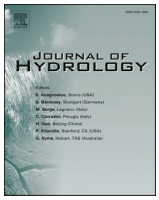
Journal of Hydrology 620 (2023) 129563

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol



Research papers

How well does the IMERG satellite precipitation product capture the timing of precipitation events?

Runze Li [a,*], Clement Guilloteau [a], Pierre-Emmanuel Kirstetter [b,c], Efi Foufoula-Georgiou [a,d]

[a] Department of Civil and Environmental Engineering, University of California, Irvine, Irvine, CA, USA
[b] Hydrometeorology and Remote Sensing Laboratory, University of Oklahoma, Norman, OK, USA
[c] NOAA/Severe Storms Laboratory, Norman, OK, USA
[d] Department of Earth System Science, University of California, Irvine, Irvine, CA, USA

ARTICLE INFO

Keywords:
Precipitation
Satellite precipitation product
Sub-daily evaluation
GPM
IMERG
Event-based performance

ABSTRACT

Precipitation occurs in the form of discrete "events" and the event characteristics (event duration, depth, peak intensity, start/end time) significantly influence the hydrologic response of a basin. Despite this importance, event-based performance of satellite precipitation products has still not been fully investigated to assess limitations in the retrieval algorithms, guide future improvements, and inform hydrologic applications. In this study, we evaluate the precipitation event performance of the GPM IMERG product using as reference the high-resolution ground gauge–radar dataset (GV-MRMS) in the Continental United States (CONUS) at the native IMERG resolution ($0.1^\circ \times 0.1^\circ$, 0.5 h), with a primary focus on the detectability and timing of events. Our results show that IMERG generally overestimates the event duration but underestimates the mean event precipitation intensity in the summer, while the opposite is true for winter. This discrepancy is mostly attributed to the under-representation of short-duration intense events in the summer and long-duration moderate events in the winter in IMERG. In terms of the detection of individual events, about 50% of the reference events are properly detected by IMERG, and conversely, 50% of IMERG events do not match a reference event. However, nearly 40% of the missed or false events result from temporal mismatching of less than 3 h between the retrieved and the reference event. The remaining 60% comes from IMERG not detecting an existing event or inventing a nonexistent event. When IMERG successfully detects an event, the average temporal overlap with the reference event is about 70% of its total duration, which mostly stems from the mistiming of IMERG-derived events. IMERG events tend to start, peak, and end earlier than GV-MRMS events, with national mean shifts of −26 min, −17 min, and −7min, respectively. For about only 20% of all situations the starting time of the event is correctly reproduced by IMERG, and the same applies to the peak time and end time. Our results provide guidance for applications of IMERG at sub-daily scales, as well as new insights for the improvement of satellite retrieval algorithms.

1. Introduction

Our ability to comprehend and monitor precipitation and its global variability is essential to water resource management, food security, and ecological sustainability, with important economic, social and political implications (Hsiang et al., 2013; Kotz et al., 2022; Lobell et al., 2011; Siepielski et al., 2017; Sinha et al., 2017). An important feature that makes precipitation different from most other environmental variables (e.g., temperature, soil moisture, humidity) is that it occurs discretely in space and time in the form of coherent and continuous objects which we call "events" (Haile et al., 2011; Hanel and Maca, 2014; Lochbihler et al., 2017). The occurrence and characteristics of precipitation events are directly linked to multi-scale, multi-type dynamic, thermodynamic, and microphysical atmospheric processes (Berg et al., 2013; Lamjiri et al., 2017; Lochbihler et al., 2017; Wasko et al., 2015) and depending on the characteristics of these events, land surface processes like runoff, infiltration, and soil erosion are affected (Dunkerley, 2017; Dunkerley, 2021; Dunkerley, 2019; Guan et al., 2016). Therefore, studying the properties of precipitation events is important for deepening our understanding of precipitation physics and accurately modeling precipitation-driven processes.

High-resolution multi-satellite precipitation products have become indispensable in recent decades for providing quantitative precipitation estimates globally (McCabe et al., 2017; Skofronick-Jackson et al.,

* Corresponding author.
E-mail address: runzl11@uci.edu (R. Li).

https://doi.org/10.1016/j.jhydrol.2023.129563
Received 29 December 2022; Received in revised form 7 April 2023; Accepted 19 April 2023
Available online 25 April 2023

2017). Consequently, knowing how well they capture the characteristics of precipitation events is crucial to assess their reliability for hydrological applications and provide insight for future improvement of precipitation retrieval algorithms (Freitas et al., 2020; Tadesse and Anagnostou, 2009). However, most published studies assess satellite precipitation products by lumping together all hourly or daily data and computing statistical indicators (Asong et al., 2017; Huang et al., 2016; Smith et al., 2006; Tang et al., 2016), an approach that largely erases the crucial information in the properties of individual precipitation events (Li et al., 2021). Conversely, the existing event-based evaluations are still limited (Freitas et al., 2020; Li et al., 2021; Maranan et al., 2020), resulting in an insufficient understanding of the event-related performance of satellite precipitation products.

A series of works by Nikolopoulos et al. (2013), Mei et al. (2014), and Mei et al. (2016) assessed the event-based performance of satellite products at the basin scale in Europe. Since they aimed to assess the satellites' ability to simulate flood events, they focused on basin-average statistics. More recently, Maranan et al. (2020) compared IMERG-derived events with gauge-derived events in Africa, concentrating on event duration differences. To the authors' knowledge, Freitas et al. (2020) and Li et al. (2021) are the only evaluation studies that analyzed in detail not only event duration, but also further event characteristics such as depth, mean rate, etc. These studies still did not cover all the facets of precipitation events, and many other aspects of satellite-derived precipitation events remain to be investigated. For example, precipitation event characteristics are closely related to the type of precipitation systems which significantly varies among different seasons (Mao et al., 2022), but the seasonal factor has not been analyzed in previous studies (Freitas et al., 2020; Li et al., 2021). Additionally, the potential timing errors in the start, peak, and end time of satellite-derived events have not been investigated, despite the fact that such inaccuracies can result in significant biases in flood event simulations (Dunkerley, 2012; Mei et al., 2016). Assessing event's timing in satellite products can also provide quantitative insight for improving retrieval algorithms (Guilloteau et al., 2018). Previous event-based evaluations have used gauge observations (Freitas et al., 2020; Li et al., 2021; Mei et al., 2014; Mei et al., 2016; Nikolopoulos et al., 2013) as ground reference. However, the fundamental difference between point measurements by gauges and the area-integrated satellite estimates makes the comparison between the two equivocal (unless each satellite pixel is homogeneously covered by several gauges, which is rarely the case). In contrast, the sampling geometry of ground-based weather radars is more consistent with that of satellite estimates, making radars generally more suitable for the evaluation of satellite products, provided that the radars are properly calibrated and bias-corrected with gauge data.

In this study, we present an event-based evaluation of the state-of-the-art Integrated MultisatellitE Retrieval for Global Precipitation Measurement (GPM) (IMERG) product over the Continental United States (CONUS), using as reference the Ground Validation-Multi-Radar/Multi-Sensor (GV-MRMS) product which combines ground radar and gauge measurements. We focus on assessing the performance of IMERG in retrieving the following three progressive event attributes: (1) the spatial variability and statistical distribution of event characteristics (depth, duration, mean event precipitation rate, peak intensity, etc.), (2) the detection rate (and false detection rate) of precipitation events, and (3) the timing (start, peak, and end time) of the detected events. Thanks to the dense ground radar network in the U.S., this study is not only the first systematic event-based evaluation in the U.S., but also the first event-based evaluation with the ground radar-based data as reference. Our analysis reveals seasonal and regional differences in the performance of IMERG and points out to systematic biases that could be addressed in future algorithm developments.

2. Data and methods

2.1. Data

2.1.1. IMERG precipitation dataset

IMERG is the level-3 research-quality gridded global multi-satellite merged precipitation product developed by the U.S. GPM team (Hou et al., 2014; Huffman et al., 2020). The version we use is IMERG V06B Final Run at a resolution of 0.1°, 30 min (Huffman et al., 2019b). The V06 of IMERG is available both in the TRMM era (June 2000–May 2014) and the GPM era (June 2014-present). IMERG integrates multiple sensor sources, including all available passive microwave (PMW) sensors in the GPM constellation of low-Earth-orbit satellites and infrared (IR) sensors on geostationary orbits. Among them, all PMW estimates are globally calibrated against the Combined Radar-Radiometer Algorithm (CORRA) product, which relies on the combined measurements of the Dual-frequency Precipitation Radar (DPR) and the GPM Microwave Radiometer (GMI), both onboard the GPM Core Observatory. All IR estimates are then calibrated by the PMW sensors (Huffman et al., 2020). Such intercalibrations efficiently enforce the uniformity of the multi-source merged IMERG product across space and time.

For constructing continuous gridded estimates, the calibrated PMW estimates and the IR estimates produced by PERSIANN-CCS (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Cloud Classification System) are interpolated using a quasi-Lagrangian time interpolation scheme similar to that of the Climate Prediction Center (CPC) Morphing-Kalman Filter (CMORPH-KF) algorithm (Huffman et al., 2019a). Finally, the gridded estimates are calibrated by the Global Precipitation Climatology Centre (GPCC) gauge monthly analysis. IMERG is among the most accurate state-of-the-art high-resolution satellite precipitation datasets (Guilloteau et al., 2021; Pradhan et al., 2022; Tang et al., 2020), and has been used in a wide range of applications (Orland et al., 2022; Yin et al., 2022; Zhang and Wang, 2023).

2.1.2. GV-MRMS gauge–radar data

The high-quality MRMS quantitative precipitation estimation (QPE) is used here as the ground reference dataset. It integrates precipitation observations from 180 Weather Surveillance Radar 88 Doppler (WSR-88D) and Canadian C-band radar data and approximately 7,000 hourly rain gauges, and also incorporates model analyses as ancillary data (Zhang et al., 2016). Based on MRMS, the further quality-controlled and adjusted GV-MRMS product forms a standardized reference for GPM ground validation (Kirstetter et al., 2014; Kirstetter et al., 2012; Kirstetter et al., 2018). The gridded GV-MRMS QPE used in this study matches the IMERG resolution of 0.1° and 30-minutes and a spatial domain from 20° to 55° N and 60° to 130° W. A Radar Quality Index (RQI, 0–100), which accounts for the distance to the radar, beam blockage and environmental parameters such as the altitude of the freezing level, is applied to filter out lower-quality estimates (RQI less than 60) (Petersen et al., 2020). The gridboxes with an average RQI less than 60 during the study period were entirely eliminated from our analyses. The study period is 2018–2020.

2.2. Methods

2.2.1. Precipitation event definition

In this study, the definition of a precipitation event is based on an Eulerian approach, that is, at each pixel in space, events are defined as continuous periods of half-hourly rain rates of no less than 0.1 mm/h (Li et al., 2021). This simple definition allows an easy-to-implement analysis (as compared with the Lagrangian approach which requires tracking the space–time evolution of precipitation features) that can reveal important information relevant to the hydrologic response of a basin (Ignaccolo and De Michele, 2010). Moreover, the revealed biases in the retrieval could be more directly linked to the local surface type, climate

zone, and latitude, which could be helpful in attributing errors and improving retrieval algorithms.

2.2.2. Event properties

Following Li et al. (2021), for each event, we depict a set of event properties, namely, *event duration* (***T***, *h*), *event depth* (***D***, *mm*) and *mean event precipitation rate* (***R***, *mm/h*), *relative peak position* (***P***), and *peak-to-mean intensity ratio* (***O***). Specifically, ***P*** is the relative position of the peak time of an event, which is expressed as (see Li et al. (2021) for details):

$$P = \frac{t_p - 0.5}{T - 0.5} \tag{1}$$

where t_p (*h*) is the time to peak. Note that here "0.5" replaces "1" in the accordant formula in Li et al. (2021), considering the temporal resolution of 0.5 h used in this study. ***O*** is the ratio of peak precipitation rate (r_p, *mm/h*) to the mean event precipitation rate ***R***:

$$O = \frac{r_p}{R} \tag{2}$$

In this work, we analyze the *number of events* (***N***, *events/year or events/season*) and the statistical distributions of the above-mentioned event properties and their spatial and seasonal variability, as compared between GV-MRMS and IMERG.

