However, this analysis shows a consistent pattern in that for all three scores and across most regions, IMERG returns systematically higher POD and ACC, but lower FAR scores for all thresholds compared to TMPA at the respective spatiotemporal scale. This suggests an improved rainfall detection ability by IMERG compared to TMPA.

In terms of regional differences, both products show weakest performance (low POD, high FAR) in PCS (Fig. 7). This can be attributed to the infrequent rainfall in this arid region, combined with, at times, elevated levels of humidity that do not result in precipitation. ACC shows very high values in this region, as a result of being dominated by correct zeros, which does not necessarily represent an improvement in terms of rainfall detection.

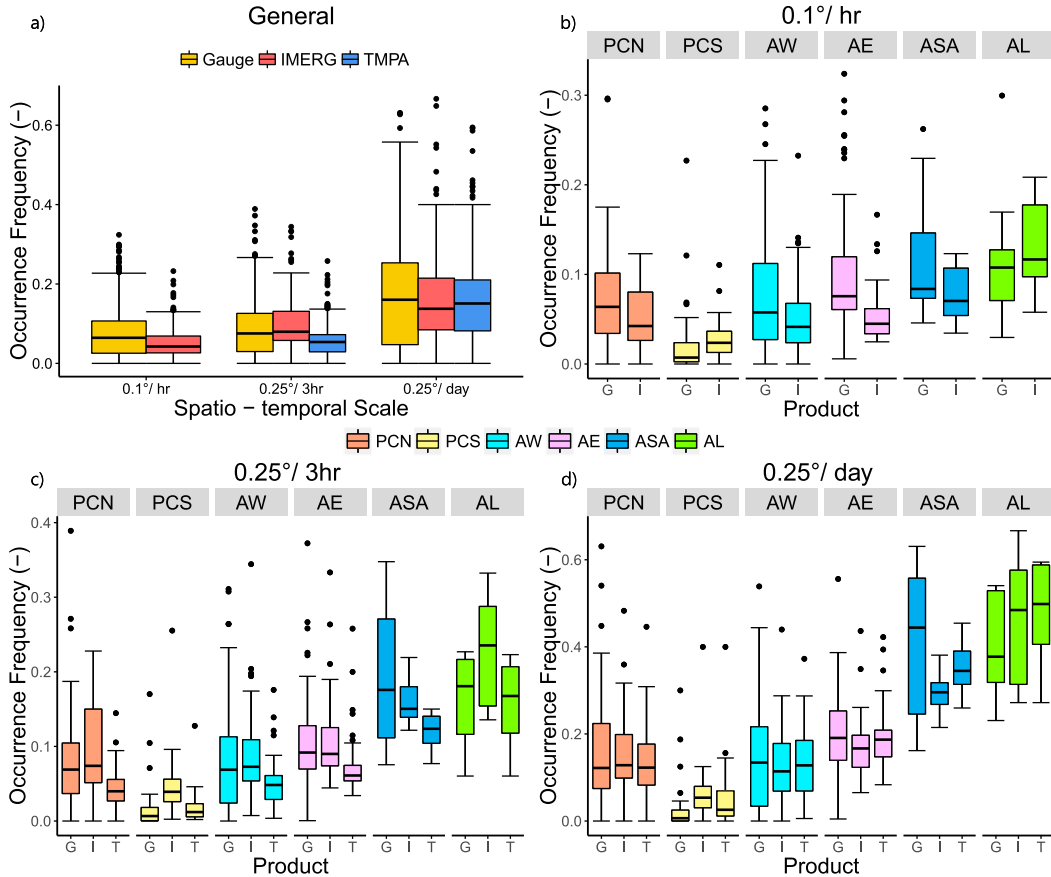


FIG. 6. ROF (a) for the entire study region and by subregion and product for three spatiotemporal resolutions: (b) 0.1°/1 h, (c) 0.25°/3 h, and (d) 0.25°/daily. In the subregional plots, *G* represents the gauge result, *I* the IMERG value, and *T* the TMPA value. For definition of the subregions, see section 2a.

d. Rainfall intensity errors

As shown in Fig. 8, the Pearson linear correlation coefficient increases with increasing spatiotemporal scale. Subregions experiencing high precipitation levels generally exhibit higher median correlation results with little spread, whereas in the dry PCS region, there is a wide spread of results. IMERG systematically outperforms TMPA, with improved satellite–gauge correlation being most pronounced in regions subject to high precipitation rates (PCN, ASA, AL) and less so in the arid subregion (PCS). RMSE scores show consistent reduction in random error by IMERG compared to TMPA across all subregions; however, RMSE scores of both SPPs are highest for those subregions experiencing highest levels of precipitation (i.e., ASA and AL). However, in agreement with the other detection and quantitative error statistics, rRMSE is most elevated in dry regions, especially PCS, showing an increase for IMERG over TMPA under these conditions. In terms of systematic error, TMPA overestimates gauge rainfall observations

(large positive PBIAS) in the Andean regions (AW and AE) and the PCN. In all three regions, PBIAS for IMERG is substantially reduced, suggesting large contributions of systematic error have been eliminated. Bias is highest in arid conditions with infrequent rainfall (PCS) for both SPPs, while it is generally lowest in the wet Amazonian regions (ASA and AL) where both return comparable results. In contrast to all other subregions, in PCS IMERG did not improve on TMPA with PBIAS for IMERG exceeding that of TMPA at 0.25°/daily, thus implying that in dry conditions IMERG estimation accuracy has not improved over TMPA at corresponding space–time scales.

e. Statistical probability distributions of rainfall intensities

With respect to the cumulative probability distribution of estimated precipitation intensities (Fig. 9a), both SPPs overestimate corresponding gauge quantiles across the entire intensity distribution and at all scales,

