A comprehensive five-year evaluation of IMERG late run precipitation estimates over the Netherlands

Bogerd, Linda; Overeem, Aart; Leijnse, Hidde; Uijlenhoet, Remko

SPECIAL

Global Precipitation Measurement (GPM)

COLLECTION

A Comprehensive Five-Year Evaluation of IMERG Late Run Precipitation Estimates over the Netherlands

LINDA BOGERD,[a,b] AART OVEREEM,[b,a] HIDDE LEIJNSE,[b,a] AND REMKO UIJLENHOET[c,a]

[a] *Hydrology and Quantitative Water Management Group, Wageningen University and Research, Wageningen, Netherlands*
[b] *R&D Observations and Data Technology, Royal Netherlands Meteorological Institute (KNMI), De Bilt, Netherlands*
[c] *Department of Water Management, Delft University of Technology, Delft, Netherlands*

ABSTRACT: Applications like drought monitoring and forecasting can profit from the global and near-real-time availability of satellite-based precipitation estimates once their related uncertainties and challenges are identified and treated. To this end, this study evaluates the IMERG V06B Late Run precipitation product from the Global Precipitation Measurement mission (GPM), a multisatellite product that combines space-based radar, passive microwave (PMW), and infrared (IR) data into gridded precipitation estimates. The evaluation is performed on the spatiotemporal resolution of IMERG ($0.1° \times 0.1°$, 30 min) over the Netherlands over a 5-yr period. A gauge-adjusted radar precipitation product from the Royal Netherlands Meteorological Institute (KNMI) is used as reference, against which IMERG shows a large positive bias. To find the origin of this systematic overestimation, the data are divided into seasons, rainfall intensity ranges, echo top height (ETH) ranges, and categories based on the relative contributions of IR, morphing, and PMW data to the IMERG estimates. Furthermore, the specific radiometer is identified for each PMW-based estimate. IMERG's detection performance improves with higher ETH and rainfall intensity, but the associated error and relative bias increase as well. Severe overestimation occurs during low-intensity rainfall events and is especially linked to PMW observations. All individual PMW instruments show the same pattern: overestimation of low-intensity events and underestimation of high-intensity events. IMERG misses a large fraction of shallow rainfall events, which is amplified when IR data are included. Space-based retrieval of shallow and low-intensity precipitation events should improve before IMERG can become accurate over the middle and high latitudes.

KEYWORDS: Rainfall; Radars/Radar observations; Satellite observations; Precipitation; Remote Sensing

1. Introduction

Precipitation observations are required for environmental applications that are highly embedded in the contemporary society, such as crop yield and flash flood forecasting, water management, and drought monitoring. However, the global coverage of ground-based precipitation measurements is limited, especially over Africa, South America, parts of Asia, and regions that are difficult to access (e.g., oceans, mountainous areas, polar regions; Lorenz and Kunstmann 2012; Saltikoff et al. 2019). Besides being limited in their spatial representation, ground-based measurements retrieved from rain gauges are restricted in their temporal resolution as well. Satellites can potentially overcome these limitations, as they are able to provide global, uniformly distributed, and quasi-real-time precipitation observations.

Satellite-based estimates, however, have challenges of their own. Examples are the retrieval of shallow precipitation (Petković and Kummerow 2016; Arulraj and Barros 2017) and snowfall (Foster et al. 2012; Casella et al. 2017).

Both precipitation types frequently occur over the middle and high latitudes, including areas that are not covered by ground-based measurements such as oceans and regions above the polar circles. Those deficiencies can cause unreliable precipitation estimates, while accurate estimates are vital for environmental applications. Hence, to be able to improve mid- and high-latitude precipitation retrievals, it is of fundamental importance to identify the sources of error and quantify the uncertainties related to precipitation estimations retrieved from satellites.

The Global Precipitation Measurement mission (GPM) (Hou et al. 2014; Skofronick-Jackson et al. 2018) is one of the recent efforts to improve space-based precipitation estimates. The GPM *Core Observatory* satellite has an orbit extending from 65°S to 65°N and carries the Dual-Frequency Precipitation Radar (DPR) and the GPM Microwave Imager (GMI) on board. Besides the core satellite, which provides high-quality estimates approximately once a day for a given location (Hou et al. 2014), GPM consists of a constellation of partner satellites carrying radiometers. Additionally, GPM incorporates observations from geostationary satellites with infrared (IR) sensors to collect as much data as possible. Due to the large spatial coverage and the inclusion of numerous

satellites, GPM provides the opportunity to elaborately study space-based precipitation measurements over different climates and from various instruments.

Apart from offering observations as retrieved from individual instruments, the GPM products also include the gridded Integrated Multisatellite Retrievals for GPM (IMERG) precipitation estimates. IMERG combines all GPM data available (including reanalysis data that are used for the morphing component since version V06B) to create a half-hourly precipitation product with a $0.1° \times 0.1°$ spatial resolution (Huffman et al. 2020). Three different IMERG products exist: two near-real-time (NRT) runs (Early, IMERG-E and Late, IMERG-L) and one post-real-time run (Final, IMERG-F). IMERG-F has been extensively evaluated over different geographical areas and at different time scales (e.g., Rios Gaona et al. 2016; Chen and Li 2016; Asong et al. 2017; Tan et al. 2017; Dezfuli et al. 2017; Ramsauer et al. 2018; Prakash et al. 2018; Cui et al. 2019; Maranan et al. 2020; Freitas et al. 2020), as IMERG-F is claimed to provide the most accurate space-based precipitation estimates on such high spatial and temporal resolution currently available (Huffman et al. 2019).

The higher accuracy of IMERG-F compared to the NRT runs is attributed to the implementation of Monthly Global Precipitation Climatology Centre (GPCC) rain gauges (Foelsche et al. 2017; Tapiador et al. 2019; Hosseini-Moghari and Tang 2020; Mohammed et al. 2020). However, the inclusion of rain gauges is not necessarily beneficial for those areas where satellite observations could add most value: regions where ground observations are scarce or even absent. Furthermore, it takes several months after the observation before IMERG-F is available. This latency is reduced to 4 and 14 h for IMERG-E and IMERG-L, respectively. Hence, NRT runs are more feasible to use for operational meteorology or water management despite their lower accuracy compared to IMERG-F.

The aforementioned studies also show that IMERG-L outperforms IMERG-E. This better performance is attributed to 1) the use of both forward and backward propagation in the IMERG-L run while IMERG-E only extrapolates forward in time and 2) the inclusion of additional data that are not delivered in time to be included in IMERG-E but are in time to be included in IMERG-L due to its additional 10 h of latency. Because of the combination of higher accuracy compared to IMERG-E and the exclusion of ground-based observations and the associated shorter latency compared to IMERG-F, IMERG-L is selected to evaluate and unravel the discrepancies between the reference and a satellite-only precipitation product.

Thus far, in-depth evaluations focusing on IMERG-L are scarce, especially over the midlatitudes (Gebregiorgis et al. 2018; Wang and Yong 2020). The seasonal performance of IMERG-L is only briefly considered in the aforementioned studies, despite the different types of precipitation that characterize the different seasons (e.g., Attema and Lenderink 2014; Navarro et al. 2019; Tapiador et al. 2019). Additionally, it is expected that the performance of IMERG-L is related to rainfall intensity, as already proved to be the case for IMERG-F (Xu et al. 2017). This relation is anticipated to be even stronger for IMERG-L: the implementation of gauges in

IMERG-F may mask errors associated with satellite-based observations, such as attenuation of the signal during high-intensity rainfall events and difficulties to detect low-intensity rainfall events. Each instrument that contributes to IMERG has its own deficiencies concerning these precipitation characteristics. Hence, identification of the source available during the moment of observation provides valuable insights into the required adaptations to improve the overall performance of IMERG.

This study aims to validate IMERG-L precipitation estimates over the period January 2015–December 2019 over the Netherlands. A high-resolution gauge-adjusted weather radar dataset, which is available over the entire period and research area, is used as reference. As far as the authors are aware, this is the first study validating IMERG over such an extended time period over a midlatitude country, performing an in-depth analysis of IMERG-L V06B and its constituents. Furthermore, the influence of 1) seasons, 2) rainfall intensity, 3) vertical extent of precipitation, and 4) the relative contribution of the IR and PMW sources on IMERG's performance are evaluated.

2. Measurement and methods

a. Data

The three datasets used in this study were all available over the research area, the Netherlands (50.78°–53.68°N, 3.38°–7.38°E; 35 000 km^2) and for the evaluation period from 1 January 2015 to 31 December 2019. Each dataset is briefly described in the following subsections.

1) SATELLITE RAINFALL ESTIMATES: IMERG V06B

This study evaluates the most recent version (V06B) of IMERG, the gridded multisatellite precipitation product of GPM. To obtain the high spatiotemporal resolution of IMERG ($0.1° \times 0.1°$, 30 min), precipitation estimates from various GPM partner satellites with passive microwave (PMW) sensors on board are combined. Additionally, a morphing algorithm is applied to fill time gaps between PMW observations using motion vectors. If the time gap between two subsequent PMW observations is larger than ~30 min, infrared (IR) observations are additionally included to update the final precipitation estimates (Huffman et al. 2019, 2020).

The key difference between IMERG V06B and its previous versions is a modification in the morphing algorithm. While motion vectors were derived from cloud top observations retrieved from IR measurements in previous versions, they are derived from reanalysis data (MERRA-2 for IMERG-F, GEOS FP for IMERG-E and IMERG-L) in V06B. Additionally, IMERG-L V06B does not involve climatological calibrations based on Global Precipitation Climatology (GPCC) gauges, which makes the calibrated and uncalibrated precipitation observations identical and therefore independent of direct ground observations. More details about (the recent changes in) IMERG are described in Tan et al. (2019) and Huffman et al. (2020). For the remaining of this paper, the IMERG-L V06B product is referred to as IMERG.