2.2.3. Event detectability

Even if IMERG has the same number of events with GV-MRMS, it would not be guaranteed that every IMERG-derived event temporally matches with a reference event. To assess this, we introduce the metrics of event detectability, which describe the hit, false, and miss events. Based on the temporal overlap (or non-overlap), and eventually on the temporal delays between the events across the two datasets, we classify all the events into three categories, which are diagramed in Fig. 1. The first category is the "*hit event*", which represents an event in a dataset that overlaps with at least one time step (here 0.5 h) with an event in the

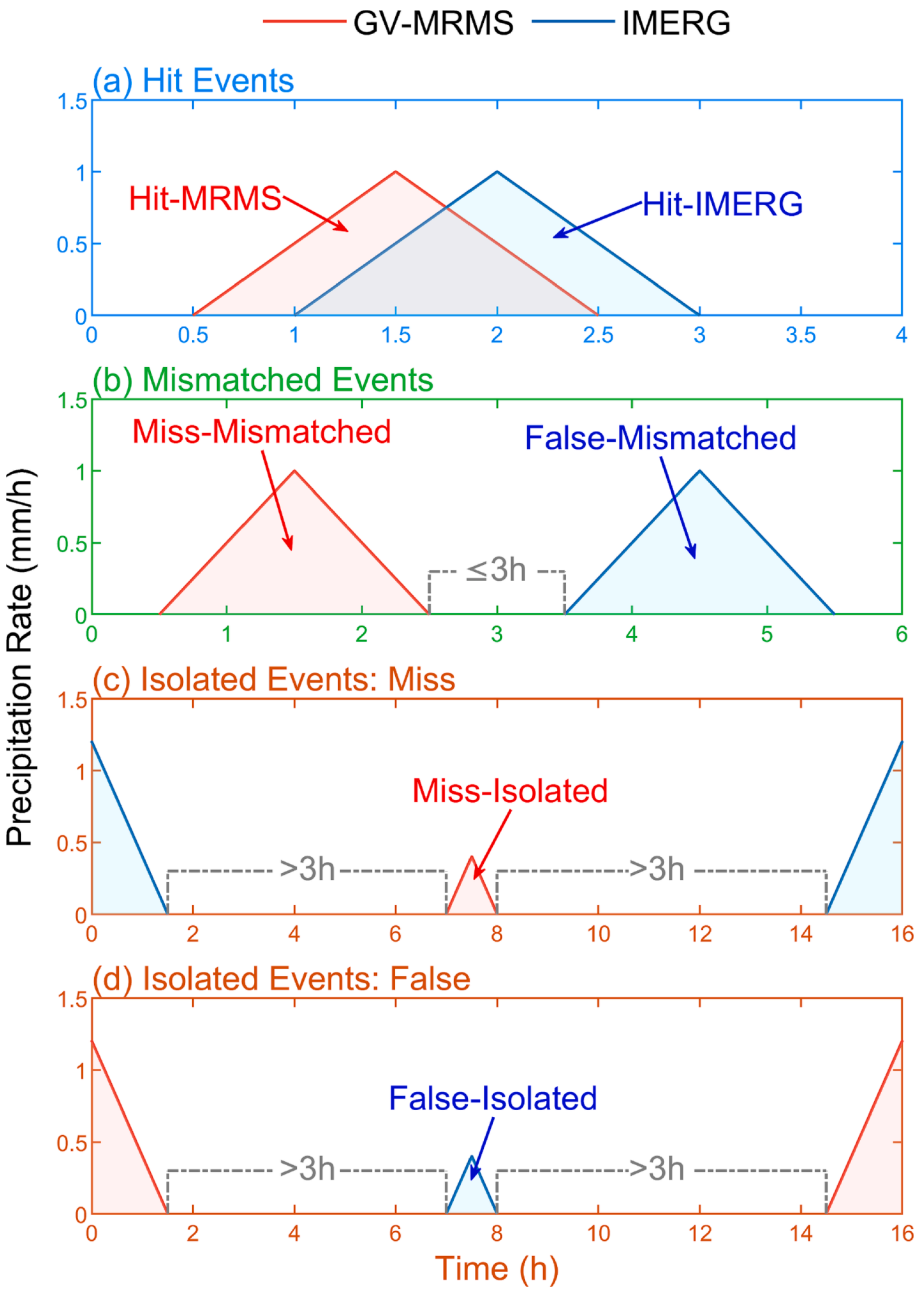


Fig. 1. Diagram of the three categories of events from IMERG and GV-MRMS based on their relative temporal positions. (a) *hit events* (overlap with an event in the other dataset with at least one time step), (b) the *mismatched events* (do not overlap with any event in the other dataset, but there is at least one event in the other dataset occurring within 3 h of their start or end times), and (c-d) the *isolated events* (do not overlap with any event in the other dataset, and there is also no event in the other dataset occurring within 3 h of their start or end times).

other dataset. This situation indicates that IMERG successfully detects a reference event, and the overlapped events in GV-MRMS and IMERG are named as *Hit-MRMS event* and *Hit-IMERG event*, respectively. The second category is the "*mismatched event*", which is defined as an event in a dataset that does not overlap with any event in the other dataset, but there is at least one event in the other dataset occurring within 3 h of its start or end time. This situation indicates that IMERG actually senses the existence of an event but narrowly misses it, and the two adjacent but non-overlapping events in GV-MRMS and IMERG are termed as "*Miss-mismatched event*" and "*False-mismatched event*", respectively. The third category is the "*isolated event*", which represents an event in a dataset that does not overlap with any event in the other dataset, and there is also no event in the other dataset occurring within 3 h of its start or end time. This situation indicates that IMERG is completely unaware of the existence of an event or makes up an utterly nonexistent event. Accordingly, such events are named respectively as "*Miss-isolated events*" and "*False-isolated events*". The "3-hour" threshold is chosen because we consider that timing errors (due in particular to the vertical variability of the precipitation process and to the dynamical interpolation based on motion vectors in IMERG) are unlikely to be more than 3 h in magnitude (Guilloteau et al., 2018; Utsumi et al., 2019; You et al., 2019).

We apply the traditional detectability error metrics, namely *Probability of Detection* (**POD**) and *False Alarm Ratio* (**FAR**) to the event detectability of the satellite product (Ebert et al., 2007), and define *Event Probability of Detection* (**EPOD**) and *Event False Alarm Ratio* (**EFAR**). Specifically,

$$\mathbf{EPOD} = \frac{N_{hM}}{N_{hM} + N_{mm} + N_{mi}} \tag{3}$$

$$\mathbf{EFAR} = \frac{N_{fm} + N_{fi}}{N_{hI} + N_{fm} + N_{fi}} \tag{4}$$

where N_{hM}, N_{mm}, and N_{mi} are the number of *hit-MRMS*, *miss-mismatched*, and *miss-isolated events* from GV-MRMS, respectively, and N_{hI}, N_{fm}, and N_{fi} are the numbers of *hit-IMERG*, *false-mismatched*, and *false-isolated events* from IMERG, respectively.

2.2.4. Timing of the detected events

Moreover, even if IMERG correctly detects an event present in the GV-MRMS dataset, this pair of *hit events* is still likely not to match perfectly in terms of the starting time, ending time and peak-intensity time. To assess this, we take the *hit-MRMS* events as the reference, and investigate the *start time shift* (dt_s), *peak time shift* (dt_p), and *end time shift* (dt_e) of the corresponding *hit-IMERG* events.

3. Results

3.1. Event properties

We analyze the statistical distribution of event properties in each

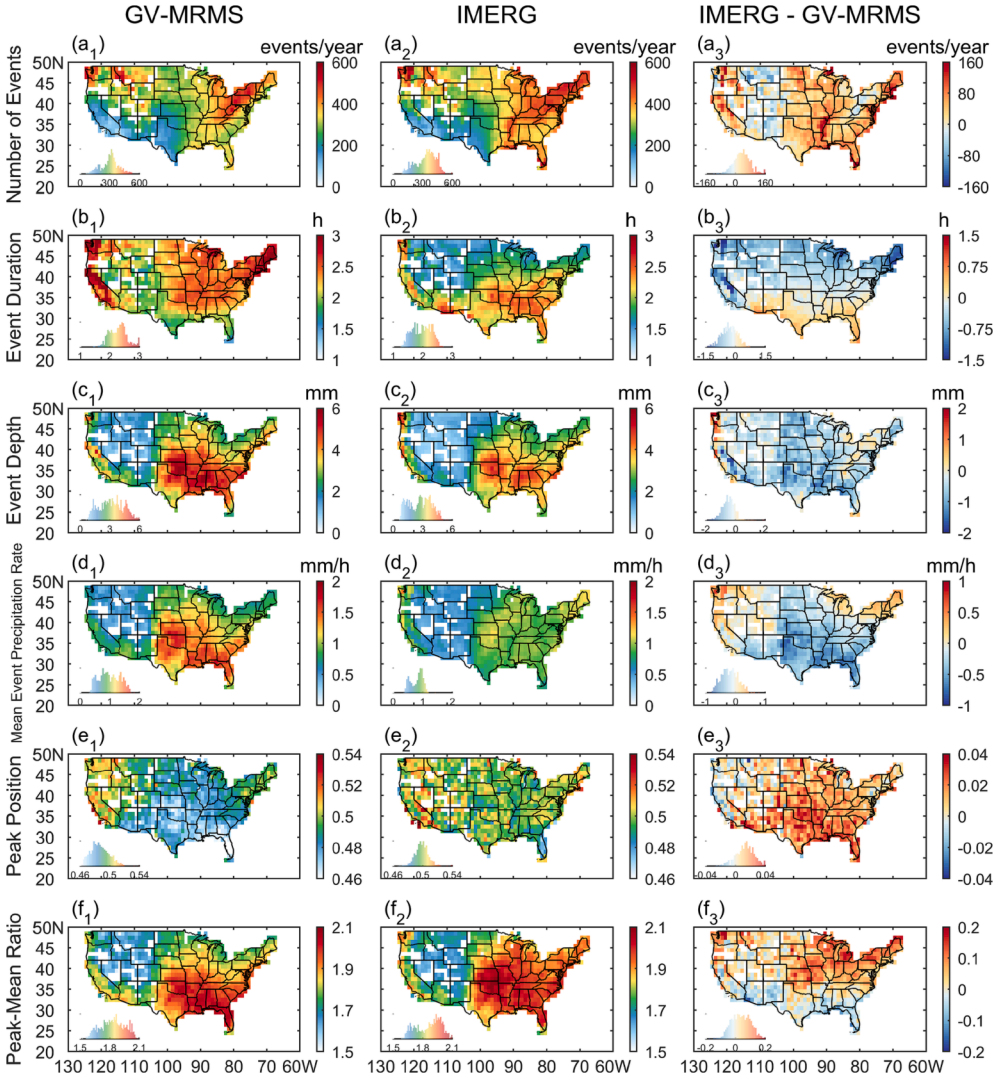


Fig. 2. The spatial patterns of (a_{1-3}) *number of events (events/year)*, as well as the mean values of (b_{1-3}) *event duration (h)*, (c_{1-3}) *event depth (mm)*, (d_{1-3}) *mean event precipitation rate (mm/h)*, (e_{1-3}) *relative peak position (time to peak divided by duration)*, and (f_{1-3}) *peak-mean ratio (peak intensity divided by mean event precipitation rate)* during the study period of 2018–2020 from GV-MRMS and IMERG data and their differences (IMERG - GV-MRMS) for 1° × 1° gridboxes in the U.S. The histogram of each variable over CONUS is shown in the left bottom of each panel.

pixel for GV-MRMS and IMERG. Fig. 2 shows maps of the number of events per year along with maps of the mean value for each one of the event properties. To reduce sampling noise, the metrics are exhibited as the average values over 1° × 1° gridboxes. From this map, regional differences in the event properties between IMERG and GV-MRMS are salient. In terms of the number of events, event depth, and peak-mean-ratio, a spatial consistency between GV-MRMS and IMERG is observed, with cross correlation coefficients of 0.88, 0.93, and 0.87, respectively (Figs. 2a1-2, d1-2, and f1-2), although there are still biases in their quantitative values (Fig. 2a3, d3, and f3). For example, IMERG overestimates the number of events in most regions of CONUS where the mean event duration revealed by GV-MRMS is also relatively high (Fig. 2a3 and b1). This may be due to IMERG intermittently missing precipitation during the events, which causes long-duration events to be split into several parts. The exception is over the Rocky Mountains, where the lower number of events in IMERG as compared to GV-MRMS might be related to the documented low detection rate of satellites for orography-induced rainfall and over arid/semiarid areas with hot background surfaces (Derin and Kirstetter, 2022; Derin and Yilmaz, 2014; Dinku et al., 2010; Maggioni et al., 2016). In contrast to the duration, the mean event depth showed a nationwide underestimation by IMERG (Fig. 2d3), which might also be owing to the event split mentioned above. Considering that the average event depth multiplied by the number of events in a year is the total annual precipitation, which is constrained by the GPCC annual values (Huffman et al. 2019a), an overestimated number of events is almost certainly accompanied by an underestimated event depth as a compensation.