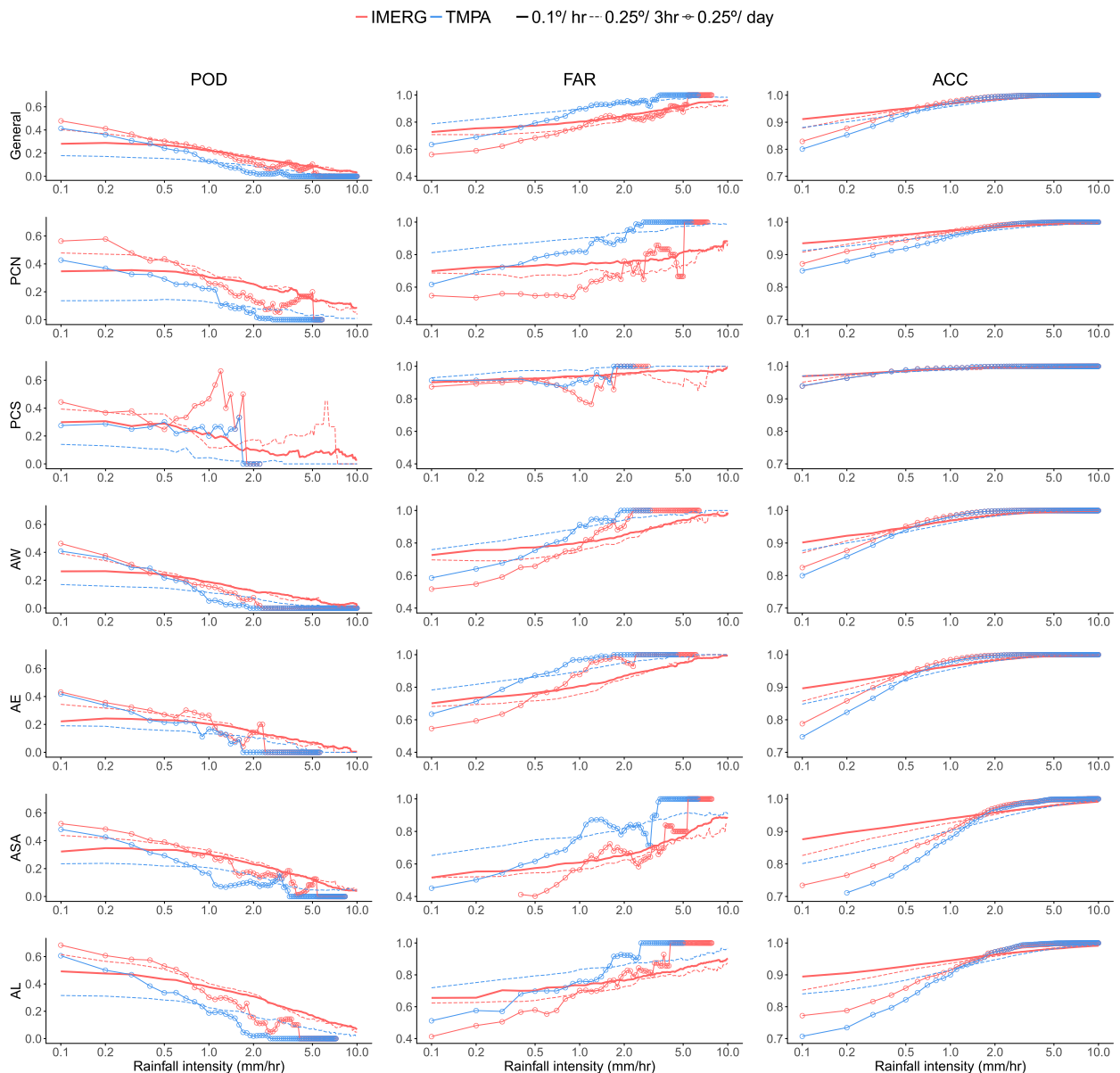


FIG. 7. Detection scores (POD, FAR, and ACC) for each spatiotemporal resolution for the (top row) entire study area and across the different subregions.

except for IMERG at 0.25°/3 h between the 40th and 90th percentiles. However, for identical spatiotemporal scales, IMERG returns a CDF ratio closer to 1.0 than TMPA. At their native resolution, both SPPs overestimate gauge quantiles substantially with a maximum of factor 2.8 (IMERG) and 2.4 (TMPA) at approximately the 30th percentile and decreasing thereafter. Unlike for TMPA, the CDF ratio for IMERG increases above the 90th percentile. As shown in the subregional plots (Figs. 9b–d), this increase was found to be isolated to the PCS subregion. This finding agrees with the

positive bias for IMERG observed in PBIAS (Fig. 8) and the time series analysis (Fig. 5), which showed strong overestimation of peak rainfall by IMERG in PCS. At the daily aggregation, both SPPs show gradually increasing, positive CDF ratios with a maxima of approximately 1.7 at the 99th percentile. However, for the majority of the cumulative probability distribution and especially below the 40th percentile, the CDF ratio for IMERG is far lower than that of TMPA, suggesting superior ability of IMERG in estimating the gauge rainfall probability distribution. Overestimation of the

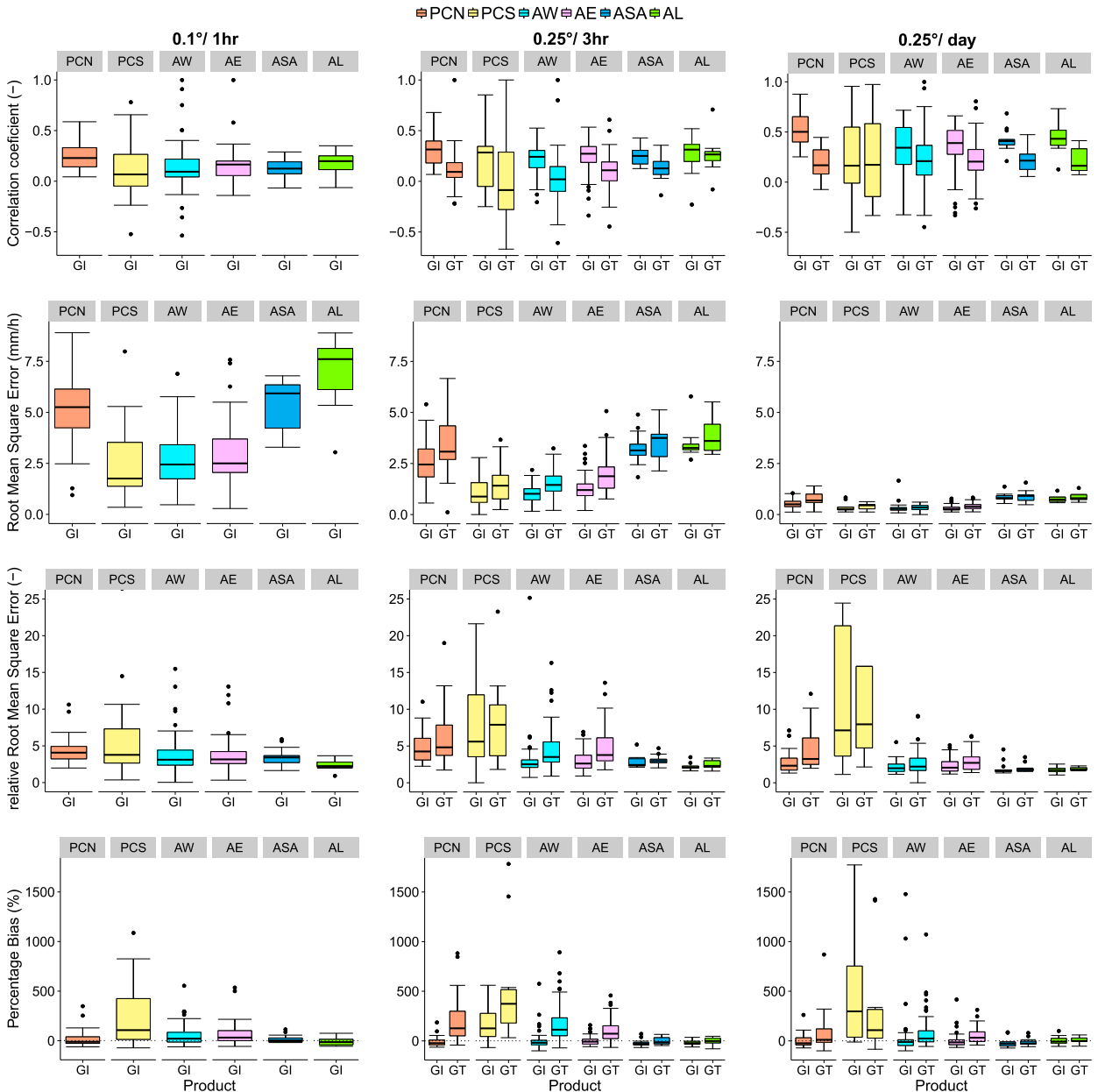


FIG. 8. Precipitation intensity estimation errors where precipitation is correctly detected by subregion and spatiotemporal resolution. GI represents the gauge – IMERG analysis, whereas GT represents the gauge – TMPA analysis. For definition of the subregions, see section 2a.

gauge quantiles by TMPA is most evident at the TMPA native resolution (0.25°/3 h) and most pronounced in ASA, AL, and PCN, that is, under conditions of high frequency and high total rainfall. At this resolution IMERG underestimates gauge rainfall in almost all subregions, especially between the 50th and 80th percentiles, which may be an artifact of spatiotemporal aggregation, as IMERG estimation accuracy is improved both at finer and coarser spatiotemporal scales. However, irrespective

of scale, medium to high quantiles are underestimated by IMERG for the high AL and ASA regions.