To study whether the source (i.e., PMW, morphing and/or IR) of observation affects the accuracy of precipitation estimates,

four categories were distinguished [using a similar approach as applied in Tan et al. (2016) and Maranan et al. (2020)]. The observations were categorized based on the availability of PMW observations (stored in the IMERG data field HQPrecipSource) and the weighted percentage of IR observations included to retrieve the final precipitation estimate (stored in the IMERG data field IRKalmanFilterWeight). The following four categories were defined: 1) PMW observations (HQPrecipSource $\neq$ 0, referred to as PMW), 2) spatially advected PMW observations (HQPrecipSource = 0 and 0% IRKalmanFilterWeight, referred to as PMW-morph), and two categories for morphed estimates distinguished on the weighted contribution of IR data, namely, 3) mostly morphed (<50% IRKalmanFilterWeight, referred to as morph+IR < 50%), or 4) mostly IR (>50% IRKalmanFilterWeight, referred to as morph+IR > 50%). Because the frequency of estimates solely based on IR observations was found to be small (they mostly occur over areas where PMW estimates are unreliable, such as snow surfaces), no separate category for "only IR observations" was created.

Five radiometers contributed to IMERG estimates over the Netherlands during the studied period: two sounders [Microwave Humidity Sounder (MHS) and Advanced Technology Microwave Sounder (ATMS)] and three imagers [GPM Microwave Imager (GMI), Advanced Microwave Scanning Radiometer (AMSR), Special Sensor Microwave Imager/Sounder (SSMIS)]. Note that if a particular type of sensor is mounted on multiple satellites, they are grouped together. The performance of PMW estimates might be sensor dependent, as exact specifications vary among the radiometers. Hence, five additional categories representing the specific instruments were added to the source categories.

2) GROUND-BASED RAINFALL ESTIMATES: GAUGE-ADJUSTED RADAR

A gauge-adjusted radar dataset obtained from the Royal Netherlands Meteorological Institute (KNMI) was used as reference to validate the IMERG precipitation estimates. This gridded dataset completely covers the land surface of the Netherlands at a spatial resolution of ~1 km^2 and a 5-min temporal resolution. Radar data were unavailable for some time steps within the studied time period due to, for instance, maintenance of the radar systems. Time gaps of 5 min were filled by means of linear interpolation. If more than one 5-min radar sample were missing in half an hour, this half-hour was removed before further analysis for all datasets. Still, the temporal coverage remains larger than 99% for the considered period.

The rainfall estimates are based on composites of two C-band radars, which measure instantaneous rainfall every 5 min. Until 2017, four elevation scans were used (0.3°, 1.1°, 2.0°, and 3.0°). This is reduced to three elevation scans (0.3°, 0.8°, and 2.0°) from 2017 onward. These scans are used to construct the pseudoCAPPI (pseudo–constant altitude plan position indicator) at a height of 1500 m. Subsequently, the CAPPIs from both radars are combined using a range-dependent weighting factor. The weights decrease with distance from the radar, except close to the radar, where the weights become smaller to mitigate the effects of residual clutter and the cone of silence. As the majority of the land surface of the Netherlands is covered by at least one scan with a height that is below or at the CAPPI level, the impacts of (large) overestimations of rainfall intensity that occur due to bright band effects is limited (Overeem et al. 2009b, 2020). Subsequently, this combined composite is adjusted with gauge data from KNMI (31 automatic and 325 manual gauges). These gauge adjustments provide corrections during weather circumstances that reduce the accuracy of radar estimates, such as overshooting or variability of the drop size distribution. The combination of CAPPIs, distance weighting, and gauge adjustments yields a high-quality dataset. More detailed information about this dataset (including an assessment of its quality) can be found in Overeem et al. (2009a,b, 2011).

3) GROUND-BASED ECHO TOP HEIGHT OBSERVATIONS: RADAR

As briefly mentioned, one of the persistent challenges for satellite-based precipitation monitoring is the detection of shallow precipitation events (i.e., precipitation from clouds in the lower parts of the atmosphere). Therefore, radar echo top height (ETH) data were used to examine the influence of the vertical extent of a precipitating area on IMERG's performance. The ETH observations are retrieved from the same C-band radars as described in the previous subsection. However, while precipitation estimates are based on three or four vertical elevation scans, fifteen elevations (ranging from 0.3° to 12.0°) are used for the ETH product.

ETH is defined as the maximum height at which a reflectivity threshold of 7 dBZ is exceeded. This low detection threshold combined with residual clutter, as well as the vertical sampling by the radar (especially at long ranges overshooting may occur as a consequence of the increasing height of observation), can induce unrealistically high or low ETH values. Therefore, extremely low (below 1 km) and high (above 15 km) ETH observations were removed before further analysis [more information about the ETH product and an evaluation can be found in Beekhuis and Holleman (2008) and Aberson (2011)].

b. Spatiotemporal aggregation

The spatiotemporal resolution of the KNMI (gauge-adjusted) radar datasets was aggregated to match the coarser IMERG resolution. First, the 5-min KNMI estimates were aggregated into 30-min estimates for each radar pixel by summing the precipitation estimates and averaging the ETH observations. Then, each radar pixel was allocated to an IMERG pixel based on the minimum distance between the grid centers. The arithmetic mean of all radar pixels belonging to a certain IMERG pixel was assumed to be representative for spatial aggregation to IMERG resolution. Additionally, the IMERG dataset was converted from intensity (mm h^{-1}) to rainfall depth (mm). Once all datasets had the same spatiotemporal resolution, a mask was applied to make sure only pixels (partly) over the land surface of the Netherlands were considered for further analysis. Due to the limited availability of rain gauges over the sea region, this area was excluded from further analysis to ensure the reference product had a consistent spatial performance.

c. Validation

Five widely applied statistics and contingency metrics were adopted to evaluate the performance of IMERG. The overall error magnitude was calculated with the mean absolute error (MAE). To be able to compare the MAE across different seasons and observation sources, its normalized version (NMAE) was also computed. The relative bias (RB) was used to indicate the direction (i.e., underestimation or overestimation) of the systematic bias. The RB, MAE, and NMAE are defined as follows:

$$RB = \frac{\sum_{i=1}^{n}(R_{\mathrm{IMERG},i} - R_{\mathrm{radar},i})}{\sum_{i=1}^{n} R_{\mathrm{radar},i}}, \tag{1}$$

$$MAE = \frac{\sum_{i=1}^{n}|R_{\mathrm{IMERG},i} - R_{\mathrm{radar},i}|}{n}, \tag{2}$$

$$NMAE = \frac{\sum_{i=1}^{n}|R_{\mathrm{IMERG},i} - R_{\mathrm{radar},i}|}{\sum_{i=1}^{n} R_{\mathrm{radar},i}}, \tag{3}$$

where n represents the number of pixels at all time steps available for the studied period or corresponding to a certain selection (e.g., season, intensity). All three metrics would yield a value of 0 if the observations of IMERG would be identical to the radar observations.

Furthermore, two contingency metrics were adopted to evaluate the capability of IMERG to distinguish between wet and dry pixels. These are the probability of detection (POD) and probability of false alarm (POFA, often referred to as the false alarm ratio). The POD and POFA are defined as

$$POD = \frac{\text{hits}}{\text{hits} + \text{misses}}, \tag{4}$$

$$POFA = \frac{\text{false alarms}}{\text{hits} + \text{false alarms}}, \tag{5}$$

where "hit" means that both IMERG and the reference (the radar estimates) identify a pixel as "rainy," "false alarm" means that IMERG identifies a pixel as rainy while the reference identifies the pixel as "dry," and "miss" means that the reference identifies the pixel as rainy while IMERG identifies the pixel as dry. The threshold to distinguish between dry and wet pixels was 0.1-mm rainfall depth in 30 min. If IMERG would always correctly identify rainy pixels, POFA would be 0 and POD would be 1.

Except for the contingency metrics, the values in Tables 1–4 in the results section are based on observations where both radar and IMERG exceeded the threshold of 0.1-mm rainfall depth. With all observations included where only the radar exceeded the threshold, the RB, MAE, and NMAE scores were lower due to the inclusion of misses (i.e., the scores improved as the values smaller than 0.1 mm reduce the (positive) bias). Because this compensation effect was similar for all statistics and selections and the number of misses can be deduced from the POD, only the scores based on the subsets of hits are shown in the tables.

TABLE 1. IMERG performance on a pixel-by-pixel basis (0.1° × 0.1° spatial resolution, 30-min temporal resolution) for the entire studied period (January 2015–December 2019). The calculations of μ_{radar} (mean of the radar estimates), μ_{IMERG} (mean of the IMERG estimates), RB, MAE, and NMAE are based on paired observations where both radar and IMERG exceed the threshold value of 0.1 mm per 30 min (i.e., only hits were considered).

Season	μ_{radar} (mm)	μ_{IMERG} (mm)	RB (%)	MAE (mm)	NMAE (—)	POD	POFA
All	0.77	1.28	66	0.94	1.22	0.51	0.46
DJF	0.68	1.39	104	1.01	1.48	0.47	0.45
MAM	0.70	1.38	96	1.07	1.43	0.52	0.47
JJA	0.95	1.38	46	1.05	1.11	0.59	0.47
SON	0.76	1.00	32	0.73	0.97	0.50	0.44

Last, we compared the cumulative distribution functions (CDF) of IMERG and radar estimates. The CDF was calculated on both occurrence (i.e., the relative contribution of a certain rainfall depth to the total precipitation occurrence) and volume (i.e., the relative contribution of a certain rainfall depth to the total volume).