The spatial distribution of the mean event duration of IMERG is significantly different compared to that of GV-MRMS (Fig. 2b1-3). For GV-MRMS, events last longer in all the Midwest and the Northeast U.S. but are shorter in the Rocky Mountain, Gulf Coast, and Florida (Fig. 2b1), while the duration of IMERG-derived events inversely increases from the North to the South (Fig. 2b2). Attributing a cause to such a difference in the patterns is not easy when simply examining these two distribution maps. We note, however, that the bias shows an evident regularity, with an apparent meridional gradient pattern from negative in the North to positive in the South (Fig. 2b3). This pattern might be related to the gradually reduced frequency of PMW overpasses as the latitude decreases, eventually leading to the increased influence of the morphing scheme in IMERG (Tan et al., 2021), which tends to increase the event duration (Ayat et al., 2021). For the relative peak position, the peaks of the reference events are skewed to the early part of the event and have a clearer regional variation than IMERG-derived events, while the relative peak position of the IMERG-derived events is closer to the event center with less regional differences (Fig. 2e1-2). Such bias may arise from the interpolation procedures in IMERG that weaken the asymmetry (Li et al., 2021). For the mean event precipitation rate, IMERG generally reproduces the spatial pattern shown by GV-MRMS with higher values in the Southeast U.S. compared to other regions. However, such contrast between different regions is not as strong as that in GV-MRMS (Fig. 2d1-2), leading to a clearly underestimated mean event precipitation rate in the Southeast U.S. (Fig. 2d3). Since the mean event precipitation rate is the result of event depth divided by event duration, such a pattern is explainable considering the overestimated duration and underestimated depth in the South U.S. (Fig. 2b3). As a result, IMERG tends to have more long-duration, light to moderate events in the South U.S., more short-duration, heavy events in the Northeast U.S., Rocky Mountain, and West Coast, but more short-duration, light to moderate events in the Great Plain.

Since different types of precipitation systems dominate in different seasons, further analysis by season is necessary for bias attribution (Fig. 3 and Fig. S1). For example, the overestimation of event duration is most severe in winter when the long-duration frontal precipitation dominates (Fig. 3b_4 and Fig. S1b_7). During the cold season, the interpretation of the passive microwave signal is more ambiguous as frozen or snow-covered land surfaces can have spectral signatures similar to those of ice clouds (Ferraro et al., 1998). To counterbalance this, during the cold season, the IMERG algorithm relies more on the IR information (Huffman et al., 2020), which is generally less accurate in terms of precipitation detection (Derin et al., 2021; Tan et al., 2016). Such deficiency tends to split the long-lasting events, causing the overestimated number of events but underestimated duration (Fig. 3a4 and b4). The missed precipitation, which is more likely to occur for low-intensity precipitation, further elevates the mean event precipitation rate (Fig. 3d4).

In contrast, the short duration but still overestimated number of events in the summer indicates the existence of false events in IMERG (Fig. 3a_2, Fig. S1b_3). Such phenomenon could also be owing to the fact that the deep convective precipitation, predominant in the summer, is reasonably well detected by passive space-borne instruments (Chen et al., 2011; Young et al., 2014). Exceptionally, the more pronounced underestimation of the number of events around the Rocky Mountains, relative to the annual counterpart (Fig. 2a3), can presumably be attributed to the elevated air temperature favoring the generation of warm rain, coupled uytwith the increased surface temperature neutralizing the cold signals of hydrometeors (Derin and Kirstetter, 2022; Dinku et al., 2010; Maggioni et al., 2016). Indeed, intermittent satellite observations missing the short-duration precipitation prevalent in the summer could be another important contributor. Due to the interpolation procedures in IMERG, a PMW estimate, usually with a high precipitation rate in the summer, is likely to fill their adjacent dry gaps with non-zero values, thus prolonging the duration but consequently reducing the mean precipitation rate (Fig. 3b_2, d_2, and S1d_3). Besides, the relative peak position has a delaying bias, not least in summer when afternoon convective precipitation dominates (Fig. 3e_2 and S2e_3). This type of precipitation generally has an early peak since the evaporative cooling suppresses the surface heating shortly after the rainfall starts (Dunkerley, 2021), but this feature fails to be reproduced by IMERG due to the weakening of rainfall asymmetry induced by the interpolation procedures in the merging algorithms.

To go further than analyzing the mean values of the different event properties, Fig. 4 depicts the joint distribution of the annual and seasonal total precipitation amount contributed by events with different durations and mean event precipitation rates, from which we can see more detail about the event-related performance of the satellites. As shown, IMERG and GV-MRMS display dissimilar patterns at all times. Specifically, the distributions for GV-MRMS are more dispersed, while those for IMERG are more concentrated around the diagonals (Fig. 4a1-a5 and b1-b5). In other words, IMERG underestimates the occurrence of short-duration, high-intensity events and long-duration, low-intensity precipitation events but overestimates the occurrence of the medium-duration, moderate-intensity events as compensation (Fig. 4c1-c5). Such bias is evident in the summer and winter (Fig. 4a3-c3 and a5-c5), when brief strong thunderstorms and long-lasting, steady frontal precipitation are known to occur, respectively. As analyzed above, the smoothing effect of the interpolation leads to IMERG's underestimation of the occurrence of short-duration heavy events, while the IR-only sources with high missing rates that frequently occur in late autumn, winter, and early spring make IMERG hard to reflect long-lasting events. Especially for the high-impact, short-duration extreme events that might cause flash floods and debris flows, their underrepresentation could largely affect the prediction or retrospective attribution of these natural hazards when employing the IMERG data.

3.2. *Event detectability*

After analyzing the difference in the mean event properties, events from GV-MRMS and IMERG are further matched in time to investigate whether IMERG can successfully capture the reference events depicted by GV-MRMS. As stated in the Methods section, events in GV-MRMS and IMERG are classified into three categories: *hit events*, *mismatched events*, and *isolated events*, whose proportions are displayed by their durations

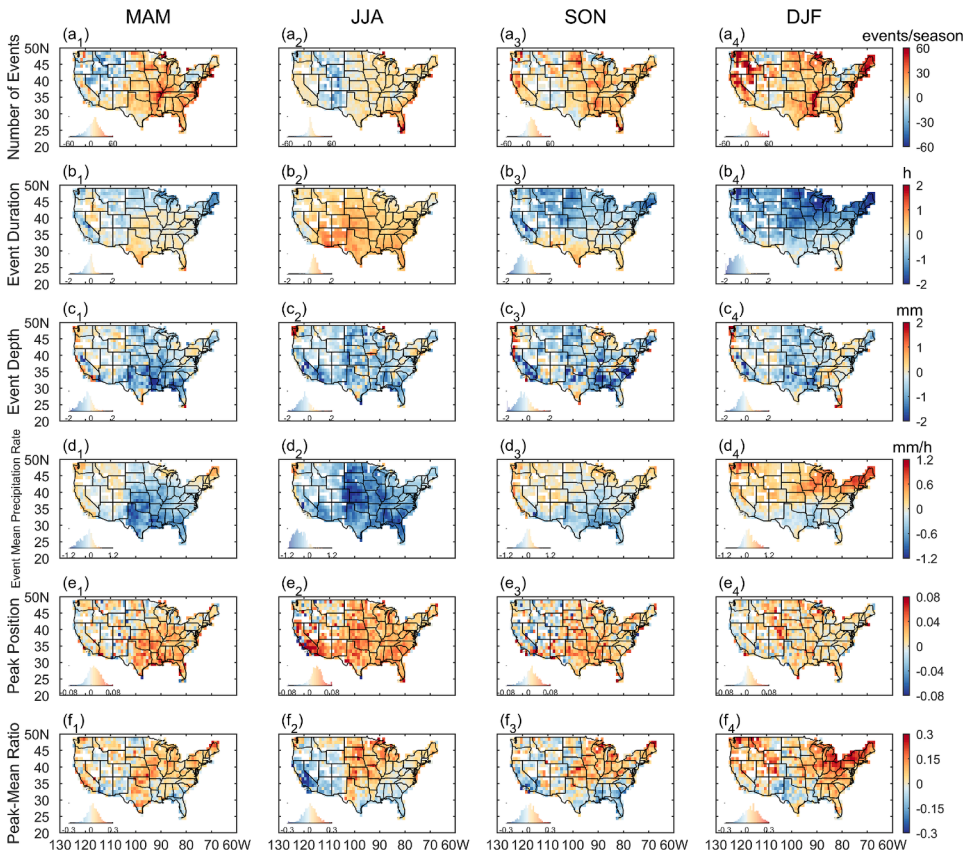


Fig. 3. The spatial patterns of the differences (IMERG - GV-MRMS) of the metrics listed in Fig. 2 for four seasons for 1° × 1° gridboxes in the U.S.

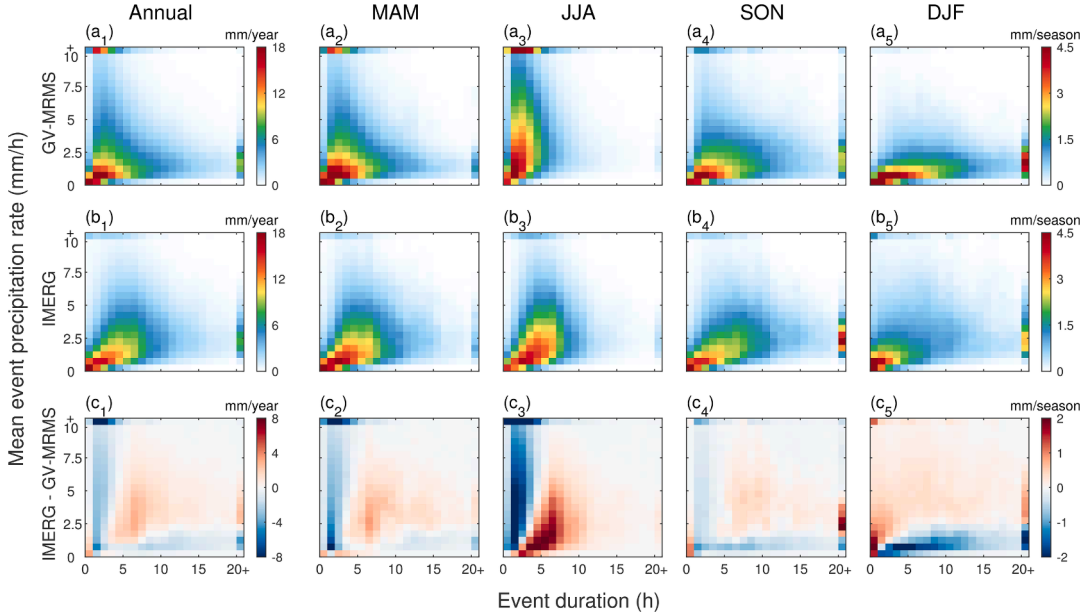


Fig. 4. Joint distribution of the national mean annual and mean seasonal precipitation amount contributed by events with different durations and mean event precipitation rates from ($a_{1\text{-}5}$) GV-MRMS and ($b_{1\text{-}5}$) IMERG data and ($c_{1\text{-}5}$) their difference (IMERG - GV-MRMS) for the full year and for each of the four seasons. Precipitation events with durations of more than 20 h or mean rates of more than 10 mm/h are counted together and marked with a "+".

(Fig. 5). For the events with all durations, the two datasets possess a similar composition, with the *hit events*, *mismatched events*, and *isolated events* making up 49%, 20%, and 31% in GV-MRMS and 45%, 19%, and 36% in IMERG. The proportion of hit events in each dataset gradually increases with the lengthening of event duration, from below 25% to above 95%. Such a phenomenon is understandable since a longer-duration event has a higher chance of being overlapped with any other event in the other dataset.