*f. Impact of spatial scale*

When comparing the bootstrapping to the standard pixel–gauge averaging, reductions in median percentage bias for IMERG at its native resolution (0.1°/1 h) over the Andean and Amazonian subregions (Fig. 10) are observed. On the other hand, in the PCS subregion bias

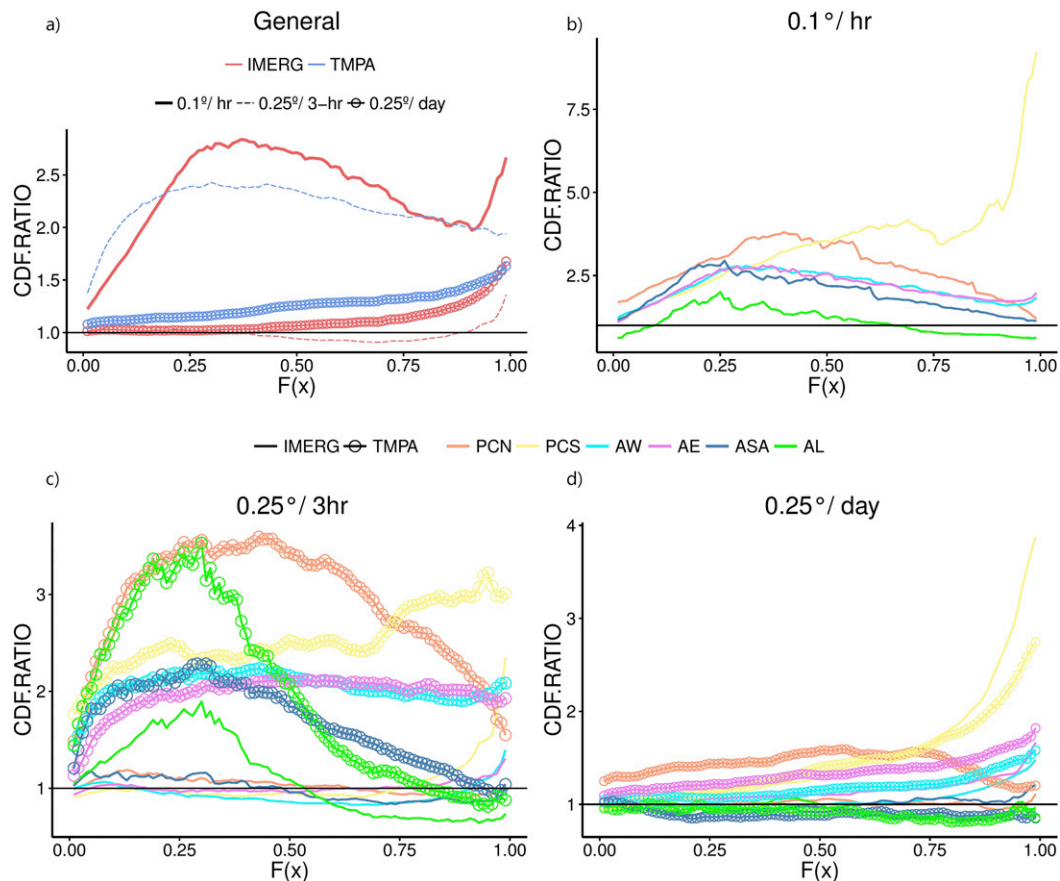


FIG. 9. CDF ratio [Eq. (9)] (a) for the entire study region and by subregion and product for three spatiotemporal resolutions: (b) 0.1°/1 h, (c) 0.25°/3 h, and (d) 0.25°/daily. For definition of the subregions, see section 2a.

increases for both SPPs at all space–time scales. The variations show the average response in the sensitivity of the bias calculation to a removal of only a single gauge at each pixel, highlighting the importance of the coverage and representativeness of the gauge network for SPP evaluation. However, the majority of pixels only contain a single gauge in both the pixel averaging and the bootstrapping approaches. Hence, the difference between the two methods is very low when averaged across all pixels, but will be larger locally, that is, for individual pixels containing more than a single gauge.

Focusing on the 0.25° pixel with the highest gauge density, Fig. 11 shows a large variation between the grid SPP estimates and individual gauge estimates within the pixel. While both IMERG and TMPA overestimate the spatial mean gauge rainfall by approximately factor two (IMERG 0.1°/1 h: 2.56, IMERG 0.25°/3 h: 2.17, TMPA: 2.78), the relationship between an individual gauge and the SPP estimates may even be negative, that is, gauge P46 (from  $-0.12$  to  $0.06$ ). This gauge is located on an east-facing slope in a north–

south-oriented, medium-elevation (2960 m MSL) valley discharging to the Amazon, whereas most gauges (labeled “JTU”) are clustered in a high-altitude region (above 4000 m MSL) to the west of the Antisana volcano (southwest of the pixel). For these gauges, the Antisana volcano acts as a barrier, restricting humidity transported by easterly trade winds from the Amazon basin. On the other hand, gauges located at lower elevations farther downstream are directly exposed to the higher humidity levels.

Finally, evaluation of the spatial correlation structure across the entire region (Fig. 12) suggests that, irrespective of the spatiotemporal scale, IMERG and TMPA estimate higher spatial correlation distances than the gauges. This implies the degree of spatial averaging (smoothing) is far higher in the SPPs than the gauge estimates. However, this behavior may also be attributed to the majority of ground-based pixels only containing a single gauge. Thus, these pixels represent point-scale rainfall, which is by nature more variable and has shorter spatial correlation distances than grid