3. Results

From the spatially averaged monthly sums (i.e., time series calculated from IMERG native resolution over the entire period of evaluation), it is clear that IMERG systematically overestimates the monthly amount of rainfall compared to the reference Fig. 1). This overestimation is generally lowest in the second half of the year, i.e., midsummer, fall, and the beginning of winter. The largest absolute and relative overestimation occurs in January and February, respectively.

The performance of IMERG on a pixel-by-pixel basis over the entire research period is summarized in the upper row of Table 1. The mean rainfall amount is noticeably higher according to IMERG (1.28 mm) compared to radar (0.77 mm). The corresponding RB is high and positive (66%), which means that IMERG systematically overestimates rainfall (as already observed in Fig. 1). Furthermore, IMERG misses approximately half of the radar rainfall events (POD = 0.51), while at the same time almost half of IMERG's rainfall events are false alarms (POFA = 0.46).

To further analyze the fluctuating monthly performance of IMERG and to identify what causes the systematic overestimation and large number of misses, four categories are defined and discussed in the subsections below. First, the seasonal performance of IMERG is analyzed in more detail. Seasons are defined as the four meteorological seasons in the Northern Hemisphere: winter (DJF), spring (MAM), summer (JJA), and fall (SON). Then, the effect of rainfall intensity and the vertical extent of precipitation on the accuracy of the IMERG estimates are considered, both strongly linked to the seasons in the Netherlands. Finally, the degree to which the source of observation influences the performance of IMERG is examined.

a. Seasonality

The results of the seasonal evaluation of IMERG are shown in the last four rows of Table 1. The mean radar rainfall depth
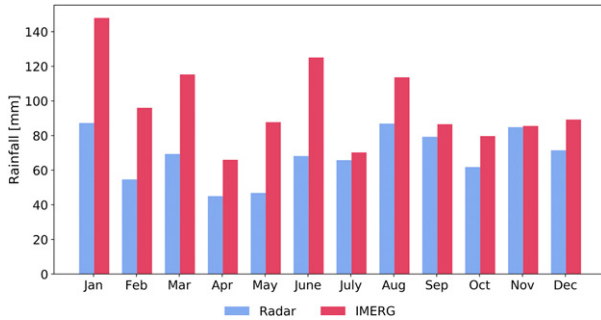
FIG. 1. Spatially averaged (time series calculated from IMERG native resolution over the entire period of evaluation) monthly rainfall accumulations according to the radar and IMERG datasets for the January 2015–December 2019 period.

shows a seasonal dependence and is smallest in winter and largest in summer. Conversely, the mean IMERG rainfall depth is similar during winter, spring, and summer. The high RB and NMAE and the low POD indicate that winter is the most challenging season for IMERG. Consistent with Fig. 1, the mean IMERG rainfall depth, RB, and NMAE values are smallest in fall. An analysis of the individual years yields similar conclusions (not shown).

A comparison between the seasonal rainfall maps of IMERG and radar reveals that, except for fall, IMERG overestimates the amount of rain over the entire surface of the Netherlands (Fig. 2). A certain geographical dependence can be observed as IMERG consistently shows higher values over the Waddeneilanden (small islands in the north of the Netherlands), Zeeland (southwest), and Limburg (southeast) compared to the reference observations, especially in winter and spring. Furthermore, the RB (last row of Fig. 2) seems reduced in the middle of the country at a larger distance from the North Sea, again especially during winter and spring. Compared to the other seasons, the bias is small during fall and even almost zero in the middle of the country. The small spatial bias during fall indicates that IMERG is able to correctly capture the higher amounts of rainfall near the coast of the Netherlands while the amount gradually decreases toward the east of the country.

b. Rainfall intensity

Dividing the radar estimates into five intensity classes clearly shows that the detection performance of IMERG improves for higher intensities (Table 2). More than half of the pixels with rainfall amounts between 0.1 and 1 mm are not detected, while the correctly identified wet pixels are highly overestimated.
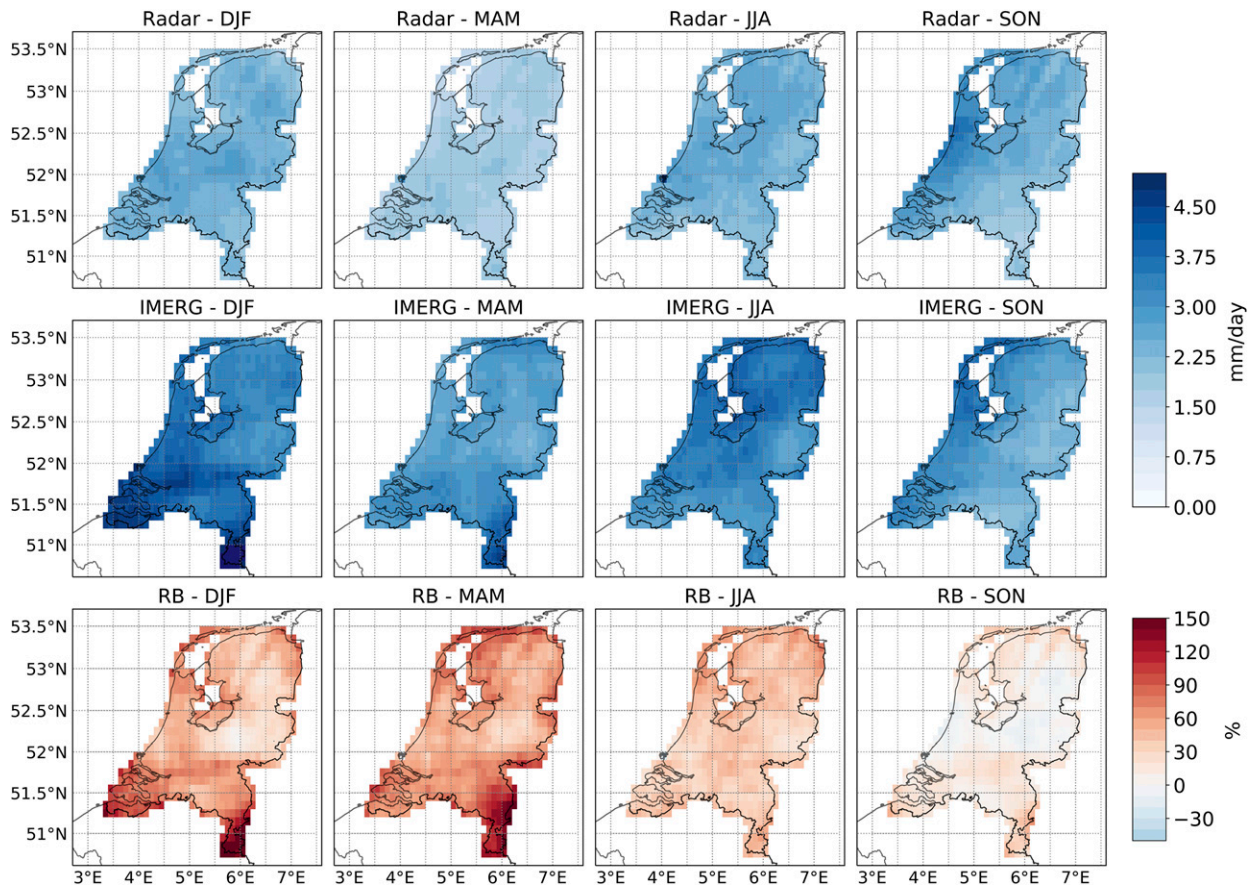

FIG. 2. Seasonally averaged precipitation maps (mm day^{-1}) according to (top) radar and (middle) IMERG, and (bottom) the relative bias (RB) for the January 2015–December 2019 period.

TABLE 2. IMERG performance on a pixel-by-pixel basis decomposed into 30-min rainfall intensity intervals. Because the intervals are applied to radar rainfall depths and start from 0.1 mm, NMAE (divided by the sum of radar observations) is omitted. POFA is calculated based on 30-min IMERG observations. Other details are the same as stated in Table 1.

Rainfall intensity	μ_{radar} (mm)	μ_{IMERG} (mm)	RB (%)	MAE (mm)	NMAE (—)	POD	POFA
0.1–1 mm	0.42	1.00	137	0.72	1.72	0.47	0.55
1–2 mm	1.38	1.90	38	1.36	0.99	0.75	0.21
2–4 mm	2.66	2.76	3	2.09	0.79	0.83	0.14
4–8 mm	5.18	3.65	−30	3.35	0.65	0.88	0.10
>8 mm	10.90	4.86	−55	7.00	0.64	0.94	0.07

Remarkably, more than half of IMERG's observations between 0.1 and 1 mm are false alarms (i.e., the corresponding radar rainfall depths are smaller than 0.1 mm). In general, IMERG tends to overestimate lower intensities and underestimate higher intensities.

A comparison of the cumulative distribution functions (CDF) of radar and IMERG rainfall depths clearly indicates that IMERG underestimates both the occurrence (Fig. 3, solid line) as well as the volume (Fig. 3, dashed line) of rainfall amounts smaller than 1 mm (not shown). Although the underestimation is present in all seasons, it is amplified in winter and spring (Fig. 3, upper row): more than 80% (less than 70%) of the radar (IMERG) rainfall amounts are smaller than 1 mm, contributing more than 55% (less than 30%) of the total rainfall volume. Furthermore, less than 1% of the IMERG rainfall amounts are larger than 10 mm, while they contribute more than 10% of the total rainfall volume in winter and spring. Therefore, the systematic overestimation visible in those seasons seems to be related to a small number of very large rainfall values. Although less extreme, similar conclusions apply to summer and fall.