We further apply the *Event Probability of Detection* (EPOD) and *Event False Alarm Ratio* (EFAR) to quantify IMERG's detectability of

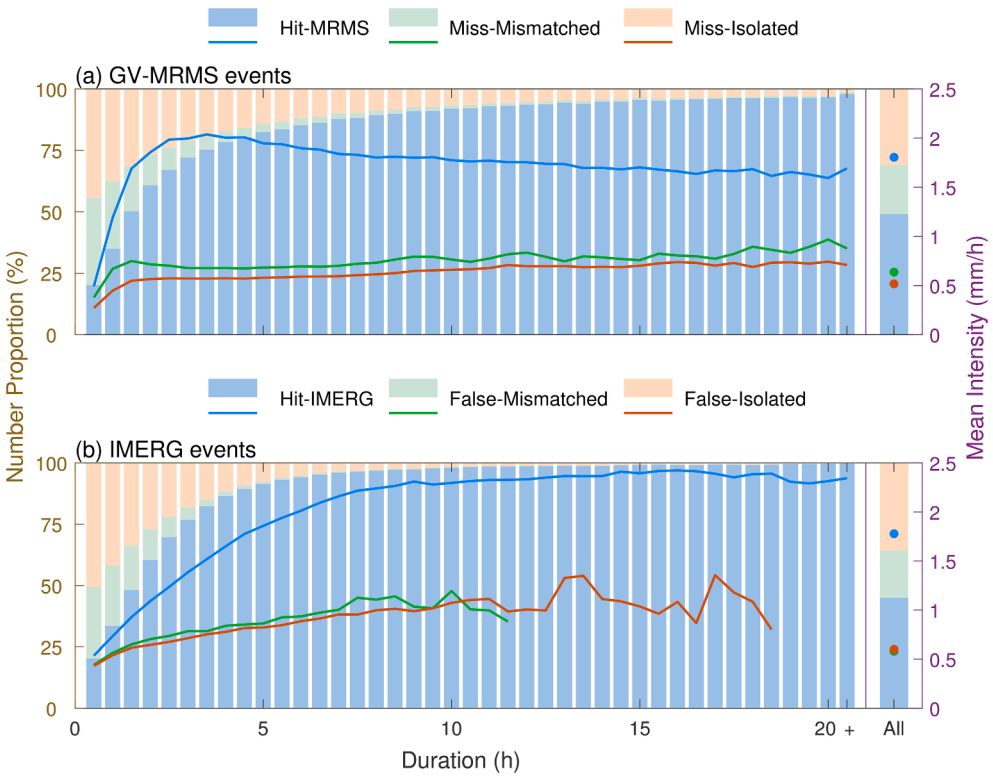


Fig. 5. Proportions of number of events (bars, left y-axis) and mean half-hourly intensity (lines, right y-axis) of the *hit events*, the *mismatched events*, and the *isolated events* diagramed in Fig. 1 from (a) GV-MRMS and (b) IMERG by their durations. The proportions are calculated from all the data during the study period of 2018–2020 over CONUS. The label "+" denotes durations above 20 h and "All" means all durations. The intensities are displayed only if the number of events in the sample is above 50.

precipitation events (Fig. 6). Overall, IMERG does not have a high level of detectability. The mean EPOD equals 0.49 and EFAR equals 0.55 across the CONUS and at the annual time scale (Fig. 6a1 and b1), which might be largely contributed by the misdetection of short-lived events (lasting 0.5–3 h) as analyzed above (Fig. 5). The EPOD and EFAR show differential spatial patterns compared to the bias of the annual and seasonal average number of events (Figs. 6 2a, and 3a1-a4), which indicates that the similarity in the number of events between IMERG and GV-MRMS does not necessarily demonstrate IMERG's good event detectability, further bearing out the necessity of the comparison based on event matching. For example, although IMERG-derived events have a consistent overestimation in the number of events in both the Northeast and Southeast U.S. at the annual time scale (Fig. 2a3), the Southeast exhibits a relatively high EPOD and low EFAR while the opposite is true for the Northeast (Fig. 6a1 and b1). More missed and false events in the Northeast at the same time cancel out each other and lead to a comparable performance in the number of events compared with that in the Southeast. Additionally, the summer exhibits the best detectability among all seasons, with EPOD over 0.7 and EFAR below 0.3 in some regions (Fig. 6a3 and b3); while the winter performs worst, and the EPOD and EFAR could be below 0.2 and above 0.8, respectively (Fig. 6a5 and b5).

As analyzed above, the *miss-isolated* and *false-isolated* events not only comprise a significant proportion of the bias but are also not easy to address in the retrieval, thus worthy of further investigation. Fig. 7 looks in more detail into the characteristics of these events, involving the contributions to the total number of events, total amount, and total frequency, as well as their mean duration and intensity. Overall, the regions with severe miss- or false-isolated events are still in the Western U.S., where arid/semiarid and mountainous areas characterized by generally poor satellite performance dominate (Derin and Kirstetter, 2022; Dinku et al., 2010; Maggioni et al., 2016). The isolated events account for a large proportion of events both in GV-MRMS and IMERG, with their national mean values equaling 31% and 36%, respectively (Fig. 7a1 and a2). However, their contributions to the total frequency are smaller (19% for GV-MRMS and 18% for IMERG nationally, Fig. 7b1 and b2). Compared with the miss-isolated events, those false-isolated events are marked by a larger number but shorter durations (Fig. 7a1-a2 and d1-d2). Although their contribution to the total amount during the study period is even less (7% for both GV-MRMS and IMERG nationally) than to the total frequency, the percentages in the total amount could still exceed 20% in the Mountain West areas (Fig. 7c1 and c2). In terms of intensity, the miss-isolated events are comparatively high in the South U.S. but progressively decrease with higher latitude, while the false-isolated events are higher, especially in the Northeast and Northwest U.S. The fundamental causes for these biases need to be further investigated for possible improvements in the retrieval algorithms.

3.3. Timing of the detected events

Even if IMERG successfully detects a precipitation event, it does not warrant that IMERG can exactly reproduce the whole event in terms of its timing, duration, and depth. That is, the retrieved events may mistime the beginning and end, preventing them from wholly matching the reference events. Fig. 8 displays the matching percentage of the hit events from IMERG and GV-MRMS in terms of duration and depth as a function of event duration. Results show that the pairs of hit events from the two datasets can only match about 70% of the total durations on average, but the percentages can rise up to about 80% on average for the total depth due to the relatively small intensity of the unmatched rainy hours. Besides, it is easily explainable that the matching percentage of the 0.5 h-events always equals 100%, and the values naturally drop with the longer duration for the brief events (<3h). However, the percentages of GV-MRMS-derived events keep decreasing (e.g., from 70% to 60% for the matching percentages of event duration), but those of IMERG-derived events inversely increase with longer duration (e.g., from 60% to 80% for the matching percentages of event duration). Such a phenomenon indicates that the hit-IMERG events tend to have fewer false time steps but more miss time steps with increasing event duration. This further implies IMERG's more significant underestimation of the event duration for longer "actual" events. These results are consistent with those revealed in Figs. 1-3.

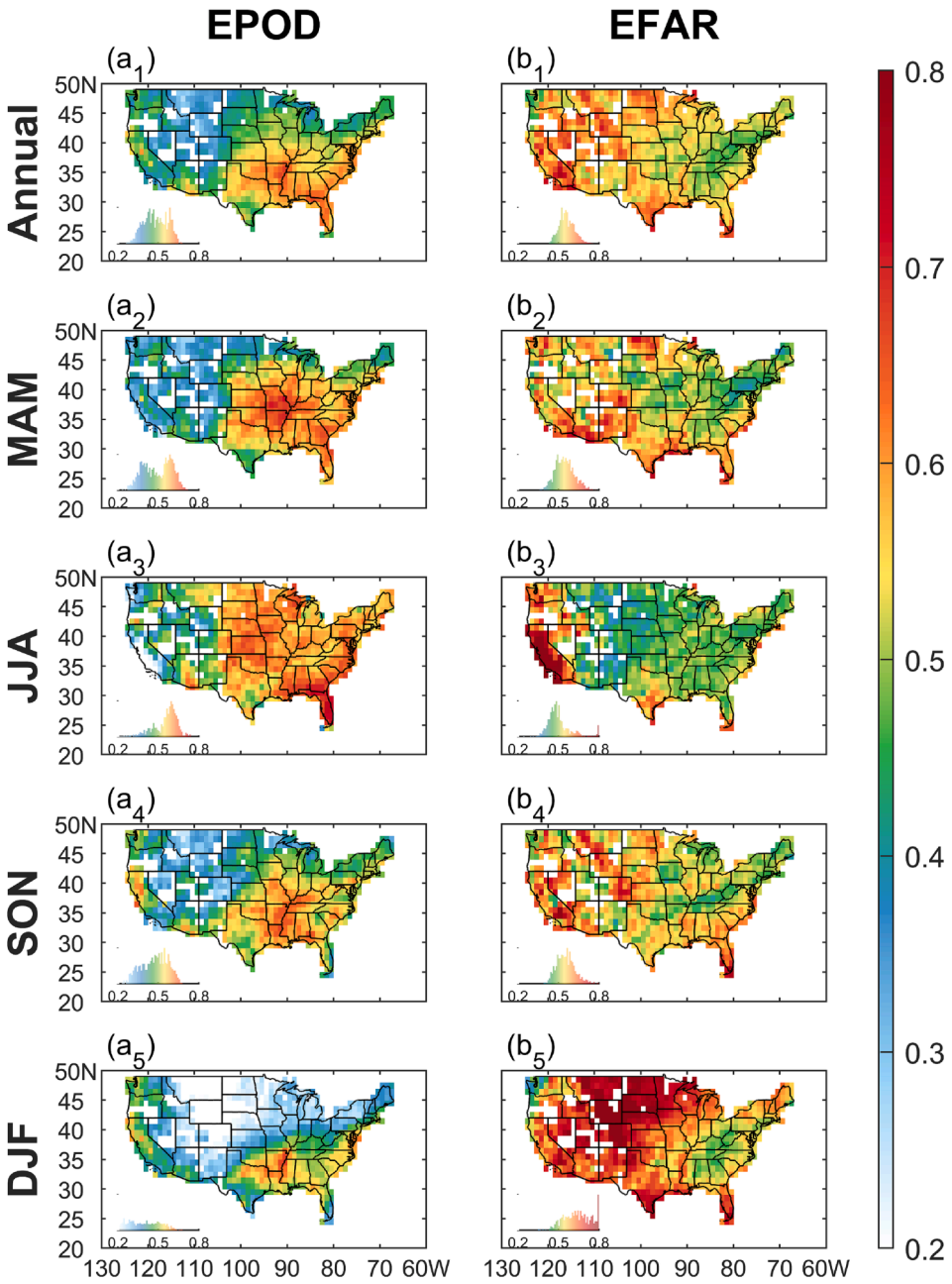


Fig. 6. The spatial patterns of *Event Probability of Detection* (**EPOD**) and *Event False Alarm Ratio* (**EFAR**) for (a_1 and b_1) the full year and (a_2-a_5 and b_2-b_5) four seasons.

Fig. 9 further looks into the start, peak, and end time shifts of the hit-IMERG events compared with their corresponding hit-MRMS events in the form of the Probability Density Functions (PDFs). From the graphs, all three PDFs are negatively skewed, which means that IMERG tends to start, peak, and end all earlier than the reference events. Among the three event properties, the time shift is most severe for the start time, with an average of −0.44 h (26 min), which is followed by peak time and end time with averages of −0.29 h (17 min) and −0.12 h (7 min). Such a difference is also reflected in the cumulative probabilities, with the contrast between the percent of total events with advances (time shift $\geq$ 0.5 h) and delays (time shift $\leq$ −0.5 h) at 57% versus 19%, respectively, for the start time but 44% versus 38% for the end time. Events with unbiased start, peak, and end times (time shifts of 0 h) only account for about 20% in each PDF, and the chance of a −0.5 h shift even exceeds that of no shift for the peak and end times (Fig. 9b and c). Such results further demonstrate the systematic time shifts of the IMERG-derived events.