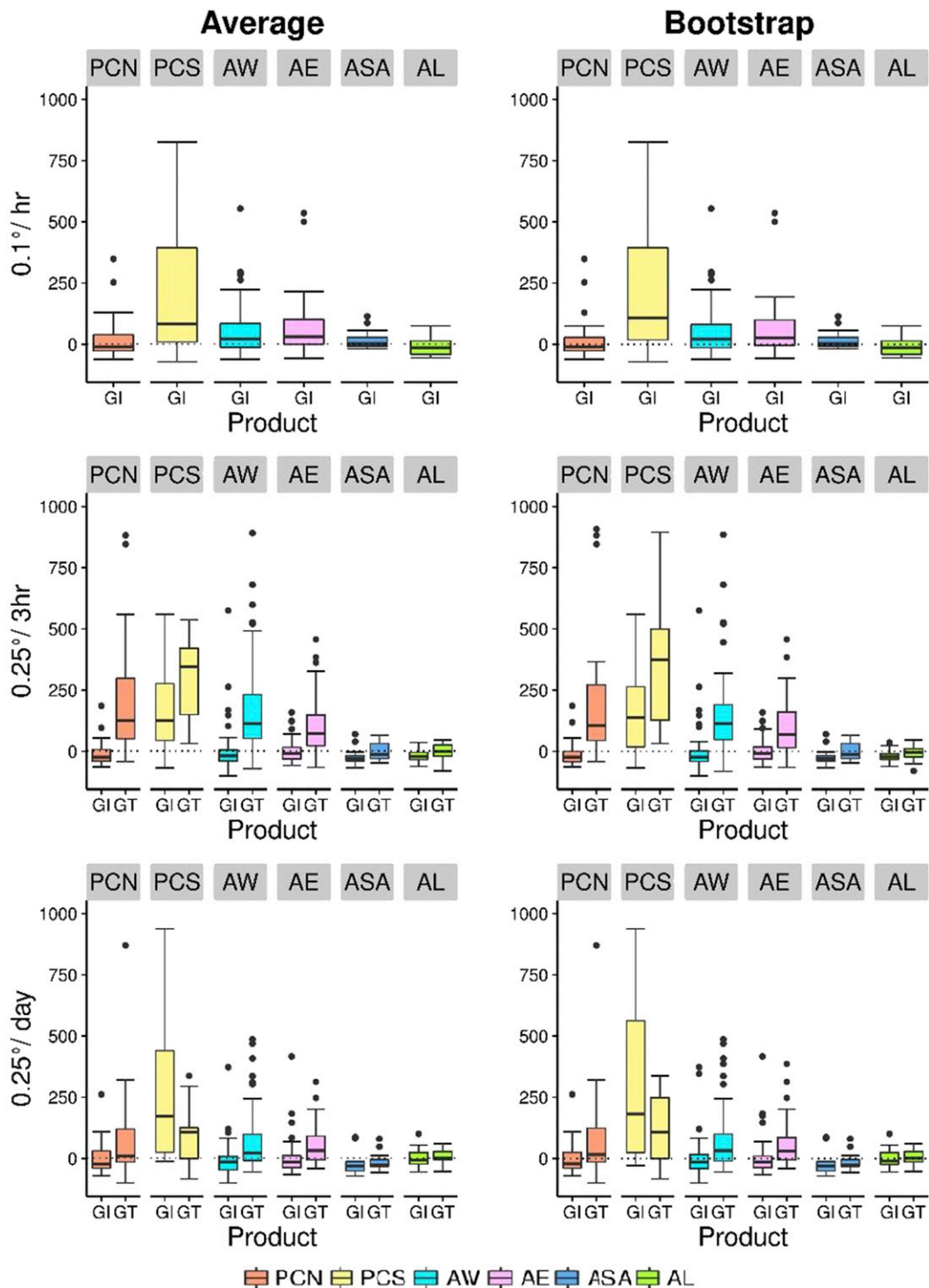


FIG. 10. Percentage bias as identified during (left) the standard pixel average analysis and (right) the bootstrap sampling analysis. GI represents the gauge – IMERG bias, whereas GT represents the gauge – TMPA bias. For definition of the subregions, see section 2a.

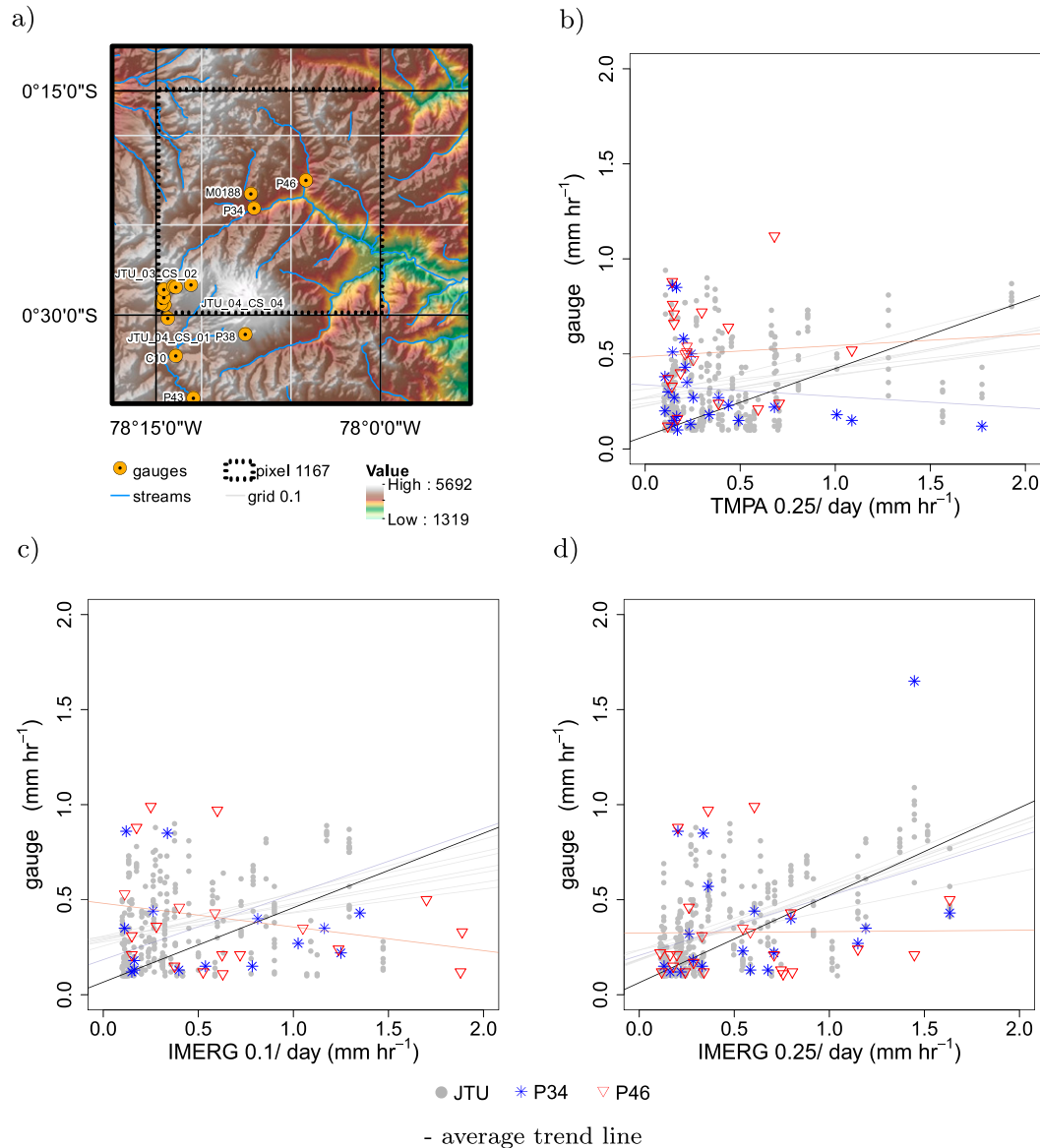


FIG. 11. Analysis of the impact of scale on a single satellite pixel (ID 1167): (a) the topography and gauge network as well as superimposed grids at  $0.1^\circ$  and  $0.25^\circ$  resolution and scatterplots of the individual gauges (and their areal grid mean) against the corresponding SPP estimate for (b) TMPA at  $0.25^\circ$ , (c) IMERG at  $0.1^\circ$ , and (d) IMERG at  $0.25^\circ$  with best-fit lines using linear regression. The gauge M0188 is not included in the scatterplots as no “hits” were observed.

average rainfall as reported by the SPPs. Analysis of the spatial correlation of pixels containing at least two gauges as shown in Figs. 12d–f) suggests similar spatial correlation results as the SPPs, thus supporting the notion that observed differences in spatial correlation can be predominantly attributed to the point-area difference in spatial support of the gauges versus the SPPs. Considering only the spatial correlation results for pixels with at least two gauges, IMERG and TMPA still overestimate spatial correlation with the differences between

the SPPs smaller than satellite–gauge differences in spatial correlation across all scales.