Table 2 shows that IMERG underestimates large radar rainfall depths, while Fig. 3 reveals that the contribution of high rainfall amounts to IMERG's total rainfall accumulation is relatively large. To illustrate the potential relation between the rainfall rate (based on the radar rainfall amount) and IMERG's under- or overestimations, the residuals (IMERG − radar) are analyzed as a function of radar estimates (Fig. 4). The underestimation of large rainfall depths is visible for all seasons, except during winter when higher intensities are absent. For radar rainfall depths larger than 15 mm, the difference is almost always below the 0 mm line. On the other hand, IMERG overestimates low intensities, especially in winter and spring.

c. Echo top height

To study the relationship between IMERG's performance and the vertical extent of the precipitation system, the rainfall estimates are coupled (pixel by pixel) with ETH observations. The seasonal distribution of ETH observations shows that the variability of ETH values is smallest in winter and largest in summer (Fig. 5). Almost all rainfall events in winter and more than 95% of the events in spring and fall have ETH values below 6 km. The highest ETH values occur during summer, the season associated with most (convective) high-intensity events in the Netherlands.

The performance of IMERG categorized by ETH clearly shows that shallow events (low ETH, 1–3 km) are the most challenging to detect (POD = 37%, Table 3). The increasing ability of IMERG to detect rainfall for higher ETH goes at the expense of a larger positive bias and a slightly higher NMAE. To unravel the performance of IMERG per ETH category even further, Fig. 6 decomposes the results from Fig. 4 per ETH category (again both hits and misses are considered). Two conclusions can be derived from Fig. 6 which are valid for all seasons except winter, when both high-intensity events and high ETH hardly occur: 1) most observations linked with low ETH are misses or underestimations of low rainfall amounts and 2) the number of (severe) overestimations is amplified with higher ETH.

d. Source of observation

Categorizing the IMERG estimates based on its constituent available during the observation reveals that, surprisingly, both PMW-morph and morph+IR < 50% exceed the detection performance of PMW observations (Table 4). Both morphing and inclusion of IR information increase the occurrence of false alarms. Increasing the contribution weight of IR to more than 50% reduces IMERG's detection performance. All sources systematically overestimate the amount of rain once they correctly detect rainfall, although this effect is strongest for PMW observations.

Each category shows outliers of IMERG rainfall estimates exceeding 15 mm when radar estimates are smaller than 5 mm (Fig. 7). The inclusion of morphing and IR data seems to decrease the magnitude of overestimation, especially for low-intensity events. Figure 7 suggests that severe overestimations derived from PMW observations might propagate into the morphed estimates. Most of the PMW observations are performed by MHS, followed by SSMIS (Fig. 7, two upper rows). All five radiometers show a similar pattern (Fig. 7, two upper rows): they all overestimate low-intensity events and underestimate high-intensity events. Most of the observations are performed by MHS, followed by SSMIS. GMI (the radiometer aboard the core satellite) does not seem to perform better than the other constellation sensors.

Finally, the performance of each source is evaluated per ETH class (Fig. 8). Rainfall associated with low ETH is the most difficult to detect, especially for morph+IR > 50%. The POD clearly increases with higher ETH, although this increase is limited for the POD in case of medium ETH deduced from morph+IR > 50%. For medium and high ETH, the largest RB, MAE, and NMAE result from PMW observations. For low ETH, the largest values of the error metrics (RB, MAE, and NMAE) are reported for morph+IR > 50%. It should be noted that the limited number of observations when both radar and IMERG detect rainfall in the low ETH morph+IR > 50% category, which mainly originates from the winter season, increases the uncertainty of the reported bias.

Among the five PMW instruments, SSMIS has the lowest RB and (N)MAE for all ETH categories (bottom row Fig. 8). Additionally, of all five radiometers, SSMIS best detects low and medium ETH events. GMI has the second-best POD score for both low and medium ETH, but the worst POD score for
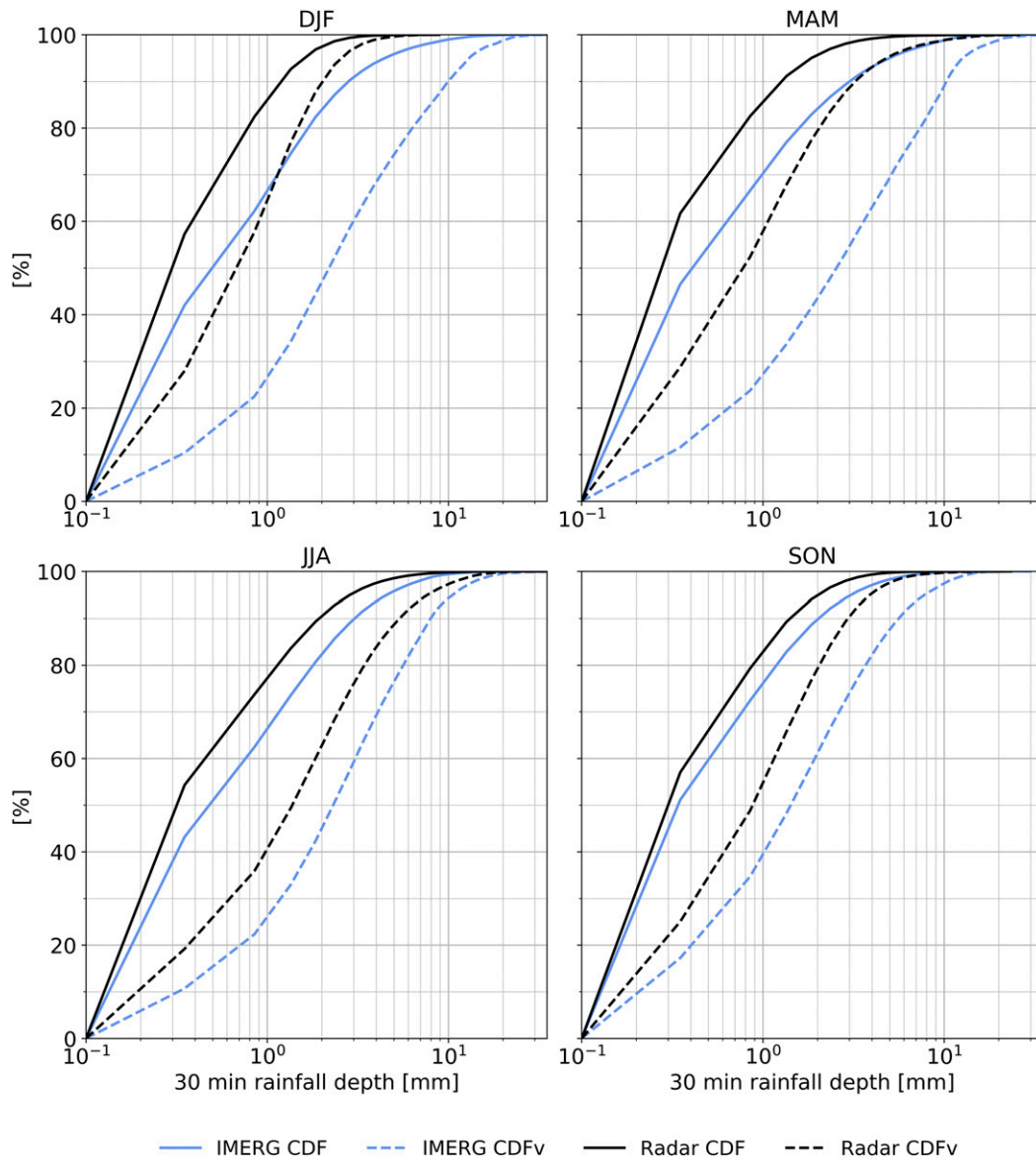
FIG. 3. Seasonal cumulative distribution functions of rainfall depth occurrence (CDF; solid line) and volume (CDF$_v$; dashed line) for radar (black) and IMERG (blue) on half-hourly time intervals for the January 2015–December 2019 period. Only paired observations where both radar and IMERG exceed 0.1 mm per 30 min are considered (hits). The CDFs are calculated with a 0.5-mm bin width.

high ETH. Furthermore, the RB and (N)MAE of GMI estimates are quite high. ASMR-2 has the highest RB and MAE for all three ETH categories. The sounders (ATMS and MHS) have low POD scores for both low and medium ETH, but the highest POD score for high ETH. Furthermore, the two sounders have similar (N)MAE and RB values, whereas these values vary among the three imagers.

4. Discussion

Previous research frequently reported IMERG's underestimation of high-intensity events (Fang et al. 2019; Freitas et al.