Fig. 10 further examines the consistency of the revealed mistiming issue of IMERG-derived events over CONUS in the form of mean time shifts. Among the three properties (start, peak, and end time), the peak time shift is most consistent with an advanced shift among almost all the regions and seasons (Fig. 10a2-e2). IMERG-derived events have advanced start times in most areas for spring, summer, and winter (Fig. 10b1-d3), while the time shifts in the event ending time could be both positive and negative for the same periods (Fig. 10b3-d3). Such facts align with the more negatively skewed PDF for the start time than the end time (Fig. 9a and c), further indicating the satellites' asymmetric performance during the events (Li et al. 2021). However, significantly different patterns are shown in the winter for both the start and end time shifts (Fig. 10e1 and e3). While the moderate shifts displayed in the South U.S. are in agreement with those in the other seasons, the North suffers from substantially delayed start times and advanced end times at the same time (Fig. 10e1 and e3). The mean delays in the North in the winter could exceed 1 h while the advances could be even more severe

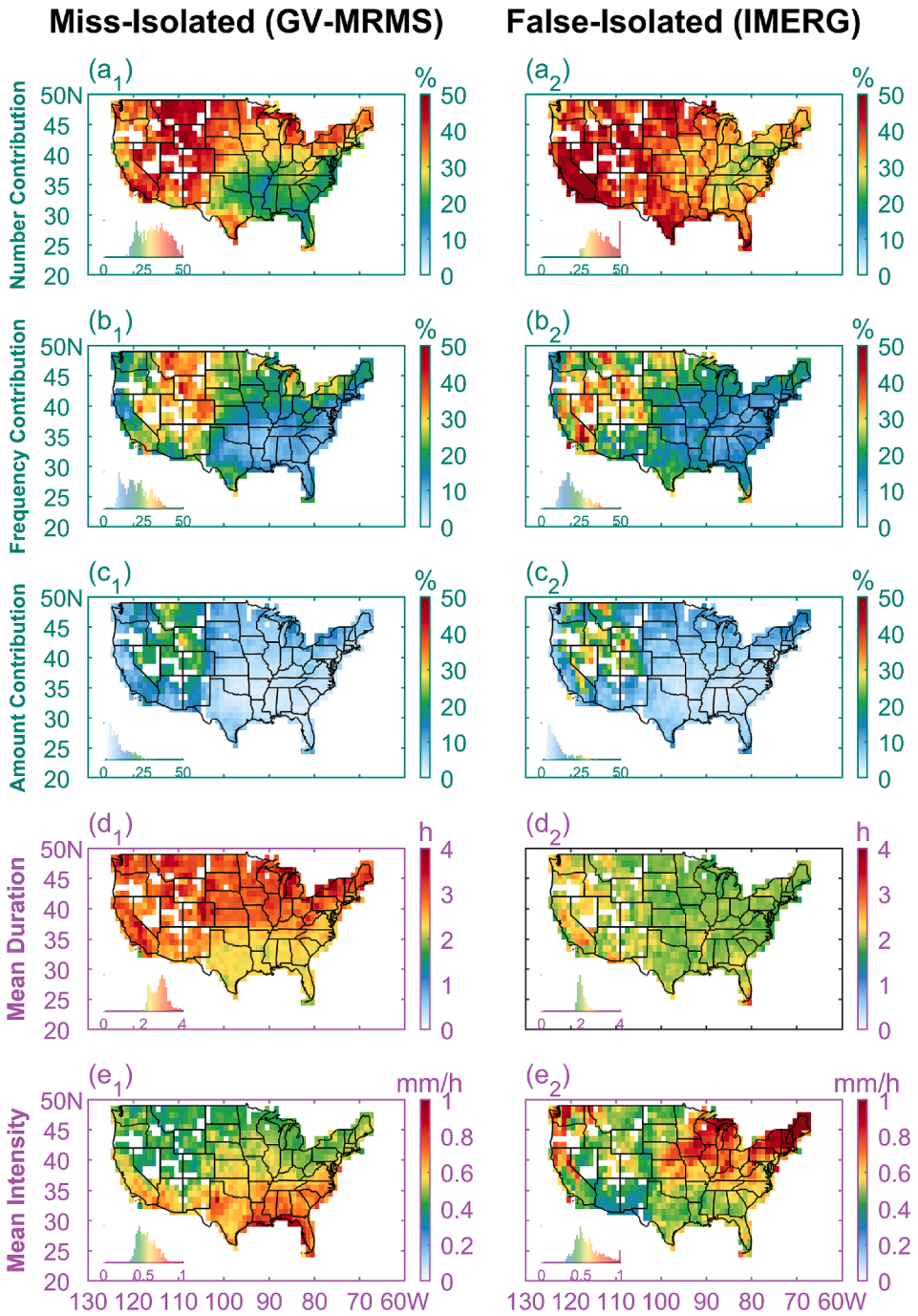


Fig. 7. The spatial patterns of the contributions of the *miss-isolated* and *false-isolated* events to the total (a_1–a_2) number, (b_1–b_2) frequency, (c_1–c_2) amount during the study period, as well as their (d_1–d_2) mean duration and (e_1–e_2) mean intensity from GV-MRMS and IMERG during the study period of 2018–2020 for 1° × 1° gridboxes in the U.S.

and surpass 2 h, which may well be due to the inaccurate IR sources in the retrieval process, as discussed before.

Next, we further examine the time shifts of IMERG when detecting the events of different durations. As shown in Fig. 11, the curves of the start, peak, and end time shifts all exhibit smooth and monotonic variations with increasing duration, which indicates the significant dependency of the mistiming on the event duration. Specifically, IMERG-derived events tend to start earlier for short-duration events but later for long-duration events, while they end later for short-duration events but earlier for long-duration events. However, the tipping point of the start time shift is "8h" while that of the end time shift is "2h", which suggests a negatively skewed tendency of the timing of IMERG-derived events in accordance with the results in Fig. 9. Such patterns indicate that the causes of mistiming might involve both symmetric and asymmetric factors that contribute to the biases in the retrieval. The curve of the peak time shifts further confirms this inference, as its values keep being negative and decreasing with longer durations, but the shape of the curve highly dovetails with that of the mean value of the start and end time shifts. This means that the start and end time shifts of IMERG-derived events might be attributed to a symmetric extension or contraction centered around the event peak time, as well as an overall systematic advance of the events. The former might result from the inherent deficiencies of the data sources, e.g., the lengthening effects of "morphing" sources and the high missing rate of IR sources which

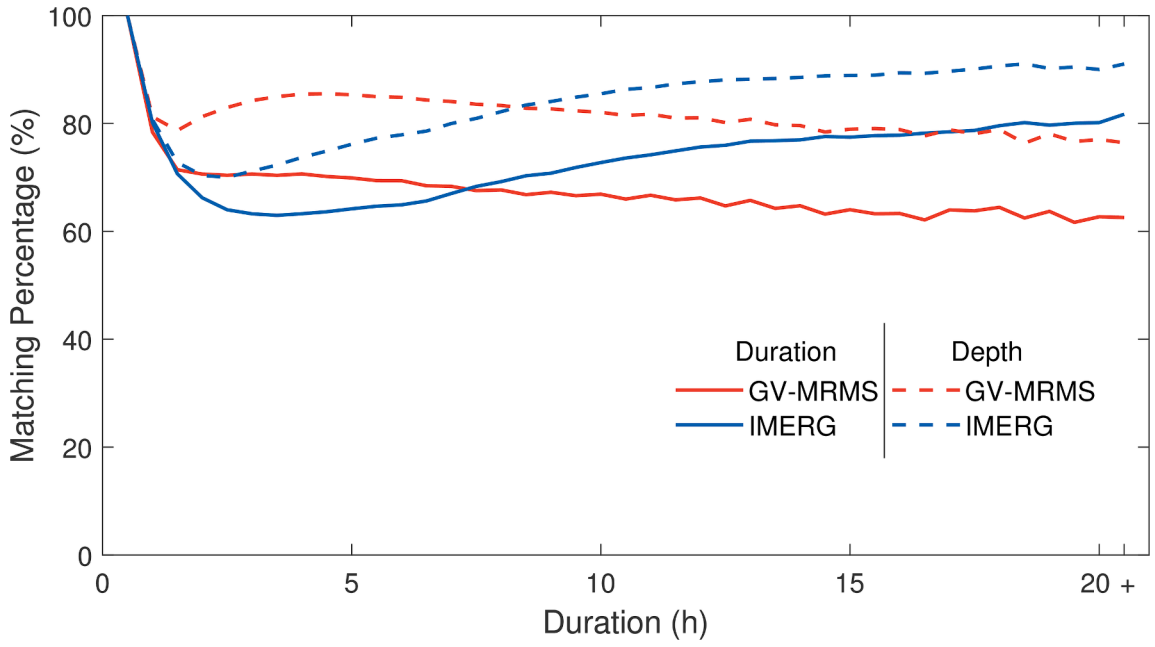


Fig. 8. Matching percentages, i.e., the proportion of the overlapping parts in the whole event duration and depth of the hit events from GV-MRMS and IMERG as a function of event duration. The percentages are calculated from all the data during the study period of 2018–2020 over CONUS.

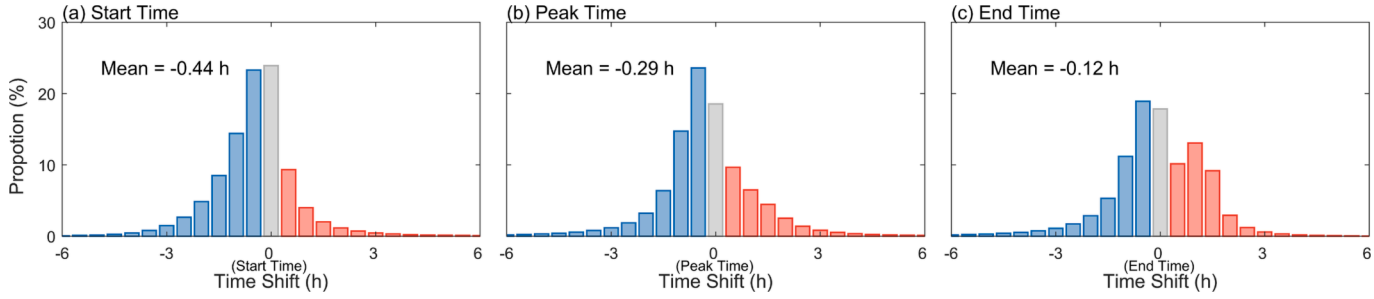


Fig. 9. Probability density function of the (a) start, (b) peak, and (c) end time shifts of the *hit-IMERG* events compared with the corresponding *hit-MRMS* events (IMERG-MRMS). The PDFs are calculated from all the data during the study period of 2018–2020 over CONUS. The negatives indicate advances in timings (IMERG starts, peaks, or ends earlier than GV-MRMS) while the positives indicate delays.

generally exert a comparable impact on the event start and end times (Tan et al., 2021). The latter might be caused by a consistent time difference between the satellite receiving the ice scattering signal from the sky and the precipitation falling on the ground due to instrumental effects, such as the incidence angle (Guilloteau et al., 2018). The actual situation might be more complex, also influenced by the different retrieval capabilities in response to the inherent asymmetry of the cloud physical properties during the storm evolution (Imaoka and Nakamura, 2012).

4. Discussion

Compared with the traditional evaluation methods of satellite precipitation products that treat precipitation values at each time step as independent and calculate the mean error statistics (e.g., Mean Absolute Error, Root Mean Square Error) (Amjad et al., 2020; Beck et al., 2019; Beck et al., 2017), our study examines the inherent temporal information of precipitation, that is, properties of individual events. The event-based evaluation is expected to provide additional insight into the use of satellite products for hydrological applications and also lead to targeted improvements in the retrieval algorithms. Moreover, in contrast to the previous few event-based evaluations (Freitas et al., 2020; Li et al., 2021), we go one step further to investigate the mismatching and mistiming issues of satellite-derived events, in addition to the comparison of the mean event properties. Several insights are gained through these analyses as discussed below.

For example, Fig. 11 shows an advanced start time and a slightly delayed peak time for the short-duration events but a delayed start time and a more strongly delayed peak time for the longer-duration events in IMERG compared with those in GV-MRMS. This indicates that the time from the start to the peak is overestimated in IMERG for short-duration events but underestimated for long-duration events. Previous studies have found that an event with a later peak could generate substantially more runoff than an event with an earlier peak (Dunkerley, 2012; Dunkerley, 2017), mostly because the later-peak events tend to wet the dry soil before the arrival of the most intense precipitation, thus reducing infiltration and generating more surface runoff (Dunkerley, 2021). Therefore, such systematic bias could influence the results of hydrological models when using IMERG as the input (Wu et al., 2012; Wu et al., 2014). Added to that, although the comparisons of the mean annual event properties calculated from the events of IMERG and the reference data separately in this study and in a previous studies indicate a later relative peak position of the satellites (Fig. 2e1-e3) (Li et al., 2021), further results from matching the events from the two datasets show that the absolute peak times of IMERG-derived events tend to be earlier than the real events (Fig. 9b). This result highlights the