## 5. Discussion

Implications derived from the observations of IMERG performance in this study are summarized and compared to previous studies in Table 2 and discussed in this section. In terms of the comparison of IMERG against TMPA, no improvements are evident with respect to



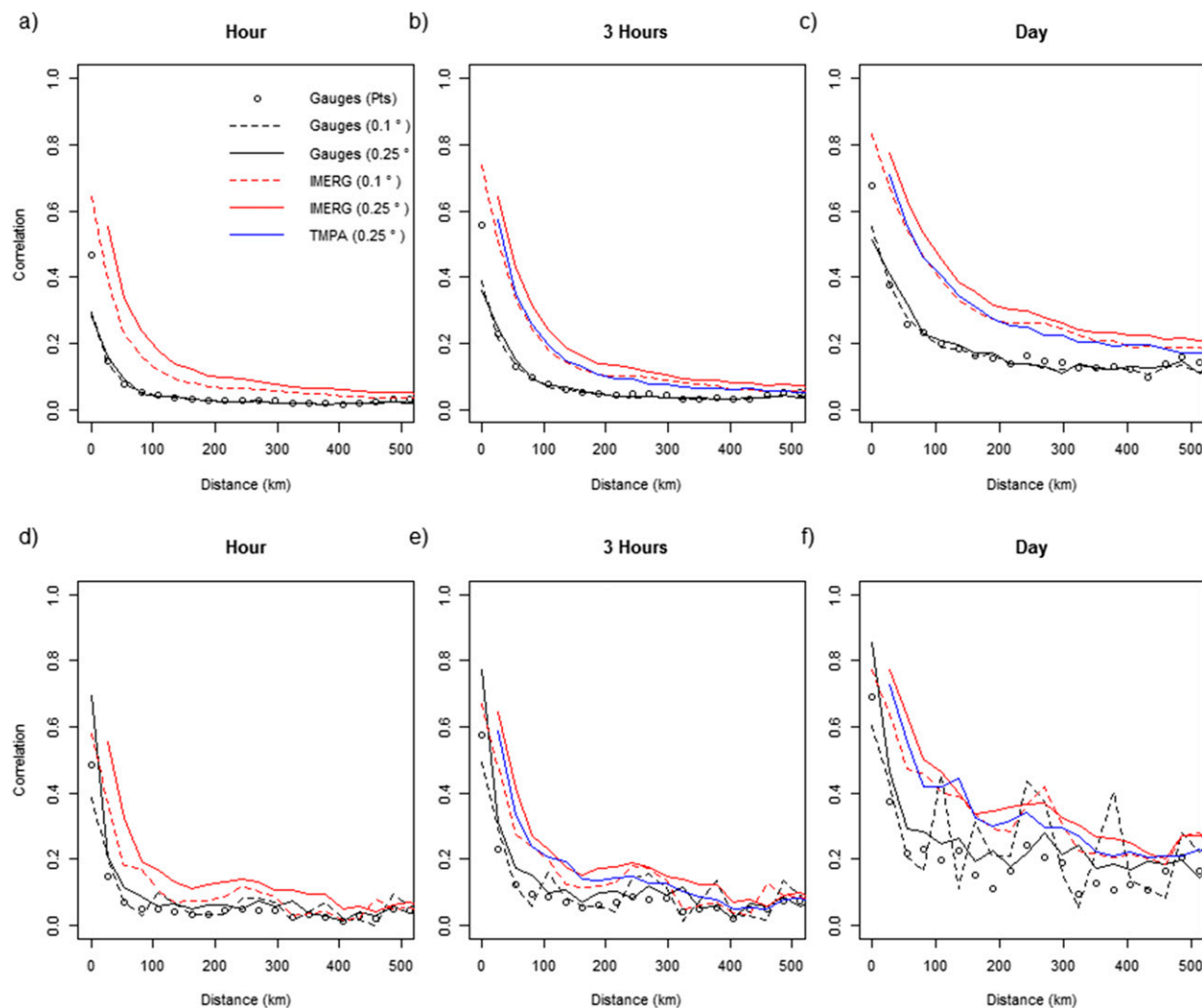


FIG. 12. Spatial correlation for each product at different temporal aggregations. (a)–(c) The spatial correlation considering all pixels inclusive of those with just one gauge per pixel for the hourly, 3-hourly, and daily scale, respectively. (d)–(f) The spatial correlation only for pixels with at least two gauges for the respective temporal scales.

mean rainfall across the study period, and TMPA even shows higher S-G correlation. However, separating rainfall detection and rainfall rate estimation accuracy reveals that IMERG has superior skill in estimating both of these precipitation components. The good agreement between TMPA and gauges in terms of the spatial mean rainfall field can be attributed to ROF being underestimated while rainfall intensities were overestimated. Furthermore, irrespective of the regionally differing precipitation regimes, a consistent observation is the reduction in false peaks by IMERG compared to TMPA, suggesting a better estimation accuracy of high rainfall intensities by IMERG. Overall, improvements are most pronounced in the high Andes, which experience a large fraction of low-intensity rainfall (Padrón et al. 2015), suggesting an improved light rainfall detection ability.

Despite improvements in estimating the majority of the rainfall intensity distribution compared to TMPA, IMERG markedly overestimates the high intensities, that is, the highest quantiles of the cumulative probability distribution. This observation has already been reported elsewhere (Sahlu et al. 2016) and is of high relevance, in particular, for applications focused on rainfall or hydrological extremes such as intensity–duration–frequency curves or hydrological flood simulation.

Additionally, IMERG continues to overestimate both the frequency of rainfall as well as rainfall intensities in extremely dry regions, as in the case of PCS in this study, confirming previous findings by Tang et al. (2016), Sharifi et al. (2016), and Guo et al. (2016). In this region, improvements of IMERG over TMPA are lowest by all

TABLE 2. Summary of IMERG performance observations, implications, and potential causes as well as agreement with previous studies.

Observation	Implication/potential cause	Previous studies
Overestimation of high rain intensities	High relevance for rainfall/ hydrology extremes	Sahlu et al. (2016)
Reduced rainfall detection for high rain rates	Potential limitation in detecting convective rainfall	Sahlu et al. (2016)
Overestimation of rainfall frequency in arid regions (frequent false alarms)	Incorrectly transformed IR retrievals; high rates of subcloud evapotranspiration	Tang et al. (2016); Sharifi et al. (2016); Guo et al. (2016)
High quantitative errors in rain intensity (arid regions)	—	Tang et al. (2016)

considered statistics. In particular, the rainfall detection analysis showed high frequency of missed rainfall and falsely detected rainfall. Such false positives can potentially stem from incorrectly transformed infrared retrievals when high cloud cover is falsely translated to rainfall. Similarly, high rates of subcloud evaporation may account for the discrepancy between gauge recorded rainfall and SPP estimates. Quantitative errors were generally highest in the dry PCS region for both TMPA and IMERG, which agrees with previous findings in arid regions of China by Tang et al. (2016).

For all spatiotemporal scales IMERG shows a lower CDF ratio than TMPA, suggesting better estimation of the gauge probability distributions than its predecessor. However, the IMERG CDF ratio increases at high quantiles for all scales, implying that IMERG overestimates the frequency of heavy rainfall, especially in dry arid areas (i.e., subregion PCS). However, it should be noted that direct comparison of quantiles is limited as the sample size of rainfall events differs between regions depending on the rainfall controls and also between gauges and SPPs depending on the SPP estimation accuracy and the rain gauge density. For instance, the high fraction of missed rainfall and false alarms in PCS by IMERG implies sensor and retrieval algorithm limitations under these conditions. Overestimation of high percentiles therefore does not necessarily imply that the frequency of heavy rainfall events is overestimated, but could potentially be a result of proportional underestimation of the bulk of the rainfall intensities (i.e., low quantiles).