2020; Maranan et al. 2020) and its overestimation of low-intensity events (Foelsche et al. 2017; Anjum et al. 2019), features of IMERG we also observed (Fig. 4). However, the magnitude of (the overall) overestimation found in this study is considerably larger. Before we elaborate on probable causes, we would like to mention that most of these studies employed IMERG-F, in which (severe) overestimations are corrected by means of gauge adjustment (Foelsche et al. 2017; Tapiador et al. 2019; Hosseini-Moghari and Tang 2020). Our aim was to evaluate satellite-only products, as identifying the sources of error contributes to better estimates over both ungauged and gauged areas. Two possible
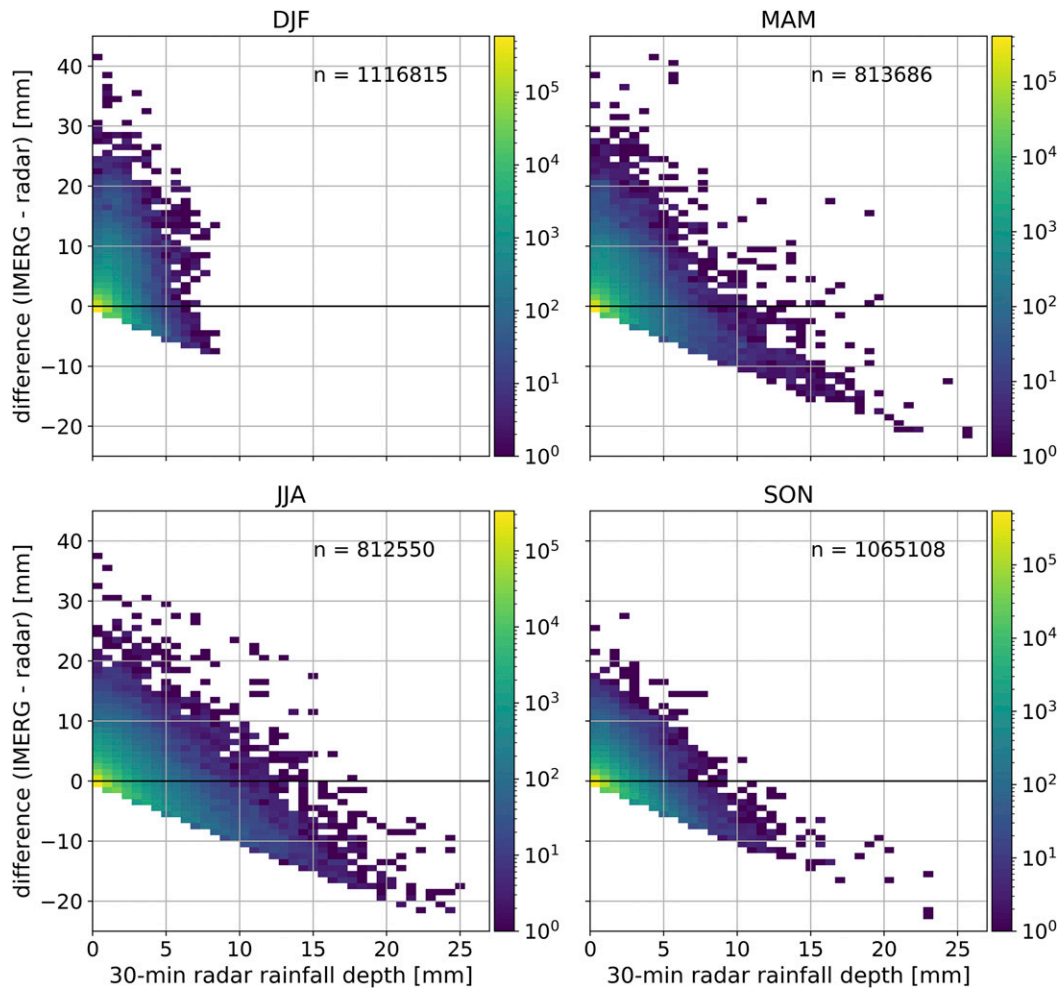
FIG. 4. Seasonal scatter density plots of the difference between IMERG and radar estimates (IMERG − radar) vs the radar rainfall depths on half-hourly time intervals for the January 2015–December 2019 period. Only paired observations where radar exceed 0.1 mm per 30 min are considered (hits and misses).

explanations are discussed, which together may cause these severe overestimations.

First, it should be emphasized that the current analysis is focused on a relatively high-latitude location compared to their more frequently studied tropical and semiarid areas. Small rainfall depths, which frequently occur in this area (Fig. 3), are systematically overestimated by IMERG (Fig. 4). In contrast, tropical precipitation is often intense and short. The current analysis shows that the intensity of convective events is frequently underestimated by IMERG. This effect inherently reduces the bias when IMERG is validated over a longer time period.

Our findings are in line with studies focusing on mid- to high latitudes. A study that validated all IMERG runs over Austria (located around 47°N), where the WEGN gridded rain gauge dataset was used as reference, reported relatively large overestimations for IMERG-L (Foelsche et al. 2017). However, this study does not mention which intensities were overestimated. Over Alaska, where daily precipitation accumulation observations

from 155 (automatic) stations were used as reference, IMERG Early Run (IMERG-E) was found to both systematically overestimate the amount of precipitation as well as to overestimate the occurrence of large rainfall depths (Gowan and Horel 2020). Their findings are similar to our results in Fig. 3. Over Germany (at similar latitudes as the Netherlands), where a gauge-adjusted quality-controlled dataset from the German Weather Service was used as reference, the systematic overestimation of IMERG-F was found to be amplified in winter (Ramsauer et al. 2018).

Second, differences (in magnitude) with previous studies might be related to the implemented IMERG version and its corresponding GPROF version. GPROF is the algorithm responsible for the PMW precipitation retrieval from all sensors belonging to the GPM constellation [detailed information can be found in Kummerow et al. (2015) and Randel et al. (2020)]. Another study over the Netherlands, which evaluated the first IMERG version (V03D), did not find a consistent overestimation in the uncalibrated IMERG-F (Rios Gaona et al. 2016). A brief evaluation of the same time series as Rios Gaona et al. (2016)
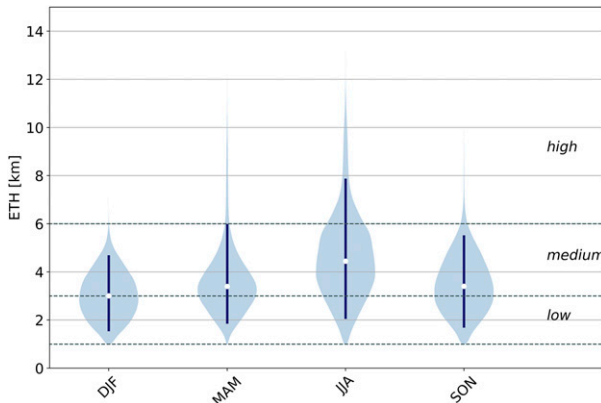
FIG. 5. Seasonal probability density function (PDF; a kernel density estimation is used to compute the PDF) of the ETH values, from January 2015 to December 2019. ETH values are only considered between 1 and 15 km and when the paired radar rainfall depth observations exceed 0.1 mm per 30 min.

TABLE 3. IMERG performance on a pixel-by-pixel basis decomposed into three (low, medium, and high) ETH categories. POFA cannot be calculated since ETH observations are only considered when the half-hourly radar estimates are above 0.1 mm. Therefore, there are no radar observations below 0.1 mm within this selection. Other details are the same as stated in Table 1.

ETH	μ_{radar} (mm)	μ_{IMERG} (mm)	RB (%)	MAE (mm)	NMAE (—)	POD
Low 1–3 km	0.46	0.72	57	0.52	1.14	0.37
Medium 3–6 km	0.81	1.30	60	0.96	1.18	0.70
High 6–15 km	1.72	3.00	75	2.18	1.27	0.93

with different IMERG versions (not shown) using the uncalibrated estimates, revealed an enhanced overestimation of the annual accumulation for each new version, except from V05 to V06 (the only version update without significant GPROF changes). Although this brief evaluation involved the uncalibrated IMERG-F and a shorter period of evaluation, a similar trend is expected for different years and IMERG runs. However, since multiple algorithms are combined within IMERG, it is hard to narrow down the change in algorithms causing the larger overestimations.

Our result that IMERG's detection performance is lowest for the morph+IR > 50% category (Fig. 8) agrees with previous studies, in which both misses and false alarms are higher for IR compared to PMW retrieval (Gebregiorgis et al. 2017). Furthermore, retrieval based on PERSIANN-CCS (the algorithm responsible for IR-based precipitation retrieval in IMERG) is found to be limited to the areas with the coldest brightness temperatures and highest rain rates (Kirstetter et al. 2018). This is in line with our results, as we found a high POD for high ETH for the morph+IR > 50% category and a (very) low POD for low and medium ETH compared to the other categories (Fig. 8). Most observations in this category occur during winter (not shown), which at least partly explains the absence of higher rainfall depths (Fig. 7).

It is likely that the overestimations from radiometers propagate into the morphed estimates due to interpolation over time. This is already shown by Tan et al. (2016), who studied the performance of the different sources contributing to IMERG. In contrast to the results of Tan et al. (2016), GMI is not found to have a more reliable performance compared to the other sensors (Figs. 7, 8). However, the number of data points in Tan et al. (2016) is much smaller (e.g., n = 438 for GMI observations, including nonrainy pixels), the reference product is different (gauges and a level-3 radar product are used separately as reference), and the implemented IMERG version and run are different (Final run, V03).

The results of the current evaluation also differ from those reported by Tang et al. (2014): they found that SSMIS has the largest overestimations of all sensors, while Table 4 and Fig. 8

show that SSMIS has the lowest RB and MAE. However, SSMIS did not have a specific algorithm yet and hence Tang and colleagues implemented a revised version of GPROF2004 for SSMIS observations. Currently, all sounders and imagers contributing to IMERG use the same GPROF algorithm. A similar kind of reasoning can explain why Tang et al. (2014) report larger differences between sounders and imagers: imagers used GPROF 2010, while sounders implemented another algorithm.

The large number of low-intensity events (in general found to be overestimated by IMERG) in combination with the limited occurrence of higher intensities (in general found to be underestimated by IMERG) may result in the poor performance of IMERG found during winter. Accurate retrieval of snow events is shown to be challenging for IMERG (Cui et al. 2019), which might explain the lower performance of IMERG during winter compared to other seasons. However, snowfall was reported over a part of the Netherlands for less than fifteen days during the study period (2015–19). Hence, we expect that role of snow is limited for our results.

The seasonal POD score follows the seasonal cycle of the ETH: lowest for winter, followed by fall, and clearly the best for summer (Table 1). As shown in Fig. 6, high radar rainfall depths are associated with high ETH. Since these events are mostly convective events, they occur during summer when temperatures are higher. The seasonal cycle in ETH as observed in our study and its relation with rainfall intensity is in agreement with the results reported by Aberson (2011). Remarkably, fall is the only season where those severe overestimations are reduced despite the frequent occurrence of low-intensity events. This reduction results in the smallest positive RB and MAE of the four seasons. In-depth research with, for instance, 3D radar reflectivity data, is needed to understand why fall is different from the other seasons for otherwise similar precipitation characteristics such as ETH and rainfall intensity.