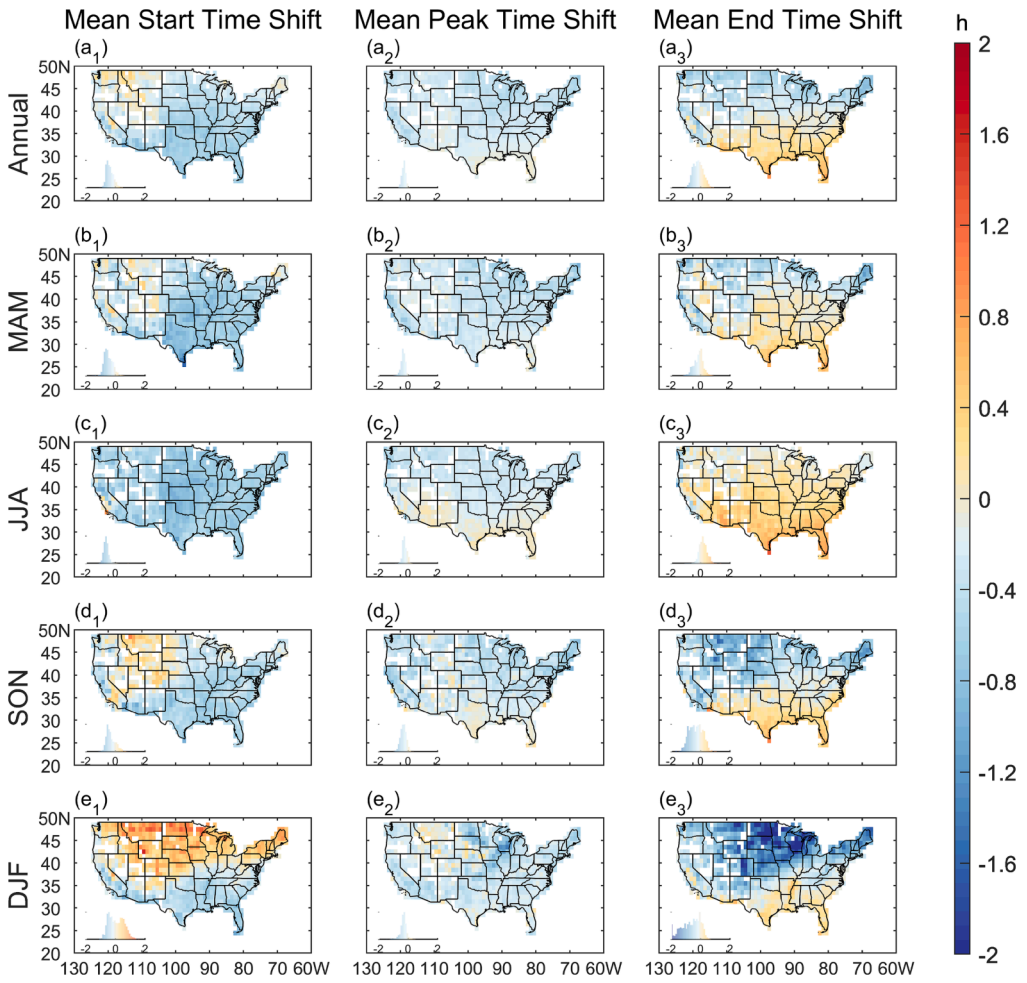


Fig. 10. The spatial patterns of the mean event (a_1-e_1) start, (a_2-e_2) peak, and (a_3-e_3) end time shift of the hit-IMERG events compared with the corresponding *hit-MRMS* events (IMERG - GV-MRMS) for the full year and for the four seasons for 1° × 1° gridboxes in the U.S. The negatives indicate the advances in timings (IMERG starts, peaks, or ends earlier than GV-MRMS) while the positives indicate delays.

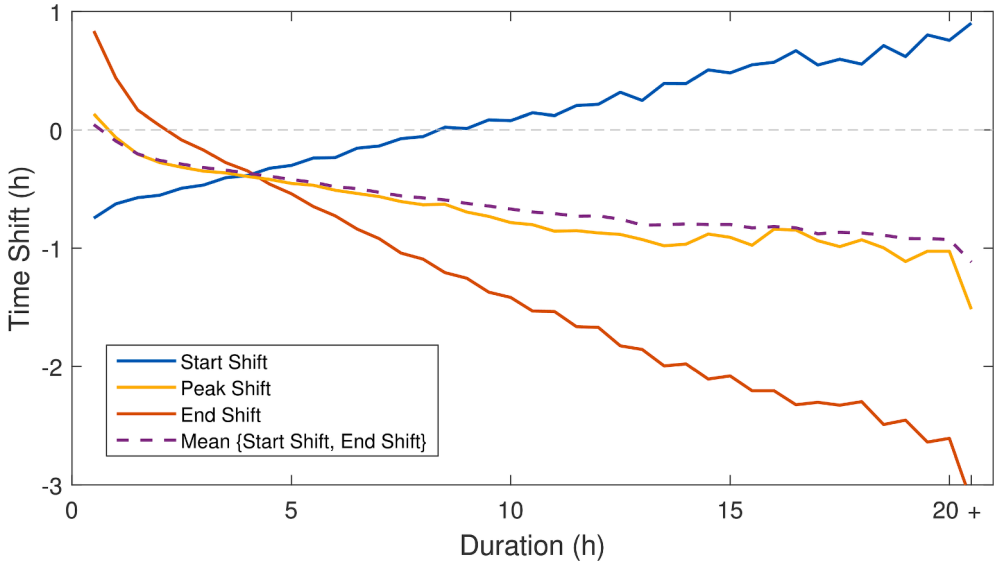


Fig. 11. The mean event start, peak, and end time shift of the *hit-IMERG* events compared with the corresponding *hit-MRMS* events (IMERG - GV-MRMS) as a function of the duration of *hit-MRMS* events. The values are calculated from all the data during the study period of 2018–2020 over CONUS. The negatives indicate the advances in timings (IMERG starts, peaks and ends earlier than GV-MRMS) while the positives indicate delays.

importance of the one-to-one comparison between the GV-MRMS- and IMERG-derived events, which has not been performed before (Li et al., 2021). In terms of intensity, IMERG underestimates the mean event precipitation rate of short, intense events, which is of high impact, notably in the summer (Fig. 4). An underestimated peak-to-mean ratio is simultaneously revealed for the summer short-duration events (Figure S2e$_3$). This characteristic could also influence the hydrological applications of IMERG, e.g., underestimating the surface influence of heavy flash events. Such findings could not be easily drawn from the traditional evaluations without considering the event-based characteristics.

Although our results show that the mean start and end time shifts are less than 0.5 h (Fig. 9), the PDFs tell us that there are only 20% of the start and end times of the real events that are accurately reflected by IMERG. In comparison, the time shifts with absolute values no less than 2 h could also account for over 20%, which is nonnegligible for fine-scale hydrological applications like flood predictions (Wu et al., 2012; Wu et al., 2014). Seasonal spatial distribution analysis further suggests that such bias could be worse in specific regions and seasons (Fig. 10). Moreover, from the error perspective, even if IMERG duplicates an event profile identical to the real case but has an overall 1-h advance, the intensities during the event could be biased in all time steps. On top of that, if a large proportion of events all have a "small shift", they could altogether contribute significantly to the total absolute error. This is the case in our results, where about 60% of the IMERG-derived events have an advanced start time (Fig. 9a). These findings imply that the mistiming issue of precipitation events detected in the satellite product should be given focused attention in future algorithm development.

As mentioned above, although the lengthening effect of the interpolation procedures in the merging algorithm on event duration could contribute to the start and end time shifts, they theoretically exert symmetric influences on both the start and end sides of the events. Therefore, other reasons should be responsible for the generally advanced event time shifts of IMERG. For example, the precipitation might fall on the ground tens of minutes after the ice-scattering signal aloft is captured by the satellites (Guilloteau et al., 2018), thus leading to the advanced timing of events. The possible mislocation of precipitation systems caused by the uncorrected parallax shifts of PMW and geostationary sensors is likely to cause a skewing in the event timing as well (Li et al., 2022). If these conjectures are further substantiated in future work, the mistiming issue might be substantially mitigated by additional adjustment procedures. Moreover, precipitating clouds could undergo substantial thermodynamical and microphysical changes in their lifetime, which might give rise to different levels of fidelity in the course of events (Petkovic and Kummerow, 2017). Future analysis should be made by combining the variation of cloud properties and environmental variables to further understand the asymmetric performance of satellite products during the events. On the other hand, the inhomogeneity of the sensor sources could bring about irregular event time shifts (Li et al., 2018), which needs to be further investigated in future work.

In addition, the bias of event duration shows significant latitude-dependent values (Fig. 3b1), which is suspected to be related to the data source composition in IMERG that is also latitude-dependent (Passive Microwave Algorithm Team Facility, 2017; Gebregiorgis et al., 2017). To verify this conjuncture, Fig. S4 compares the curves of zonal mean event duration bias and the morph source proportions as a function of latitude. The basically monotonically increasing trend of the two curves as the latitude decreases, as well as their high correlation coefficients ($>$0.85 for all situations), bears out their strong connection. The polar-orbit and low-orbit satellites revolve around the earth with the same swath width at different latitudes. This always leads to the undersampling of high-quality PMW observations at low latitudes (Passive Microwave Algorithm Team Facility, 2017). Complementarily, as shown, the proportion of the morph sources increases to fill the gap, spanning from 45% to 75% across CONUS (Fig. S4a).

The PMW sources are suggested to miss light precipitation, likely to occur both near the beginning and end of the events, thus shortening the event duration (Ayat et al., 2021). As the replacement of PMW in the ice/snow surfaces, IR sources with a high missing rate could exert a more substantial shortening effect on the duration (Tan et al., 2016). In contrast, the morph data inversely tend to prolong the event duration due to the high chance of assigning non-zero values to dry grids through the interpolation process (Li et al., 2022). At the same time, the lengthening and shortening of the duration could result in a decrease and increase of the mean event precipitation rate, respectively, which is also reflected in our results, although the intensity bias is affected by more factors (Fig. 2d3 and 2d1-d4). The underestimation of long, moderate events at high latitudes but the underestimation of short, intense events at low latitudes will be further propagated into hydrological models when using IMERG as the inputs. In sum, the uneven sampling density of the polar- and low-orbits naturally exerts a region-dependent bias, which is distinguished from the regional differences stemming from different surface types or cloud regimes (Ferraro et al., 2013; Tan et al., 2022). Therefore, efforts to improve and unify satellite retrieval capability under different climate conditions might never minimize the impact of PMW sensor sampling density that results in regional-dependent bias. A path forward is the improvement of geostationary precipitation estimates (Kirstetter et al., 2012; Upadhyaya et al., 2022). Also, this fact further emphasizes that fundamental improvement for the multi-satellite merged precipitation products will rely on increasing the spatial–temporal coverage of high-quality observations by adding more PMW sensors in the GPM constellation (Kidd et al., 2021), although the ability of PMW sensors to detect light rain needs to be further improved.

In this study, we define a precipitation event as a continuous temporal half-hourly rainy ($\geq$0.1 mm/h) series. Compared with the commonly used minimum interevent time (MIT) criterion for event definition, our definition is equivalent to "MIT = 0.5 h" because of the temporal resolution of our data (0.5 h). Due to the complexity of precipitation processes, there is no unified and standard MIT values for event definition. Since the event properties are apparently sensitive to the chosen MIT, it is necessary to know the robustness of the inferred satellite bias on the selection of MIT. Therefore, we have repeated our analysis using MIT = 3 h (the main text is for MIT = 0.5 h) and the results are shown in Figures S5 and S6. Although the details are unavoidably altered, the main bias patterns remain similar. These results also coincide with the previous event-based evaluations, where robust performance is found under different MIT values (Freitas et al., 2020; Li et al., 2021).

5. Conclusions

Precipitation naturally occurs in the form of discrete "events". Accurate depiction of precipitation events advances our understanding of the multi-scale atmospheric dynamic processes and our assessment of the land surface hydrological response to precipitation. While satellite precipitation products have been widely evaluated and applied since their inception, knowledge about their performance in terms of reproducing precipitation event characteristics and timing remains insufficient. In this study, we use 3-year GV-MRMS data to conduct an event-based evaluation of the IMERG product over CONUS from three progressive levels: (1) the spatial variability and statistical distribution of event characteristics (depth, duration, mean event precipitation rate, peak intensity, etc.), (2) the detection rate (and false detection rate) of precipitation events, and (3) the ability to accurately capture the timing (start, peak, and end time) of the detected events. The main conclusions are as follows:

1) The spatial distribution of the number of events, event depth, and peak-to-mean intensity ratio is well reproduced by IMERG as inferred by comparison to GV-MRMS over CONUS. However, in terms of the spatial pattern of event duration, mean event precipitation rate, and

relative peak position, significant discrepancies are found between IMERG and GV-MRMS. The event duration bias shows a latitude-dependent pattern from negative (IMERG underestimating mean event duration) in the North to positive in the South. IMERG underestimates the occurrence frequency of short-duration, high-intensity events, especially in the summer, but overestimates the occurrence frequency of long-duration, moderate-intensity events, especially in the winter.

2) On average over CONUS, only about 50% of the GV-MRMS events overlap (for at least 30 min) with an IMERG event, and vice versa (which are defined as the "hit events"). However, nearly half of the missed and false events result from wrong timing of events (mostly short events). The other half corresponds to isolated missed or false events, when IMERG simply does not detect any event (miss-isolated events) or invents a nonexistent event (false-isolated events). The event detectability rises with the longer event duration. The winter season suffers from the poorest performance in detecting events with EPOD below 0.2, and contrasting EFAR above 0.8 in the North Central U.S.
3) When IMERG successfully detects a GV-MRMS event, the temporal overlap between the two is about 70% on average, which reflects the imperfect timing of IMERG-derived events as compared to GV-MRMS events. Specifically, IMERG events tend to start, peak, and end all earlier than GV-MRMS events, with national mean shifts of −0.44 h (26 min), −0.29 h (17 min), and −0.12 h (7 min), respectively. For about only 20% of all situations the starting time of the event is correctly reproduced by IMERG, and the same applies to the peak time and end time. IMERG-derived events suffer from the most severe mistiming in the northern U.S. in winter, with their mean start time 1.5-hour earlier and end time 2-hour later than the real events. Duration-dependent time shifts are revealed. The event start time transits from advanced to delayed with the longer duration, while the opposite is true for the event end time.