Furthermore, despite improvements over TMPA in spatially defining high rainfall areas as well as estimating rainfall rates, IMERG fails to accurately capture the high precipitation rates in orographically enhanced regions (i.e., ASA), which is also the region with the lowest coverage of the network. Here, warm clouds that precipitate at temperatures higher than those expected based on ice particle distributions assumed in PMW-based precipitation estimates may result in underestimations of ground-observed rainfall by the SPPs (Dinku et al. 2010). This highlights the importance of further research into the estimation of tropical mountain precipitation by

SPPs and the need to increase the number of ground-based stations in data-scarce regions.

This study has also highlighted the importance of the gauge network used for validation of SPPs. For instance, inadequate estimation of precipitation in ASA is likely to be a combined result of limitations in precipitation estimation ability as well as the low and irregular distribution of the local gauge network used for evaluation (see Fig. 1). While random errors may be addressed by statistical methods such as sampling or random simulations, elimination of systematic error (i.e., bias) requires reliable validation data for satellite–gauge adjustment (Tan et al. 2016). The unrepresentative coverage of rain gauge networks is a well-established and somewhat fixed boundary condition of hydrometeorological studies and a principal motivation for extensive research into SPPs. However, the impacts of insufficient gauges and the resulting point-area differences continue to affect satellite–gauge evaluations.

An example of this is illustrated by the analysis of the individual pixel 1167 (Fig. 11). Here overestimation of the spatial gauge mean by the SPP may stem from the gauges preferentially sampling the low rainfall area in the southwest, leading to an underestimation of the true spatial mean of the area within the pixel. Similarly, the negative relationship between gauge P46 and the SPPs illustrates the impact of the point-area difference: while subgrid-scale variability in precipitation patterns is not captured by the SPP, it may strongly impact individual gauges. Disagreements between SPPs and gauge estimates therefore do not necessarily imply satellite retrieval errors but can similarly stem from low density or nonuniform gauge coverage. While restricting satellite–gauge comparisons to pixels with at least three gauges would eliminate vast regions worldwide and hamper assessment of SPPs specifically in poorly gauged regions, the representativeness of gauge networks needs to be considered when S-G agreement is poor.

## 6. Conclusions

This study performed a comparative ground validation of IMERG and TMPA against a network of 302

rain gauges in Ecuador and Peru over a 17-month period from April 2014 to August 2015. The region is influenced by a range of climatic drivers, resulting in substantial differences in spatiotemporal precipitation patterns from the dry Peruvian coast across the tropical Andes to the Amazonian lowlands. To comprehensively evaluate the SPPs, the study area was divided into six hydrometeorological subregions and rainfall occurrence frequency, detection, quantitative estimation errors, and accuracy of empirical cumulative distribution probabilities were assessed. Despite similar precipitation means over the assessment period, IMERG outperformed TMPA in most validation statistics, demonstrating lower errors in detection and quantitative rainfall rate estimation as well as a higher accuracy in estimating occurrence frequency and rainfall intensity distributions. For both products, performance increased with increasing spatiotemporal scale because of the reduction of the space–time variability of rainfall patterns.

Advances to sensor technology and retrieval algorithms of DPR and GMI have resulted in improved detection of low rainfall intensities ( $<0.7 \text{ mm h}^{-1}$ ) and higher accuracy in estimating medium and high rainfall intensities, with consequential improvements in terms of defining the statistical probability distribution of rainfall. Improvements of IMERG over TMPA are geographically most pronounced in the high Andes (AW and AE), which confirms the promising results from previous studies with respect to the potential of IMERG in high-altitude regions. Substantial errors in terms of overestimating the frequency of rainfall as well as positive bias and largest random errors persist along the dry Peruvian coastline (PCS), where IMERG does not show improvements over TMPA. Rainfall patterns characterized by infrequent rainfall events and very low mean annual precipitation totals in this subregion result in a high fraction of missed rainfall and falsely estimated rainfall, especially for high rainfall intensities. On the other hand, IMERG shows improvement in spatially defining and quantifying orographically enhanced precipitation hot spots in the Amazonian sub-Andes, although absolute gauge rainfall totals are underestimated.

The study has also highlighted the importance of utilizing a well-developed gauge network with spatiotemporally representative coverage for evaluating SPPs. While the premise of using SPPs lies in their ability to provide rainfall estimates in poorly gauged regions, assessing SPPs in these regions is complicated by the very fact that the gauge network does not allow for comparison with equal spatial support, potentially resulting in substantial point-area differences between gauge and satellite estimates. Further research should therefore focus on developing metrics to evaluate the

representativeness of gauge networks as well as understanding limitations and associated uncertainty in their estimation accuracy of *true* rainfall, for example, employing geostatistics to quantify gauge interpolation and S-G merging uncertainties (e.g., Delrieu et al. 2014) as well as S-G error frameworks that account for the nonlinear structure of rainfall and how this manifests itself in rainfall detection and rate estimation errors (Tan et al. 2016; Maggioni et al. 2014).

Further evaluation of IMERG in the tropical Andes should investigate as to what empirical conditional probability distribution is associated with each rainfall detection category (hits, misses, false alarms) and how this varies seasonally and regionally. Investigations of single-sensor (level 2) products against ground observations may help identify the performance of individual sensor retrievals (DPR, GMI, IR retrievals), and thereby support in attributing errors in the combined IMERG product (Tan et al. 2016). While the results presented here suggest IMERG has already achieved robust improvements in estimating precipitation in the tropical Andes, these should be seen in the context of the short assessment period (17 months) that overlapped with the onset of an El Niño event in mid-2015. Interannual rainfall variability in Ecuador and Peru is strongly affected by ENSO variations. Therefore, evaluation analysis over longer time periods (e.g., TRMM era retrospective since 1998) will help ascertain the influence of interannual precipitation variability and the impacts of changes in the IMERG and TMPA calibration algorithms. For the Ecuador–Peru region, comparative analysis of IMERG versions 3 and 4 as well as distributed hydrological model simulations based on IMERG products are currently being developed.

*Acknowledgments.* This research was funded by the British Council Peru Project “Towards a high-resolution gridded precipitation product for the tropical Andes.” W.B. also received funding from U.K. NERC (Grant NE-K010239-1, Mountain-EVO project). Authors S.P.B. and N.H. gratefully acknowledge the financial support provided by the Escuela Politécnica Nacional, for the development of the project PIJ-15-14: “Evaluación de la aplicación potencial de la precipitación satelital obtenida de la misión - Global Precipitation Measurement (GPM) - a la gestión integrada de los recursos hídricos en el Ecuador.” B.O.T.’s contribution was funded by NERC SSCP DTP (Grant NE/L002515/1). B.W.’s contribution was supported by USAID through the PEER Project “Strengthening resilience of Andean river basin headwaters facing global change” (PGA 084063). The authors thank INAMHI, EPMAPS, SENAMHI, CONDESAN, and all the iMHEA partners for access to gauge records. The IMERG and TMPA

datasets were provided by the NASA Goddard Space Flight Center's PMM and PPS, which develop and compute IMERG and TMPA as a contribution to GPM and TRMM, respectively.