Our findings are especially relevant for ungauged areas with similar climates or regions where light, long-duration (shallow) precipitation events frequently occur. Because IMERG relies heavily on PMW observations, we recommend further research on passive retrieval of shallow and low-intensity precipitation and how the seasonality of precipitation characteristics influences this retrieval. This could be done through coupling ground-based (3D) radar reflectivity observations with profiles from the dual-frequency precipitation radar aboard the GPM core satellite. Subsequently, these results could be related to
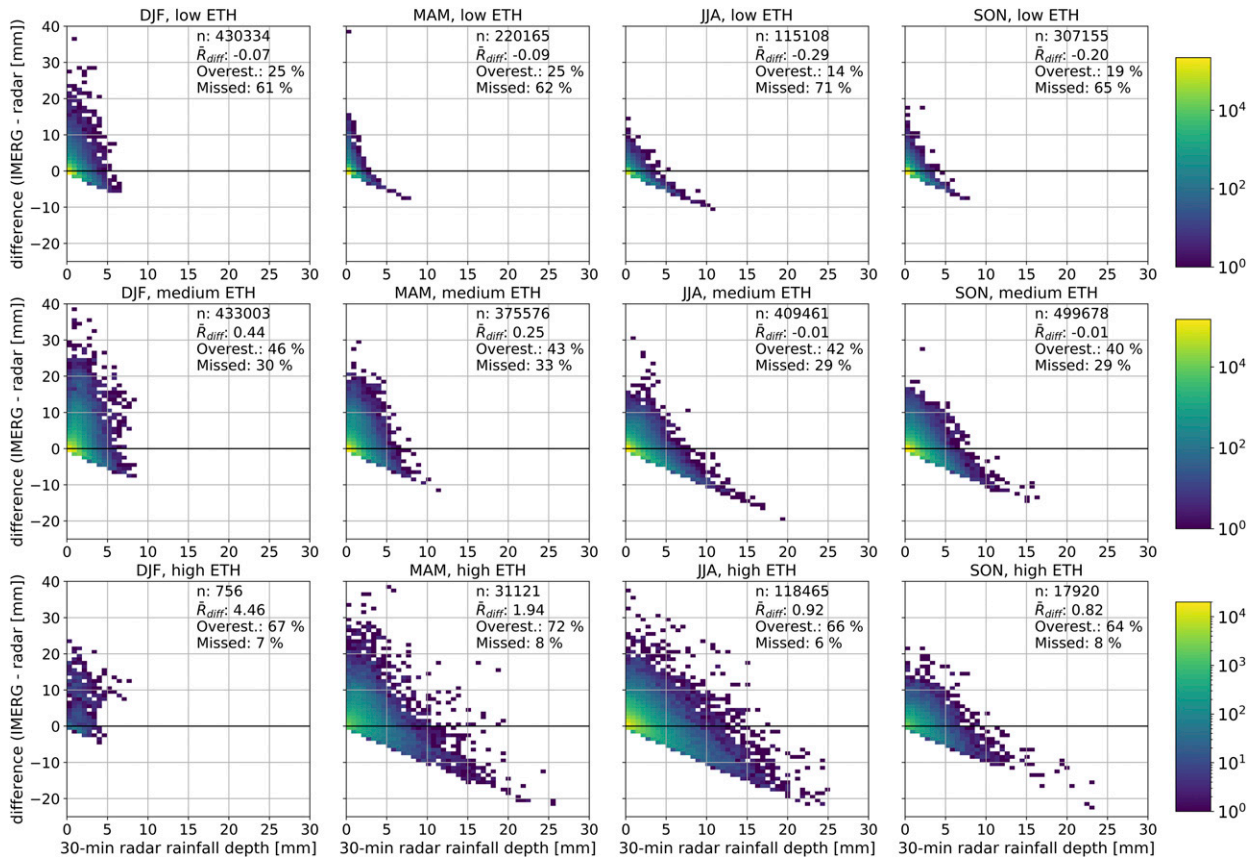
FIG. 6. As in Fig. 4, but decomposed into three different ETH categories (low, 1–3 km; medium, 3–6 km; and high, 6–15 km). The term $\bar{R}_{\mathrm{diff}}$ refers to the mean difference between the radar and IMERG estimates. Furthermore, the percentage of events overestimated (referred to as "Overest") and missed by IMERG are provided for each category (it should be noted that the percentage of overestimated events is based on events that are either "hits" or "misses").

radiometer retrieval as the core satellite also carries the GPM Microwave Imager (GMI).

This study employed the KNMI radar data as ground truth, despite its own limitations. Examples of radar artifacts are visible in Fig. 2 in the north of the Netherlands (striped pattern, especially visible during fall) and over "de Maasvlakte (2)" (blue grid cell visible in the lower panel of Fig. 2). Both are known deficiencies of the radar product. The former is most likely related to trees in the neighborhood of the radar and the latter to cranes

and containers in the Port of Rotterdam, resulting in strong backscatter in cases of radar beam superrefraction. Furthermore, ground-based weather radar is known to underestimate extreme rainfall amounts in short time intervals (Overeem et al. 2009a,b; Hazenberg et al. 2011). However, the consequences for the obtained results are expected to be limited: IMERG itself is found to have difficulties capturing high-intensity events while the reported overestimations of IMERG are related to low rainfall amounts. Therefore, a better performance of the radar is

TABLE 4. IMERG performance on a pixel-by-pixel basis decomposed into the source of observation categories as defined in section 2 (including the five radiometers). Other details are the same as stated in Table 1.

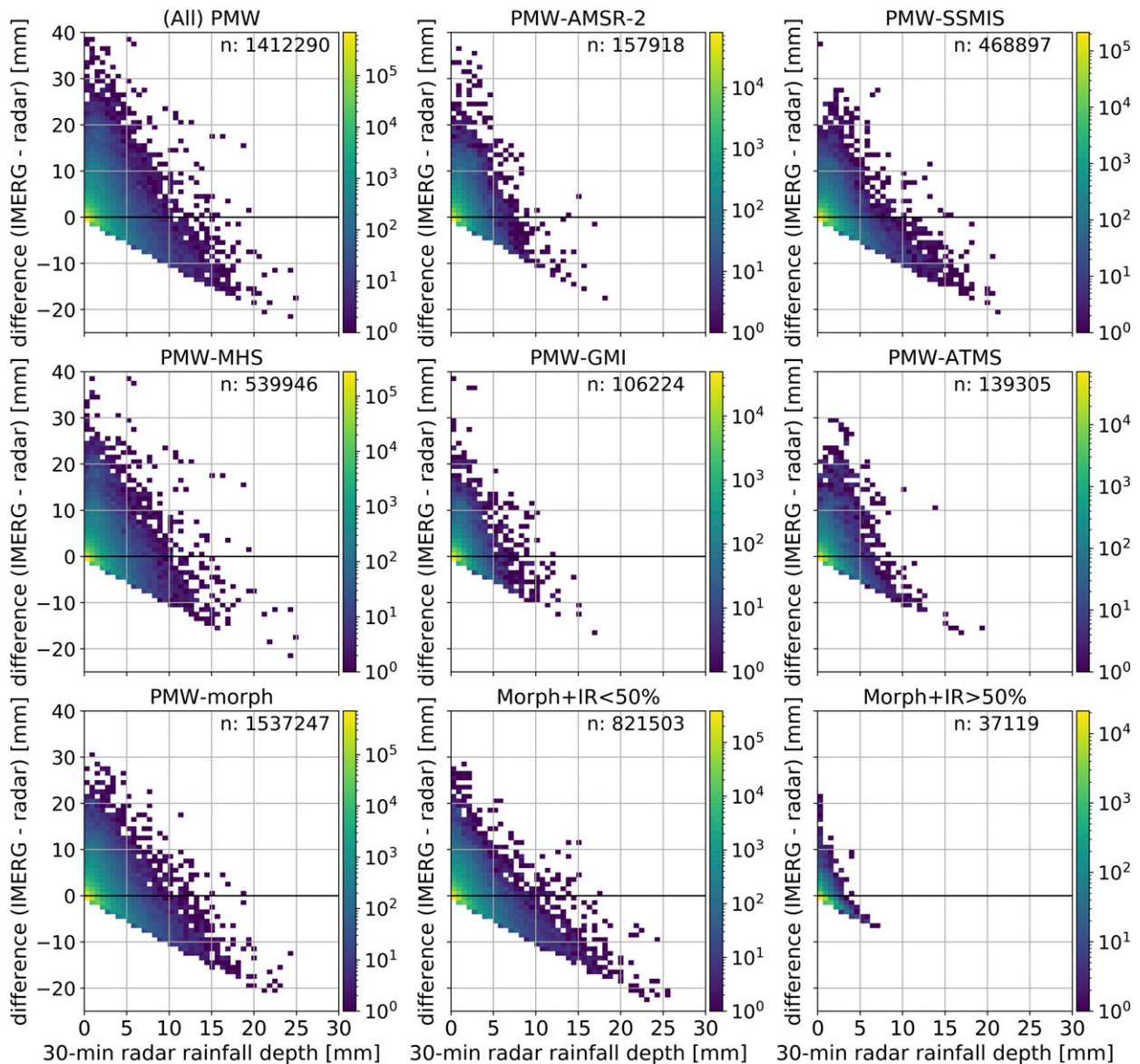| Source	μ_{radar} (mm)	μ_{IMERG} (mm)	RB (%)	MAE (mm)	NMAE (—)	POD	POFA
PMW	0.80	1.42	79	1.04	1.30	0.48	0.41
PMW-morph	0.76	1.20	58	0.88	1.16	0.55	0.45
morph + <50% IR	0.76	1.21	59	0.93	1.22	0.52	0.53
morph + >50% IR	0.68	1.01	48	0.86	1.26	0.28	0.61
AMSR-2	0.80	1.93	141	1.42	1.78	0.45	0.42
SMISS	0.77	1.19	55	0.84	1.09	0.55	0.43
GMI	0.74	1.49	100	1.06	1.43	0.52	0.32
ATMS	0.80	1.42	78	1.35	1.07	0.44	0.41
MHS	0.84	1.52	82	1.34	1.12	0.44	0.41

FIG. 7. Scatter density plots of the difference between IMERG and radar estimates (IMERG − radar) vs the 30-min radar rainfall depths categorized according to the source of observation (as defined in section 2, including the five radiometers) for the January 2015–December 2019 period. Note that the number of data points of category > 50% IR data is small. Only paired observations where radar exceed 0.1 mm per 30 min are considered (hits and misses).

expected to even amplify our finding that IMERG underestimates high-intensity rainfall.