Compared with the previous event-based evaluations of satellite precipitation products that mainly focused on the mean event property differences (Freitas et al., 2020; Li et al., 2021), this study goes one step further to investigate satellite-derived event' mismatching and mistiming issues, providing more detailed knowledge and insight on their applications and improvements. Moreover, the seasonal analyses also disentangle the markedly different features and error sources in summer and winter. These contrasting seasonal errors largely canceled out each other in the annual results presented in previous studies, thus hiding crucial information. In addition, this study is also the first systematic event-based evaluation of satellite precipitation products in the U.S. and the first to use the comparable gauge-radar merged reference dataset for comparison.

Many points encountered during the present analysis could be further explored in future works. For example, based on the mismatching and mistiming issues of satellite-derived events revealed in the study, future studies should further quantify their impact on the total precipitation amount and frequency bias to make targeted improvements in retrieval. Since the temporal mismatches between GV-MRMS and IMERG are likely to relate to the spatial inconsistency of precipitation features (e.g., size, location, and propagation speed and direction), further analysis of the spatial–temporal evolution characteristics of precipitation events is promising and expected to provide more insight into the cause of this error (Li et al., 2020). Besides, while the detected events with mistiming issues might be ameliorated with additional adjustments, the completely missed or false events, which have been suggested to make up a nonnegligible proportion of the total error, could not be tackled easily in this way. Therefore, more effort should be made to understand how these entirely ignored or fabricated events come into being. For the asymmetric satellite performances in the course of events, analysis that combines additional environmental variables and thermodynamical and microphysical factors is expected to provide insight into the underlying mechanisms and reveal approaches to improve the retrieval. Last but not least, the method proposed here could be applied to other satellite precipitation products or even to an event-based evaluation of the outputs of weather and climate models.

CRediT authorship contribution statement

Runze Li: Conceptualization, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. **Clement Guilloteau:** Formal analysis, Investigation, Methodology, Writing – review & editing. **Pierre-Emmanuel Kirstetter:** Data curation, Funding acquisition, Resources, Writing – review & editing. **Efi Foufoula-Georgiou:** Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgements

This research was supported by NASA through the Global Precipitation Measurement Missions Program (Grants 80NSSC22K0597 and 80NSSC19K0681) and the Ground Validation Program (Grant 80NSSC21K2045). The research was also partially supported by the National Science Foundation (NSF) under the TRIPODS+X program (Grant DMS-1839336).

Data availability statement

The GV-MRMS data used in this study is available at https://doi.org/10.5067/GPMGV/MRMS/DATA101. IMERG data are archived at the NASA GES DISC (https://pmm.nasa.gov/data-access/downloads/gpm).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2023.129563.

References

Amjad, M., Yilmaz, M.T., Yucel, I., Yilmaz, K.K., 2020. Performance evaluation of satellite- and model-based precipitation products over varying climate and complex topography. J. Hydrol. 584, 124707 https://doi.org/10.1016/j.jhydrol.2020.124707.

Asong, Z.E., Razavi, S., Wheater, H.S., Wong, J.S., 2017. Evaluation of integrated multisatellite retrievals for GPM (IMERG) over southern Canada against ground precipitation observations: a preliminary assessment. J. Hydrometeorol. 18 (4), 1033–1050. https://doi.org/10.1175/Jhm-D-16-0187.1.

Ayat, H., Evans, J.P., Behrangi, A., 2021. How do different sensors impact IMERG precipitation estimates during hurricane days? Remote Sens. Environ. 259, 112417 https://doi.org/10.1016/j.rse.2021.112417.

Beck, H.E., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A.I.J.M., Weedon, G.P., Brocca, L., Pappenberger, F., Huffman, G.J., Wood, E.F., 2017. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. Hydrol. Earth Syst. Sci. 21 (12), 6201–6217.

Beck, H.E., Pan, M., Roy, T., Weedon, G.P., Pappenberger, F., van Dijk, A.I.J.M., Huffman, G.J., Adler, R.F., Wood, E.F., 2019. Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. Hydrol. Earth Syst. Sci. 23 (1), 207–224.

Berg, P., Moseley, C., Haerter, J.O., 2013. Strong increase in convective precipitation in response to higher temperatures. Nat. Geosci. 6 (3), 181–185. https://doi.org/10.1038/ngeo1731.

Chen, R., Li, Z., Kuligowski, R.J., Ferraro, R., Weng, F., 2011. A study of warm rain detection using A-Train satellite data. Geophys. Res. Lett. 38 (4).

Derin, Y., Kirstetter, P.E., 2022. Evaluation of IMERG over CONUS complex terrain using environmental variables. e2022GL100186 Geophys. Res. Lett. 49 (19). https://doi.org/10.1029/2022GL100186.
Derin, Y., Kirstetter, P.E., Gourley, J.J., 2021. Evaluation of IMERG satellite precipitation over the land-coast-ocean continuum. Part I: detection. J. Hydrometeorol. 22 (11), 2843–2859. https://doi.org/10.1175/Jhm-D-21-0058.1.
Derin, Y., Yilmaz, K.K., 2014. Evaluation of multiple satellite-based precipitation products over complex topography. J. Hydrometeorol. 15 (4), 1498–1516. https://doi.org/10.1175/Jhm-D-13-0191.1.
Dinku, T., Ceccato, P., Cressman, K., Connor, S.J., 2010. Evaluating detection skills of satellite rainfall estimates over desert locust recession regions. J. Appl. Meteorol. Climatol. 49 (6), 1322–1332. https://doi.org/10.1175/2010jamc2281.1.
Dunkerley, D., 2012. Effects of rainfall intensity fluctuations on infiltration and runoff: rainfall simulation on dryland soils, Fowlers Gap, Australia. Hydrol. Processes 26 (15), 2211–2224. https://doi.org/10.1002/hyp.8317.
Dunkerley, D., 2017. An approach to analysing plot scale infiltration and runoff responses to rainfall of fluctuating intensity. Hydrol. Processes 31 (1), 191–206. https://doi.org/10.1002/hyp.10990.
Dunkerley, D.L., 2019. Rainfall intensity bursts and the erosion of soils: an analysis highlighting the need for high temporal resolution rainfall data for research under current and future climates. Earth Surf. Dyn. 7 (2), 345–360. https://doi.org/10.5194/esurf-7-345-2019.
Dunkerley, D., 2021. The importance of incorporating rain intensity profiles in rainfall simulation studies of infiltration, runoff production, soil erosion, and related landsurface processes. J. Hydrol. 603, 126834 https://doi.org/10.1016/j.jhydrol.2021.126834.
Ebert, E.E., Janowiak, J.E., Kidd, C., 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. Bull. Am. Meteorol. Soc. 88 (1), 47–64. https://doi.org/10.1175/Bams-88-1-47.
Ferraro, R.R., Smith, E.A., Berg, W., Huffman, G.J., 1998. A screening methodology for passive microwave precipitation retrieval algorithms. J. Atmos. Sci. 55 (9), 1583–1600. https://doi.org/10.1175/1520-0469(1998)055<1583:Asmfpm>2.0.Co;2.
Ferraro, R.R., Peters-Lidard, C.D., Hernandez, C., Turk, F.J., Aires, F., Prigent, C., Lin, X., Boukabara, S.-A., Furuzawa, F.A., Gopalan, K., Harrison, K.W., Karbou, F., Li, L., Liu, C., Masunaga, H., Moy, L., Ringerud, S., Skofronick-Jackson, G.M., Tian, Y., Wang, N.-Y., 2013. An evaluation of microwave land surface Emissivities over the continental United States to benefit GPM-Era precipitation algorithms. IEEE Trans. Geosci. Remote Sens. 51 (1), 378–398.
Freitas, E.d.S., Coelho, V.H.R., Xuan, Y., Melo, D.d.C.D., Gadelha, A.N., Santos, E.A., Galvão, C.d.O., Ramos Filho, G.M., Barbosa, L.R., Huffman, G.J., Petersen, W.A., Almeida, C.D.N., 2020. The performance of the IMERG satellite-based product in identifying sub-daily rainfall events and their properties. J. Hydrol. 589, 125128. https://doi.org/10.1016/j.jhydrol.2020.125128.
Gebregiorgis, A.S., Kirstetter, P.-E., Hong, Y.E., Carr, N.J., Gourley, J.J., Petersen, W., Zheng, Y., 2017. Understanding overland multisensor satellite precipitation error in TMPA-RT products. J. Hydrometeorol. 18 (2), 285–306. https://doi.org/10.1175/Jhm-D-15-0207.1.
Guan, M.F., Sillanpaa, N., Koivusalo, H., 2016. Storm runoff response to rainfall pattern, magnitude and urbanization in a developing urban catchment. Hydrol. Processes 30 (4), 543–557. https://doi.org/10.1002/hyp.10624.
Guilloteau, C., Foufoula-Georgiou, E., Kummerow, C.D., Petkovic, V., 2018. Resolving surface rain from GMI high-frequency channels: limits imposed by the three-dimensional structure of precipitation. J. Atmos. Ocean Technol. 35 (9), 1835–1847. https://doi.org/10.1175/JTECH-D-18-0011.1.
Guilloteau, C., Foufoula-Georgiou, E., Kirstetter, P.E., Tan, J., Huffman, G.J., 2021. How well do multi-satellite products capture the space-time dynamics of precipitation? Part I: five products assessed via a wavenumber-frequency decomposition. J. Hydrometeorol. 22 (11), 2805–2823. https://doi.org/10.1175/jhm-d-21-0075.1.
Haile, A.T., Rientjes, T.H.M., Habib, E., Jetten, V., Gebremichael, M., 2011. Rain event properties at the source of the Blue Nile River. Hydrol. Earth Syst. Sci. 15 (3), 1023–1034. https://doi.org/10.5194/hess-15-1023-2011.
Hanel, M., Maca, P., 2014. Spatial variability and interdependence of rain event characteristics in the Czech Republic. Hydrol. Processes 28 (6), 2929–2944. https://doi.org/10.1002/hyp.9845.
Hou, A.Y., Kakar, R.K., Neeck, S., Azarbarzin, A.A., Kummerow, C.D., Kojima, M., Oki, R., Nakamura, K., Iguchi, T., 2014. The global precipitation measurement mission. Bull. Am. Meteorol. Soc. 95 (5), 701–722.
Hsiang, S.M., Burke, M., Miguel, E., 2013. Quantifying the influence of climate on human conflict. Science 341 (6151), 1235367. https://doi.org/10.1126/science.1235367.
Huang, A., Zhao, Y., Zhou, Y., Yang, B., Zhang, L., Dong, X., Fang, D., Wu, Y., 2016. Evaluation of multisatellite precipitation products by use of ground-based data over China. J. Geophys. Res.-Atmos. 121 (18), 10654–10675.
Huffman, G.J., et al., 2020. Integrated Multi-satellite Retrievals for the Global Precipitation Measurement (GPM) Mission (IMERG). In: Levizzani, V. (Ed.), Satellite Precipitation Measurement. Advances in Global Change Research. Springer International Publishing, Cham, pp. 343–353. https://doi.org/10.1007/978-3-030-24568-9_19.
Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Stocker, E.F., Tan, J., 2019b. V06 IMERG Release Notes. 1 pp. [Available online at https://pps.gsfc.nasa.gov/Documents/IMERG_V06_release_notes_190604.pdf.].
Huffman, G.J. et al., 2019a. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellitE Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) Version 6.0. 34 pp. [Available online at https://pps.gsfc.nasa.gov/Documents/IMERG_ATBD_V5.1.pdf.].
Ignaccolo, M., De Michele, C., 2010. A point based Eulerian definition of rain event based on statistical properties of inter drop time intervals: an application to Chilbolton data. Adv. Water Resour. 33 (8), 933–941. https://doi.org/10.1016/j.advwatres.2010.04.002.
Imaoka, K., Nakamura, K., 2012. Statistical analysis of the life cycle of isolated tropical cold cloud systems using MTSAT-1R and TRMM data. Mon. Weather Rev. 140 (11), 3552–3572. https://doi.org/10.1175/Mwr-D-11-00364.1.
Kidd, C., Huffman, G., Maggioni, V., Chambon, P., Oki, R., 2021. The global satellite precipitation constellation: current status and future requirements. Bull. Am. Meteorol. Soc. 102 (10), E1844–E1861. https://doi.org/10.1175/Bams-D-20-0299.1.
Kirstetter, P.-E., Hong, Y., Gourley, J.J., Chen, S., Flamig, Z., Zhang, J., Schwaller, M., Petersen, W., Amitai, E., 2012. Toward a framework for systematic error modeling of spaceborne precipitation radar with NOAA/NSSL ground radar based national mosaic QPE. J. Hydrometeorol. 13 (4), 1285–1300.
Kirstetter, P.E., Karbalaee, N., Hsu, K.L., Hong, Y., 2018. Probabilistic precipitation rate estimates with space-based infrared sensors. Q. J. R. Meteorolog. Soc. 144, 191–205. https://doi.org/10.1002/qj.3243.
Kirstetter, P.E. et al., 2014. Research Framework to Bridge from the Global Precipitation Measurement Mission Core Satellite to the Constellation Sensors Using Ground-Radar-Based National Mosaic QPE, Remote Sensing of the Terrestrial Water Cycle. Geophysical Monograph Series, pp. 61-79. https://doi.org/10.1002/9781118872086.ch4.
Kotz, M., Levermann, A., Wenz, L., 2022. The effect of rainfall changes on economic production. Nature 601 (7892), 223–227. https://doi.org/10.1038/s41586-021-04283-8.
Lamjiri, M.A., Dettinger, M.D., Ralph, F.M., Guan, B., 2017. Hourly storm characteristics along the US West Coast: role of atmospheric rivers in extreme precipitation. Geophys. Res. Lett. 44 (13), 7020–7028. https://doi.org/10.1002/2017gl074193.
Li, Z., Wright, D.B., Zhang, S.Q., Kirschbaum, D.B., Hartke, S.H., 2020. Object-Based Comparison of Data-Driven and Physics-Driven Satellite Estimates of Extreme Rainfall. 21(12): 2759-2776. https://doi.org/10.1175/jhm-d-20-0041.1.
Li, R.Z., Wang, K.C., Qi, D., 2018. Validating the integrated multisatellite retrievals for global precipitation measurement in terms of diurnal variability with hourly gauge observations collected at 50,000 stations in China. J. Geophys. Res.-Atmos. 123 (18), 10423–10442. https://doi.org/10.1029/2018jd028991.
Li, R.Z., Wang, K.C., Qi, D., 2021. Event-based evaluation of the GPM multisatellite merged precipitation product from 2014 to 2018 over China: methods and results. e2020JD033692 J. Geophys. Res.-Atmos. 126 (1). https://doi.org/10.1029/2020JD033692.
Li, R.Z., Qi, D., Zhang, Y., Wang, K.C., 2022. A new pixel-to-object method for evaluating the capability of the GPM IMERG product to quantify precipitation systems. J. Hydrol. 613, 128476 https://doi.org/10.1016/j.jhydrol.2022.128476.
Lobell, D.B., Schlenker, W., Costa-Roberts, J., 2011. Climate trends and global crop production since 1980. Science 333 (6042), 616–620. https://doi.org/10.1126/science.1204531.
Lochbihler, K., Lenderink, G., Siebesma, A.P., 2017. The spatial extent of rainfall events and its relation to precipitation scaling. Geophys. Res. Lett. 44 (16), 8629–8636. https://doi.org/10.1002/2017gl074857.
Maggioni, V., Meyers, P.C., Robinson, M.D., 2016. A review of merged high-resolution satellite precipitation product accuracy during the tropical rainfall measuring mission (TRMM) era. J. Hydrometeorol. 17 (4), 1101–1117. https://doi.org/10.1175/Jhm-D-15-0190.1.
Mao, Y.N., Wu, G.C., Xu, G.Z., Wang, K.C., 2022. Reduction in Precipitation Seasonality in China from 1960 to 2018. J. Clim. 35 (1), 227–248. https://doi.org/10.1175/Jcli-D-21-0324.1.
Maranan, M., Fink, A.H., Knippertz, P., Amekudzi, L.K., Atiah, W.A., Stengel, M., 2020. A process-based validation of GPM IMERG and its sources using a mesoscale rain gauge network in the West African Forest Zone. J. Hydrometeorol. 21 (4), 729–749. https://doi.org/10.1175/Jhm-D-19-0257.1.
McCabe, M.F., Rodell, M., Alsdorf, D.E., Miralles, D.G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N.E.C., Franz, T.E., Shi, J., Gao, H., Wood, E.F., 2017. The future of earth observation in hydrology. Hydrol. Earth Syst. Sci. 21 (7), 3879–3914. https://doi.org/10.5194/hess-21-3879-2017.
Mei, Y.W., Anagnostou, E.N., Nikolopoulos, E.I., Borga, M., 2014. Error analysis of satellite precipitation products in mountainous basins. J. Hydrometeorol. 15 (5), 1778–1793. https://doi.org/10.1175/Jhm-D-13-0194.1.
Mei, Y.W., Nikolopoulos, E.I., Anagnostou, E.N., Zoccatelli, D., Borga, M., 2016. Error analysis of satellite precipitation-driven modeling of flood events in complex alpine terrain. Remote Sens. 8 (4), 293. https://doi.org/10.3390/rs8040293.
Nikolopoulos, E.I., Anagnostou, E.N., Borga, M., 2013. Using high-resolution satellite rainfall products to simulate a major flash flood event in Northern Italy. J. Hydrometeorol. 14 (1), 171–185. https://doi.org/10.1175/Jhm-D-12-09.1.
Orland, E., Kirschbaum, D., Stanley, T., 2022. A scalable framework for post fire debris flow hazard assessment using satellite precipitation data. e2022GL099850 Geophys. Res. Lett. 49 (18). https://doi.org/10.1029/2022GL099850.
Passive Microwave Algorithm Team Facility, 2017. GLOBAL PRECIPITATION MEASUREMENT (GPM) MISSION Algorithm Theoretical Basis Document GPROF2017 Version 1 and Version 2 (used in GPM V5 processing). 43 pp. [Available online at https://pps.gsfc.nasa.gov/Documents/ATBD_GPM_V5B_April15_2018.pdf.].
Petersen, W.A., Kirstetter, P.E., Wang, J., Wolff, D.B., Tokay, A., 2020. The GPM Ground Validation Program. In: Levizzani, V. (Ed.), Satellite Precipitation Measurement. Advances in Global Change Research. Springer International Publishing, Cham, pp. 471–502. https://doi.org/10.1007/978-3-030-35798-6_2.