## REFERENCES

- Boers, N., B. Bookhagen, N. Marwan, J. Kurths, and J. Marengo, 2013: Complex networks identify spatial patterns of extreme rainfall events of the South American monsoon system. *Geophys. Res. Lett.*, **40**, 4386–4392, doi:10.1002/grl.50681.
- , R. V. Donner, B. Bookhagen, and J. Kurths, 2015: Complex network analysis helps to identify impacts of the El Niño Southern Oscillation on moisture divergence in South America. *Climate Dyn.*, **45**, 619–632, doi:10.1007/s00382-014-2265-7.
- Bookhagen, B., and M. R. Strecker, 2008: Orographic barriers, high-resolution TRMM rainfall, and relief variations along the eastern Andes. *Geophys. Res. Lett.*, **35**, L06403, doi:10.1029/2007GL032011.
- Buytaert, W., R. Celleri, P. Willems, B. De Bievre, and G. Wyseure, 2006: Spatial and temporal rainfall variability in mountainous areas: A case study from the south Ecuadorian Andes. *J. Hydrol.*, **329**, 413–421, doi:10.1016/j.jhydrol.2006.02.031.
- Chen, F., and X. Li, 2016: Evaluation of IMERG and TRMM 3B43 monthly precipitation products over mainland China. *Remote Sens.*, **8**, 472, doi:10.3390/rs8060472.
- Ciach, G. J., and W. F. Krajewski, 1999: On the estimation of radar rainfall error variance. *Adv. Water Resour.*, **22**, 585–595, doi:10.1016/S0309-1708(98)00043-8.
- Delrieu, G., A. Wijbrans, B. Boudevillain, D. Faure, L. Bonnifait, and P.-E. Kirstetter, 2014: Geostatistical radar–raingauge merging: A novel method for the quantification of rain estimation accuracy. *Adv. Water Resour.*, **71**, 110–124, doi:10.1016/j.advwatres.2014.06.005.
- Demaria, E. M. C., D. A. Rodriguez, E. E. Ebert, P. Salio, F. Su, and J. B. Valdes, 2011: Evaluation of mesoscale convective systems in South America using multiple satellite products and an object-based approach. *J. Geophys. Res.*, **116**, D08103, doi:10.1029/2010JD015157.
- Derin, Y., and Coauthors, 2016: Multiregional satellite precipitation products evaluation over complex terrain. *J. Hydrometeorol.*, **17**, 1817–1836, doi:10.1175/JHM-D-15-0197.1.
- Dinku, T., F. Ruiz, S. J. Connor, and P. Ceccato, 2010: Validation and intercomparison of satellite rainfall estimates over Colombia. *J. Appl. Meteor. Climatol.*, **49**, 1004–1014, doi:10.1175/2009JAMC2260.1.
- Duan, Y., A. M. Wilson, and A. P. Barros, 2015: Scoping a field experiment: Error diagnostics of TRMM precipitation radar estimates in complex terrain as a basis for IPHEX2014. *Hydrol. Earth Syst. Sci.*, **19**, 1501–1520, doi:10.5194/hess-19-1501-2015.
- Espinoza, J. C., S. Chavez, J. Ronchail, C. Junquas, K. Takahashi, and W. Lavado, 2015: Rainfall hotspots over the southern tropical Andes: Spatial distribution, rainfall intensity, and relations with large-scale atmospheric circulation. *Water Resour. Res.*, **51**, 3459–3475, doi:10.1002/2014WR016273.
- Espinoza Villar, J. C., and Coauthors, 2009: Spatio-temporal rainfall variability in the Amazon basin countries (Brazil, Peru, Bolivia, Colombia, and Ecuador). *Int. J. Climatol.*, **29**, 1574–1594, doi:10.1002/joc.1791.
- Falck, A., V. Maggioni, J. Tomasella, D. Vila, and F. Diniz, 2015: Propagation of satellite precipitation uncertainties through a distributed hydrologic model: A case study in the Tocantins Araguaia basin in Brazil. *J. Hydrol.*, **527**, 943–957, doi:10.1016/j.jhydrol.2015.05.042.
- 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.
- 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.
- Hsu, K.-L., H. V. Gupta, X. Gao, and S. Sorooshian, 1999: Estimation of physical variables from multichannel remotely sensed imagery using a neural network: Application to rainfall estimation. *Water Resour. Res.*, **35**, 1605–1618, doi:10.1029/1999WR900032.
- 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.
- , R. F. Adler, D. T. Bolvin, and E. J. Nelkin, 2010: The TRMM Multi-Satellite Precipitation Analysis (TMPA). *Satellite Rainfall Applications for Surface Hydrology*, M. Gebremichael and F. Hossain, Eds., Springer, 3–22.
- , D. T. Bolvin, and E. J. Nelkin, 2015a: Integrated Multi-satellite Retrievals for GPM (IMERG) technical documentation. NASA/GSFC Code 612 Tech. Doc., 48 pp. [Available online at [http://pmm.nasa.gov/sites/default/files/document\\_files/IMERG\\_doc.pdf](http://pmm.nasa.gov/sites/default/files/document_files/IMERG_doc.pdf).]
- , —, D. Braithwaite, K. Hsu, R. Joyce, C. Kidd, E. J. Nelkin, and P. Xie, 2015b: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). IMERG ATBD, version 4.5, 30 pp. [Available online at [https://pmm.nasa.gov/sites/default/files/document\\_files/IMERG\\_ATBD\\_V4.5.pdf](https://pmm.nasa.gov/sites/default/files/document_files/IMERG_ATBD_V4.5.pdf).]
- Jiang, S., L. Ren, Y. Hong, B. Yong, X. Yang, F. Yuan, and M. Ma, 2012: Comprehensive evaluation of multi-satellite precipitation products with a dense rain gauge network and optimally merging their simulated hydrological flows using the Bayesian model averaging method. *J. Hydrol.*, **452–453**, 213–225, doi:10.1016/j.jhydrol.2012.05.055.
- Joyce, R. J., 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.
- 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.
- Lavado-Casimiro, W. S., J. Ronchail, D. Labat, J. C. Espinoza, and J. L. Guyot, 2012: Basin-scale analysis of rainfall and runoff in Peru (1969–2004): Pacific, Titicaca and Amazonas drainages. *Hydrol. Sci. J.*, **57**, 625–642, doi:10.1080/02626667.2012.672985.
- Li, L., and Coauthors, 2009: Evaluation of the real-time TRMM-based multi-satellite precipitation analysis for an operational flood prediction system in Nzoia basin, Lake Victoria, Africa. *Nat. Hazards*, **50**, 109–123, doi:10.1007/s11069-008-9324-5.
- Li, N., G. Tang, P. Zhao, Y. Hong, Y. Gou, and K. Yang, 2016: Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang River basin. *Atmos. Res.*, **183**, 212–223, doi:10.1016/j.atmosres.2016.07.020.
- Ma, Y., G. Tang, D. Long, B. Yong, L. Zhong, W. Wan, and Y. Hong, 2016: Similarity and error intercomparison of the GPM and its predecessor-TRMM multisatellite precipitation analysis using