5. Conclusions

This study validated IMERG Late Run V06B precipitation estimates at the spatiotemporal resolution of IMERG (0.1° × 0.1°, 30 min) over the Netherlands. A gauge-adjusted radar dataset over a five-year period (2015–19) was employed as reference. To the best of our knowledge, this is the first study that assessed IMERG performance over such a long period of time over a midlatitude country. Furthermore, we explored the relation between IMERG's performance and: 1) seasons, 2) rainfall intensity, 3) echo top height (ETH), and 4) the source of observation.

IMERG systematically overestimates low-intensity rainfall, more pronounced during winter and spring. Simultaneously, IMERG is found to underestimate higher rainfall intensities. IMERG's detection performance increases with higher ETH and higher rainfall intensity. Hence, the probability of detection is relatively low in winter (frequent occurrence of low-intensity and shallow precipitation events) and high in summer (when most convective events occur). The probability of false alarms decreases with higher intensity.
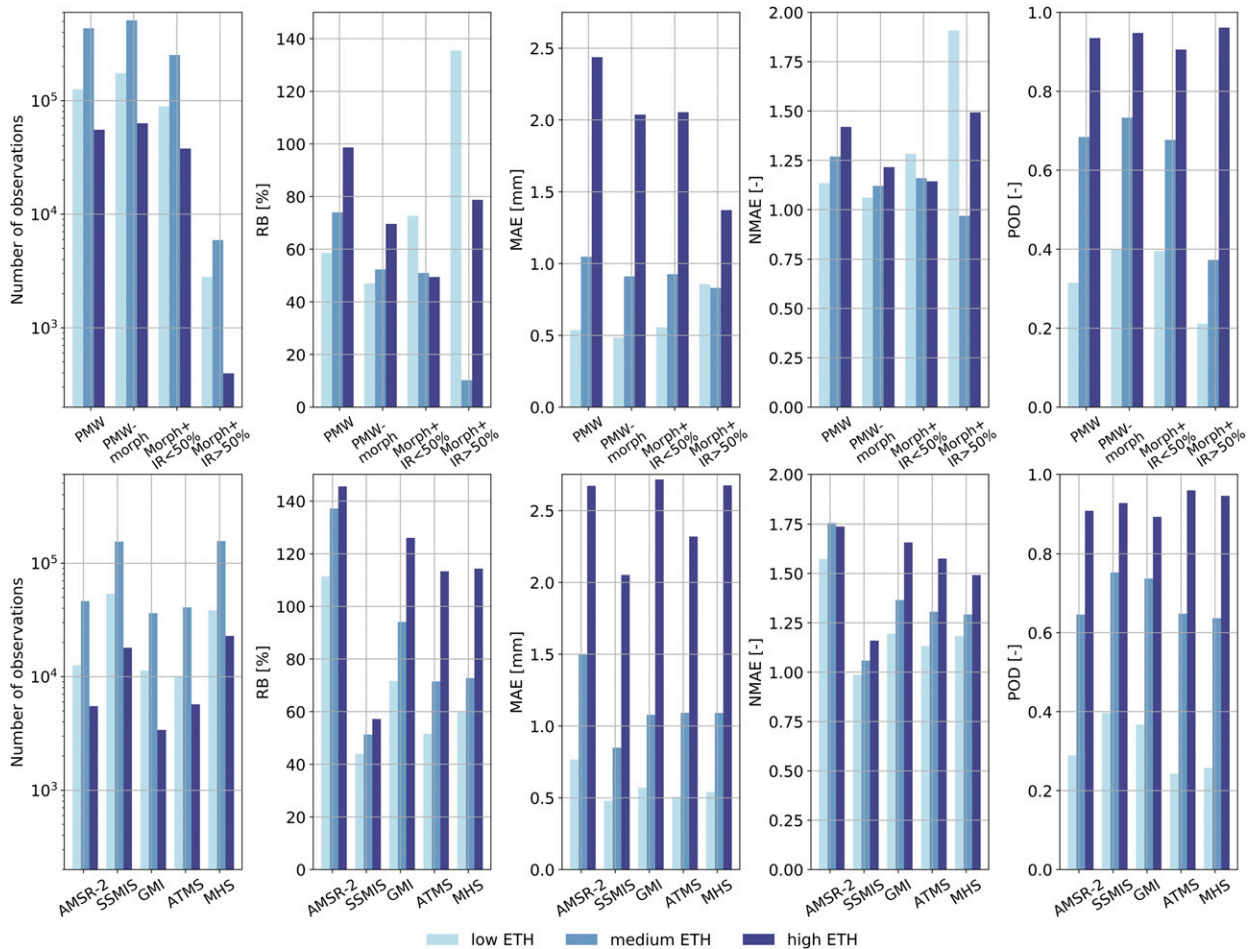
FIG. 8. Number of observations, error metrics, and validation metrics for the four different observation sources and radiometers (both metrics and sources as defined in section 2) with the same ETH categories as defined in Fig. 6. Only paired observations where both radar and IMERG exceed 0.1 mm per 30 min are considered (hits).

PMW-based precipitation estimates are prone to overestimations and the inclusion of IR data is found to decrease the detection performance of IMERG. All sources, including the five radiometers, either miss or highly overestimate low-intensity events. The performance among the imagers (ASMR-2, SSMIS, and GMI) varies: while SSMIS has the lowest RB and (N)MAE, ASMR-2 has the highest RB and MAE of all five instruments. In contrast, the sounders (ATMS and MHS) have a comparable performance. Furthermore, both sounders have a low POD score for low and medium ETH compared to the imagers. For all sources, both the correct identification of wet pixels as well as the accuracy of the rainfall amount are most challenging for shallow rainfall events. Hence, we identify space-based shallow and low-intensity precipitation retrieval as an important topic for future research.

REFERENCES

Aberson, K., 2011: The spatial and temporal variability of the vertical dimension of rainstorms and their relation with precipitation intensity. Internal Rep. IR 2011-03, KNMI, 38 pp., https://cdn.knmi.nl/knmi/pdf/bibliotheek/knmipubIR/IR2011-03.pdf.

Anjum, M. N., and Coauthors, 2019: Assessment of IMERG-V06 precipitation product over different hydro-climatic regimes in the Tianshan mountains, north-western China. *Remote Sens.*, **11**, 2314, https://doi.org/10.3390/rs11192314.

Arulraj, M., and A. P. Barros, 2017: Shallow precipitation detection and classification using multifrequency radar observations and model simulations. *J. Atmos. Oceanic Technol.*, **34**, 1963–1983, https://doi.org/10.1175/JTECH-D-17-0060.1.

Asong, Z. E., S. Razavi, H. S. Wheater, and J. S. Wong, 2017: Evaluation of Integrated Multisatellite Retrievals for GPM (IMERG) over southern Canada against ground precipitation observations: A preliminary assessment. *J. Hydrometeor.*, **18**, 1033–1050, https://doi.org/10.1175/JHM-D-16-0187.1.

Attema, J. J., and G. Lenderink, 2014: The influence of the North Sea on coastal precipitation in the Netherlands in the present-day and future climate. *Climate Dyn.*, **42**, 505–519, https://doi.org/10.1007/s00382-013-1665-4.

Beekhuis, H., and I. Holleman, 2008: From pulse to product: Highlights of the digital-IF upgrade of the Dutch national radar network. Fifth European Conf. on Radar in Meteorology and Hydrology, Helsinki, Finland, Finnish Meteorological Institute, 3 pp., https://cdn.knmi.nl/system/data_center_publications/files/000/068/061/original/erad2008drup_0120.pdf?1495621011.

Casella, D., G. Panegrossi, P. Sanò, A. C. Marra, S. Dietrich, B. T. Johnson, and M. S. Kulie, 2017: Evaluation of the GPM-DPR snowfall detection capability: Comparison with CloudSat-CPR. *Atmos. Res.*, **197**, 64–75, https://doi.org/10.1016/j.atmosres.2017.06.018.

Chen, F., and X. Li, 2016: Evaluation of IMERG and TRMM 3B43 monthly precipitation products over mainland China. *Remote Sens.*, **8**, 472, https://doi.org/10.3390/rs8060472.

Cui, W., X. Dong, B. Xi, Z. Feng, and J. Fan, 2019: Can the GPM IMERG final product accurately represent MCSs' precipitation characteristics over the central and eastern United States? *J. Hydrometeor.*, **21**, 39–57, https://doi.org/10.1175/JHM-D-19-0123.1.

Dezfuli, A. K., C. M. Ichoku, K. I. Mohr, and G. J. Huffman, 2017: Precipitation characteristics in West and East Africa from satellite and in situ observations. *J. Hydrometeor.*, **18**, 1799–1805, https://doi.org/10.1175/JHM-D-17-0068.1.

Fang, J., W. Yang, Y. Luan, J. Du, A. Lin, and L. Zhao, 2019: Evaluation of the TRMM 3B42 and GPM IMERG products for extreme precipitation analysis over China. *Atmos. Res.*, **223**, 24–38, https://doi.org/10.1016/j.atmosres.2019.03.001.