Petkovic, V., Kummerow, C.D., 2017. Understanding the sources of satellite passive microwave rainfall retrieval systematic errors over land. J. Appl. Meteorol. Climatol. 56 (3), 597–614. https://doi.org/10.1175/Jamc-D-16-0174.1.

Pradhan, R.K., Markonis, Y., Vargas Godoy, M.R., Villalba-Pradas, A., Andreadis, K.M., Nikolopoulos, E.I., Papalexiou, S.M., Rahim, A., Tapiador, F.J., Hanel, M., 2022. Review of GPM IMERG performance: a global perspective. Remote Sens. Environ. 268, 112754.

Siepielski, A.M., Morrissey, M.B., Buoro, M., Carlson, S.M., Caruso, C.M., Clegg, S.M., Coulson, T., DiBattista, J., Gotanda, K.M., Francis, C.D., Hereford, J., Kingsolver, J. G., Augustine, K.E., Kruuk, L.E.B., Martin, R.A., Sheldon, B.C., Sletvold, N., Svensson, E.I., Wade, M.J., MacColl, A.D.C., 2017. Precipitation drives global variation in natural selection. Science 355 (6328), 959–962.

Sinha, E., Michalak, A.M., Balaji, V., 2017. Eutrophication will increase during the 21st century as a result of precipitation changes. Science 357 (6349), 405–408. https://doi.org/10.1126/science.aan2409.

Skofronick-Jackson, G., Petersen, W.A., Berg, W., Kidd, C., Stocker, E.F., Kirschbaum, D. B., Kakar, R., Braun, S.A., Huffman, G.J., Iguchi, T., Kirstetter, P.E., Kummerow, C., Meneghini, R., Oki, R., Olson, W.S., Takayabu, Y.N., Furukawa, K., Wilheit, T., 2017. The global precipitation measurement (Gpm) mission for science and society. Bull. Am. Meteorol. Soc. 98 (8), 1679–1695. https://doi.org/10.1175/BAMS-D-15-00306.1.

Smith, T.M., Arkin, P.A., Bates, J.J., Huffman, G.J., 2006. Estimating bias of satellite-based precipitation estimates. J. Hydrometeorol. 7 (5), 841–856. https://doi.org/10.1175/Jhm524.1.

Tadesse, A., Anagnostou, E.N., 2009. The effect of storm life cycle on satellite rainfall estimation error. J. Atmos. Oceanic Technol. 26 (4), 769–777. https://doi.org/10.1175/2008jtecha1129.1.

Tan, J.C., Cho, N.Y., Oreopoulos, L., Kirstetter, P., 2022. Evaluation of GPROF V05 precipitation retrievals under different cloud regimes. J. Hydrometeorol. 23 (3), 389–402. https://doi.org/10.1175/Jhm-D-21-0154.1.

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

Tan, J., Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Rajagopal, M., 2021. SHARPEN: a scheme to restore the distribution of averaged precipitation fields. J. Hydrometeorol. 22 (8), 2105–2116. https://doi.org/10.1175/Jhm-D-20-0225.1.

Tang, G.Q., Ma, Y.Z., Long, D., Zhong, L.Z., Hong, Y., 2016. Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. J. Hydrol. 533, 152–167. https://doi.org/10.1016/j.jhydrol.2015.12.008.

Tang, G.Q., Clark, M.P., Papalexiou, S.M., Ma, Z.Q., Hong, Y., 2020. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. Remote Sens. Environ. 240, 111697 https://doi.org/10.1016/j.rse.2020.111697.

Upadhyaya, S.A., Kirstetter, P.E., Kuligowski, R.J., Searls, M., 2022. Towards improved precipitation estimation with the GOES-16 advanced baseline imager: algorithm and evaluation. Q. J. R. Meteorolog. Soc. 148 (748), 3406–3427. https://doi.org/10.1002/qj.4368.

Utsumi, N., Kim, H., Turk, F.J., Haddad, Z.S., 2019. Improving satellite-based subhourly surface rain estimates using vertical rain profile information. J. Hydrometeorol. 20 (5), 1015–1026. https://doi.org/10.1175/Jhm-D-18-0225.1.

Wasko, C., Sharma, A., Johnson, F., 2015. Does storm duration modulate the extreme precipitation-temperature scaling relationship? Geophys. Res. Lett. 42 (20), 8783–8790. https://doi.org/10.1002/2015gl066274.

Wu, H., Adler, R.F., Hong, Y., Tian, Y.D., Policelli, F., 2012. Evaluation of global flood detection using satellite-based rainfall and a hydrologic model. J. Hydrometeorol. 13 (4), 1268–1284. https://doi.org/10.1175/Jhm-D-11-087.1.

Wu, H., Adler, R.F., Tian, Y., Huffman, G.J., Li, H., Wang, J.J., 2014. Real-time global flood estimation using satellite-based precipitation and a coupled land surface and routing model. Water Resour. Res. 50 (3), 2693–2717. https://doi.org/10.1002/2013wr014710.

Yin, J., Pan, Z., Rosenfeld, D., Mao, F., Zang, L., Zhu, Y., Hu, J., Chen, J., Gong, J., 2022. Full-tracking algorithm for convective thunderstorm system from initiation to complete dissipation. J. Geophys. Res.-Atmos. 127 (22) https://doi.org/10.1029/2022JD037601.

You, Y.L., Meng, H., Dong, J., Rudlosky, S., 2019. Time-lag correlation between passive microwave measurements and surface precipitation and its impact on precipitation retrieval evaluation. Geophys. Res. Lett. 46 (14), 8415–8423. https://doi.org/10.1029/2019gl083426.

Young, M.P., Williams, C.J.R., Chiu, J.C., Maidment, R.I., Chen, S.H., 2014. Investigation of discrepancies in satellite rainfall estimates over Ethiopia. J. Hydrometeorol. 15 (6), 2347–2369. https://doi.org/10.1175/Jhm-D-13-0111.1.

Zhang, J., Howard, K., Langston, C., Kaney, B., Qi, Y., Tang, L., Grams, H., Wang, Y., Cocks, S., Martinaitis, S., Arthur, A., Cooper, K., Brogden, J., Kitzmiller, D., 2016. MULTI-RADAR MULTI-SENSOR (MRMS) QUANTITATIVE PRECIPITATION ESTIMATION Initial Operating Capabilities. Bull. Am. Meteorol. Soc. 97 (4), 621–638. https://doi.org/10.1175/Bams-D-14-00174.1.

Zhang, Y., Wang, K.C., 2023. Global precipitation system scale increased from 2001 to 2020. J. Hydrol. 616, 128768 https://doi.org/10.1016/j.jhydrol.2022.128768.