- the best available hourly gauge network over the Tibetan Plateau. *Remote Sens.*, **8**, 569, doi:10.3390/rs8070569.
- Maggioni, V., M. R. 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.
- Mantas, V. M., Z. Liu, C. Caro, and A. Pereira, 2015: Validation of TRMM multi-satellite precipitation analysis (TMPA) products in the Peruvian Andes. *Atmos. Res.*, **163**, 132–145, doi:10.1016/j.atmosres.2014.11.012.
- Manz, B., W. Buytaert, Z. Zulkafli, W. Lavado, B. Willems, L. A. Robles, and J.-P. Rodríguez-Sánchez, 2016: High-resolution satellite-gauge merged precipitation climatologies of the Tropical Andes. *J. Geophys. Res. Atmos.*, **121**, 1190–1207, doi:10.1002/2015JD023788.
- Mohr, K. I., D. Slayback, and K. Yager, 2014: Characteristics of precipitation features and annual rainfall during the TRMM era in the central Andes. *J. Climate*, **27**, 3982–4001, doi:10.1175/JCLI-D-13-00592.1.
- Nesbitt, S. W., and A. M. Anders, 2009: Very high resolution precipitation climatologies from the Tropical Rainfall Measuring Mission precipitation radar. *Geophys. Res. Lett.*, **36**, L15815, doi:10.1029/2009GL038026.
- Nikolopoulos, E. I., E. N. Anagnostou, and M. Borga, 2013: Using high-resolution satellite rainfall products to simulate a major flash flood event in northern Italy. *J. Hydrometeorol.*, **14**, 171–185, doi:10.1175/JHM-D-12-09.1.
- Ochoa, A., L. Pineda, P. Crespo, and P. Willems, 2014: Evaluation of TRMM 3B42 precipitation estimates and WRF retrospective precipitation simulation over the Pacific–Andean region of Ecuador and Peru. *Hydrol. Earth Syst. Sci.*, **18**, 3179–3193, doi:10.5194/hess-18-3179-2014.
- Ochoa-Tocachi, B. F., and Coauthors, 2016: Impacts of land use on the hydrological response of tropical Andean catchments. *Hydrol. Processes*, **30**, 4074–4089, doi:10.1002/hyp.10980.
- Padrón, R. S., B. P. Wilcox, P. Crespo, and R. Céleri, 2015: Rainfall in the Andean Páramo: New insights from high-resolution monitoring in southern Ecuador. *J. Hydrometeorol.*, **16**, 985–996, doi:10.1175/JHM-D-14-0135.1.
- Prakash, S., A. K. Mitra, D. S. Pai, and A. AghaKouchak, 2016: From TRMM to GPM: How well can heavy rainfall be detected from space? *Adv. Water Resour.*, **88**, 1–7, doi:10.1016/j.advwatres.2015.11.008.
- , —, A. AghaKouchak, Z. Liu, H. Norouzi, and D. S. Pai, 2017: A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J. Hydrol.*, doi:10.1016/j.jhydrol.2016.01.029, in press.
- Rasmussen, K. L., S. L. Choi, M. D. Zuluaga, and R. A. Houze, 2013: TRMM precipitation bias in extreme storms in South America. *Geophys. Res. Lett.*, **40**, 3457–3461, doi:10.1002/grl.50651.
- Rollenbeck, R., and J. Bendix, 2011: Rainfall distribution in the Andes of southern Ecuador derived from blending weather radar data and meteorological field observations. *Atmos. Res.*, **99**, 277–289, doi:10.1016/j.atmosres.2010.10.018.
- Romatschke, U., and R. A. Houze Jr., 2010: Extreme summer convection in South America. *J. Climate*, **23**, 3761–3791, doi:10.1175/2010JCLI3465.1.
- Sahlu, D., E. I. Nikolopoulos, S. A. Moges, E. N. Anagnostou, and D. Hailu, 2016: First evaluation of the Day-1 IMERG over the upper Blue Nile basin. *J. Hydrometeorol.*, **17**, 2875–2882, doi:10.1175/JHM-D-15-0230.1.
- Satgé, F., M.-P. Bonnet, M. Gosset, J. Molina, W. H. Y. Lima, R. P. Zol, F. Timouk, and J. Garnier, 2016: Assessment of satellite rainfall products over the Andean Plateau. *Atmos. Res.*, **167**, 1–14, doi:10.1016/j.atmosres.2015.07.012.
- Sharifi, E., R. Steinacker, and B. Saghafian, 2016: Assessment of GPM-IMERG and other precipitation products against gauge data under different topographic and climatic conditions in Iran: Preliminary results. *Remote Sens.*, **8**, 135, doi:10.3390/rs8020135.
- Shen, Y., A. Xiong, Y. Wang, and P. Xie, 2010: Performance of high-resolution satellite precipitation products over China. *J. Geophys. Res.*, **115**, D02114, doi:10.1029/2009JD012097.
- Tan, J., W. A. Petersen, and A. Tokay, 2016: A novel approach to identify sources of errors in IMERG for GPM ground validation. *J. Hydrometeorol.*, **17**, 2477–2491, doi:10.1175/JHM-D-16-0079.1.
- Tang, G., Z. Zeng, D. Long, X. Guo, B. Yong, W. Zhang, and Y. Hong, 2016: 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.
- 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.
- Tarnavsky, E., M. Mulligan, and G. Husak, 2012: Spatial disaggregation and intensity correction of TRMM-based rainfall time series for hydrological applications in dryland catchments. *Hydrol. Sci. J.*, **57**, 248–264, doi:10.1080/02626667.2011.637498.
- Tian, Y., C. D. Peters-Lidard, and J. B. Eylander, 2010: Real-time bias reduction for satellite-based precipitation estimates. *J. Hydrometeorol.*, **11**, 1275–1285, doi:10.1175/2010JHM1246.1.
- Villarini, G., 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.
- Vuille, M., R. S. Bradley, and F. Keimig, 2000: Climate variability in the Andes of Ecuador and its relation to tropical Pacific and Atlantic sea surface temperature anomalies. *J. Climate*, **13**, 2520–2535, doi:10.1175/1520-0442(2000)013<2520:CVTAAO>2.0.CO;2.
- Wang, J., and D. B. Wolff, 2010: Evaluation of TRMM ground-validation radar-rain errors using rain gauge measurements. *J. Appl. Meteor. Climatol.*, **49**, 310–324, doi:10.1175/2009JAMC2264.1.
- , B. L. Fisher, and D. B. Wolff, 2008: Estimating rain rates from tipping-bucket rain gauge measurements. *J. Atmos. Oceanic Technol.*, **25**, 43–56, doi:10.1175/2007JTECHA895.1.
- Ward, E., W. Buytaert, L. Peaver, and H. Wheeler, 2011: Evaluation of precipitation products over complex mountainous terrain: A water resources perspective. *Adv. Water Resour.*, **34**, 1222–1231, doi:10.1016/j.advwatres.2011.05.007.
- Wilson, P. S., and R. Toumi, 2005: A fundamental probability distribution for heavy rainfall. *Geophys. Res. Lett.*, **32**, L14812, doi:10.1029/2005GL022465.
- WMO, 2014: Guide to meteorological instruments and methods of observation. 8th ed. WMO 8, 1139 pp. [Available online at <http://www.wmo.int/pages/prog/www/IMOP/CIMO-Guide.html>.]
- Wohl, E., and Coauthors, 2012: The hydrology of the humid tropics. *Nat. Climate Change*, **2**, 655–662, doi:10.1038/nclimate1556.
- Zulkafli, Z., W. Buytaert, C. Onof, B. Manz, E. Tarnavsky, W. Lavado, and J.-L. Guyot, 2014: A comparative performance analysis of TRMM 3B42 (TMPA) versions 6 and 7 for hydrological applications over Andean–Amazon River basins. *J. Hydrometeorol.*, **15**, 581–592, doi:10.1175/JHM-D-13-094.1.