Foelsche, O. S. U., G. Kirchengast, J. Fuchsberger, J. Tan, and W. A. Petersen, 2017: Evaluation of GPM IMERG Early, Late, and Final rainfall estimates using WegenerNet gauge data in southeastern Austria. *Hydrol. Earth Syst. Sci.*, **21**, 6559–6572, https://doi.org/10.5194/hess-21-6559-2017.

Foster, J. L., and Coauthors, 2012: Passive microwave remote sensing of the historic February 2010 snowstorms in the Middle Atlantic region of the USA. *Hydrol. Processes*, **26**, 3459–3471, https://doi.org/10.1002/hyp.8418.

Freitas, E. S., and Coauthors, 2020: The performance of the IMERG satellite-based product in identifying sub-daily rainfall events and their properties. *J. Hydrol.*, **589**, 125–128, https://doi.org/10.1016/j.jhydrol.2020.125128.

Gebregiorgis, A. S., P.-E. Kirstetter, Y. E. Hong, N. J. Carr, J. J. Gourley, W. Petersen, and Y. Zheng, 2017: Understanding overland multisensor satellite precipitation error in TMPA-RT products. *J. Hydrometeor.*, **18**, 285–306, https://doi.org/10.1175/JHM-D-15-0207.1.

——, ——, ——, J. J. Gourley, G. J. Huffman, W. A. Petersen, X. Xue, and M. R. Schwaller, 2018: To what extent is the Day 1 GPM IMERG satellite precipitation estimate improved as compared to TRMM TMPA-RT? *J. Geophys. Res. Atmos.*, **123**, 1694–1707, https://doi.org/10.1002/2017JD027606.

Gowan, T. A., and J. D. Horel, 2020: Evaluation of IMERG-E precipitation estimates for fire weather applications in Alaska. *Wea. Forecasting*, **35**, 1831–1843, https://doi.org/10.1175/WAF-D-20-0023.1.

Hazenberg, P., H. Leijnse, and R. Uijlenhoet, 2011: Radar rainfall estimation of stratiform winter precipitation in the Belgian Ardennes. *Water Resour. Res.*, **47**, W02507, https://doi.org/10.1029/2010WR009068.

Hosseini-Moghari, S.-M., and Q. Tang, 2020: Validation of GPM IMERG V05 and V06 precipitation products over Iran. *J. Hydrometeor.*, **21**, 1011–1037, https://doi.org/10.1175/JHM-D-19-0269.1.

Hou, A. Y., and Coauthors, 2014: The Global Precipitation Measurement mission. *Bull. Amer. Meteor. Soc.*, **95**, 701–722, https://doi.org/10.1175/BAMS-D-13-00164.1.

Huffman, G. J., D. T. Bolvin, E. J. Nelkin, and J. Tan, 2019: Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation. NASA Tech. Doc., 77 pp., https://gpm.nasa.gov/sites/default/files/document_files/IMERG_doc_190909.pdf.

——, and Coauthors, 2020: NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG). Algorithm Theoretical Basis Doc., version 06, 39 pp., https://gpm.nasa.gov/sites/default/files/2020-05/IMERG_ATBD_V06.3.pdf.

Kirstetter, P.-E., N. Karbalaee, K. Hsu, and Y. Hong, 2018: Probabilistic precipitation rate estimates with space-based infrared sensors. *Quart. J. Roy. Meteor. Soc.*, **144**, 191–205, https://doi.org/10.1002/qj.3243.

Kummerow, C. D., D. L. Randel, M. Kulie, N.-Y. Wang, R. Ferraro, S. Joseph Munchak, and V. Petkovic, 2015: The evolution of the Goddard profiling algorithm to a fully parametric scheme. *J. Atmos. Oceanic Technol.*, **32**, 2265–2280, https://doi.org/10.1175/JTECH-D-15-0039.1.

Lorenz, C., and H. Kunstmann, 2012: The hydrological cycle in three state-of-the-art reanalyses: Intercomparison and performance analysis. *J. Hydrometeor.*, **13**, 1397–1420, https://doi.org/10.1175/JHM-D-11-088.1.

Maranan, M., A. H. Fink, P. Knippertz, L. K. Amekudzi, W. A. Atiah, and M. Stengel, 2020: A process-based validation of GPM IMERG and its sources using a mesoscale rain gauge network in the West African forest zone. *J. Hydrometeor.*, **21**, 729–749, https://doi.org/10.1175/JHM-D-19-0257.1.

Mohammed, S. A., M. A. Hamouda, M. T. Mahmoud, and M. M. Mohamed, 2020: Performance of GPM-IMERG precipitation products under diverse topographical features and multiple-intensity rainfall in an arid region. *Hydrol. Earth Syst. Sci. Discuss.*, https://doi.org/10.5194/hess-2019-547.

Navarro, A., E. García-Ortega, A. Merino, J. L. Sánchez, C. Kummerow, and F. J. Tapiador, 2019: Assessment of IMERG precipitation estimates over Europe. *Remote Sens.*, **11**, 2470, https://doi.org/10.3390/rs11212470.

Overeem, A., T. A. Buishand, and I. Holleman, 2009a: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resour. Res.*, **45**, W10424, https://doi.org/10.1029/2009WR007869.

——, I. Holleman, and A. Buishand, 2009b: Derivation of a 10-year radar-based climatology of rainfall. *J. Appl. Meteor. Climatol.*, **48**, 1448–1463, https://doi.org/10.1175/2009JAMC1954.1.

——, H. Leijnse, and R. Uijlenhoet, 2011: Measuring urban rainfall using microwave links from commercial cellular communication networks. *Water Resour. Res.*, **47**, W12505, https://doi.org/10.1029/2010WR010350.

——, R. Uijlenhoet, and H. Leijnse, 2020: Full-year evaluation of nonmeteorological echo removal with dual-polarization fuzzy logic for two C-band radars in a temperate climate. *J. Atmos. Oceanic Technol.*, **37**, 1643–1660, https://doi.org/10.1175/JTECH-D-19-0149.1.

Petković, V., and C. D. Kummerow, 2016: Understanding the sources of satellite passive microwave rainfall retrieval systematic errors over land. *J. Appl. Meteor. Climatol.*, **56**, 597–614, https://doi.org/10.1175/JAMC-D-16-0174.1.

Prakash, S., A. K. Mitra, A. AghaKouchak, Z. Liu, H. Norouzi, and D. S. Pai, 2018: A preliminary assessment of GPM-based multi-satellite precipitation estimates over a monsoon dominated region. *J. Hydrol.*, **556**, 865–876, https://doi.org/10.1016/j.jhydrol.2016.01.029.

Ramsauer, T., T. Weiß, and P. Marzahn, 2018: Comparison of the GPM IMERG Final precipitation product to RADOLAN weather radar data over the topographically and climatically diverse Germany. *Remote Sens.*, **10**, 2029, https://doi.org/10.3390/rs10122029.

Randel, D. L., C. D. Kummerow, and S. Ringerud, 2020: The Goddard Profiling (GPROF) precipitation retrieval algorithm. *Satellite Precipitation Measurement*, Vol. 1, V. Levizzani et al., Eds., Advances in Global Change Research, Vol. 67, Springer, 141–152, https://doi.org/10.1007/978-3-030-24568-9_8.

Rios Gaona, M. F., A. Overeem, H. Leijnse, and R. Uijlenhoet, 2016: First-year evaluation of GPM rainfall over the Netherlands: IMERG Day 1 Final Run (V03D). *J. Hydrometeor.*, **17**, 2799–2814, https://doi.org/10.1175/JHM-D-16-0087.1.

Saltikoff, E., and Coauthors, 2019: An overview of using weather radar for climatological studies: Successes, challenges, and potential. *Bull. Amer. Meteor. Soc.*, **100**, 1739–1752, https://doi.org/10.1175/BAMS-D-18-0166.1.

Skofronick-Jackson, G., D. Kirschbaum, W. Petersen, G. Huffman, C. Kidd, E. Stocker, and R. Kakar, 2018: The Global Precipitation Measurement (GPM) mission's scientific achievements and societal contributions: Reviewing four years of advanced rain and snow observations. *Quart. J. Roy. Meteor. Soc.*, **144**, 27–48, https://doi.org/10.1002/qj.3313.

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

——, ——, P.-E. Kirstetter, and Y. Tian, 2017: Performance of IMERG as a function of spatiotemporal scale. *J. Hydrometeor.*, **18**, 307–319, https://doi.org/10.1175/JHM-D-16-0174.1.

——, G. J. Huffman, D. T. Bolvin, and E. J. Nelkin, 2019: IMERG V06: Changes to the morphing algorithm. *J. Atmos. Oceanic Technol.*, **36**, 2471–2482, https://doi.org/10.1175/JTECH-D-19-0114.1.

Tang, L., Y. Tian, and X. Lin, 2014: Validation of precipitation retrievals over land from satellite-based passive microwave sensors. *J. Geophys. Res. Atmos.*, **119**, 4546–4567, https://doi.org/10.1002/2013JD020933.

Tapiador, F. J., A. Navarro, E. García-Ortega, A. Merino, J. L. Sánchez, C. Marcos, and C. Kummerow, 2019: The contribution of rain gauges in the calibration of the IMERG product: Results from the first validation over Spain. *J. Hydrometeor.*, **21**, 161–182, https://doi.org/10.1175/JHM-D-19-0116.1.

Wang, H., and B. Yong, 2020: Quasi-Global evaluation of IMERG and GSMaP precipitation products over land using gauge observations. *Water*, **12**, 243, https://doi.org/10.3390/w12010243.

Xu, R., F. Tian, L. Yang, H. Hu, H. Lu, and A. Hou, 2017: Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over southern Tibetan Plateau based on a high-density rain gauge network. *J. Geophys. Res. Atmos.*, **122**, 910–924, https://doi.org/10.1002/2016JD025418.