

Research papers

Global intercomparison and regional evaluation of GPM IMERG Version-03, Version-04 and its latest Version-05 precipitation products: Similarity, difference and improvements

Cunguang Wang^a, Guoqiang Tang^a, Zhongying Han^a, Xiaolin Guo^a, Yang Hong^{a,b,*}

^a State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing, China

^b Department of Civil Engineering and Environmental Sciences, University of Oklahoma, Norman, OK, United States

ARTICLE INFO

This manuscript was handled by Emmanouil Anagnostou, Editor-in-Chief, with the assistance of Amir AghaKouchak, Associate Editor

Keywords:
Precipitation
Satellite
GPM
IMERG

ABSTRACT

The overarching goal of this study is to intercompare the newly released Integrated Multi-satellitE Retrievals for GPM (IMERG) Version 05 (V05) products with its former Version 04 (V04) and Version 03 (V03) products and also assess any differences and improvements, with cross-evaluation against the Global Precipitation Climatology Project (GPCP) Version 2.3, Multi-Source Weighted-Ensemble Precipitation (MSWEP) Version 2.1 and the dense gauge networks in China. Firstly, the gauge-adjusted products (Final run) of V03, V04 and V05 are compared over the globe. Then, the near-real-time products without gauge adjustments (Early and Late run) and Final run products of all versions are evaluated against ground-based observations comprised of more than 30,000 gauges over Mainland China at $0.1^\circ \times 0.1^\circ$ grid and hourly and daily temporal scales. The primary conclusions are: (1) globally, both V04 and V05 Final run show significant differences and improvements from V03. Particularly, the overall mean oceanic precipitation of V04 and V05 increases by +31.36% and +28.81% respectively from that of V03 and much closer to GPCP and MSWEP; (2) over Mainland China, the Early and Late run products of the same version (V03 or V04) generally have similar performance, while V04 Early and Late run have better performance in most regions than the corresponding run of V03 except in the arid Xinjiang Province and the mountainous Tibetan Plateau; and (3) V04 and V03 Final run show comparable performance, while V05 Final run generally improves upon both V04 and V03 and has the best performance among the seven standard IMERG products. The improvement of V05 Final run is particularly evident in southeastern and western China. At a timely matter, the study provides first-hand global and regional assessment feedback to IMERG algorithm developers and also sheds insights for GPM precipitation product users across the world.

1. Introduction

Precipitation is a vital variable in the global water and energy cycle (Kidd and Huffman, 2011), and precise precipitation measurement has direct benefits on water resource management, disaster prevention, weather prediction and so on (Hou et al., 2014; Ma et al., 2017). However, precipitation measurement has always been a highly demanding yet challenging research topic for scientists. Typically, there are three main methods to measure precipitation, i.e., rain gauges, weather radars and satellite-based remote sensing (Yong et al., 2015). Satellite-based remote sensing gives a complementary perspective for precipitation measurement compared with traditional ground-based observations, and is the only effective way to obtain comprehensive precipitation measurement at the global scale (Hong et al., 2012; Hou

et al., 2014; Tang et al., 2017).

Since the launch of the Tropical Rainfall Measuring Mission (TRMM) satellite by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA) in 1997, a large number of scientists have developed diverse quasi-global satellite-based precipitation products through different techniques (Huffman et al., 2007; Joyce et al., 2004; Sorooshian et al., 2000; Hong et al., 2004; Kubota et al., 2007). As the successor of TRMM, the Global Precipitation Measurement (GPM) mission led by NASA and JAXA is the state-of-the-art precipitation measurement program in the world. On February 28, 2014, the GPM Core Observatory was successfully launched. The GPM constellation is composed of the core observatory and other approximately eight partner satellites. The GPM Core Observatory carries the first spaceborne Dual-frequency phased array

* Corresponding author at: Room A207, State Key Laboratory of Hydroscience and Engineering, Department of Hydraulic Engineering, Tsinghua University, Beijing 100084, China.

E-mail address: hongyang@tsinghua.edu.cn (Y. Hong).

Precipitation Radar, the DPR (the Ku-band at 13.6 GHz and Ka-band at 35.5 GHz) and a multichannel GPM Microwave Imager, the GMI (frequency ranging from 10 to 183 GHz). The two instruments (DPR and GMI) are the more advanced versions compared to Precipitation Radar (PR) and TRMM Microwave Imager (TMI) onboard the TRMM satellite. Therefore, the capability of detecting light rain and snow is improved and the two instruments (DPR and GMI) act as an intercalibration library to unify observations from diverse partner constellation radiometers (Tapiador et al., 2012; Hou et al., 2014). The level-3 products of GPM produced by the Integrated Multi-satellite Retrievals for GPM (IMERG) are widely used among hydrology communities.

IMERG products were firstly provided by NASA as Version 03 (V03) from March 2014. Many researchers have published a series of studies to reveal characteristics of uncertainties and errors of IMERG V03 products (Tang et al., 2016a, b; Ma et al., 2016; Li et al., 2017; Guo et al., 2016; Gaona et al., 2016; Tan et al., 2016, 2017; Asong et al., 2017; Mayor et al., 2017). Precipitation Processing System (PPS), the organization at NASA Goddard Space Flight Center (GSFC) that processes and distributes precipitation records from GPM, releases newer IMERG Version 04 (V04) and Version 05 (V05) Final run products successively on March 22 and November 20, 2017. The changes from V03 to V04 and from V04 to V05 are multifactorial.

IMERG V04 experienced several upgrades compared to the previous V03. For example, IMERG V04 employs the 2017 version 1 of the Goddard Profiling Algorithm (GPROF2017 v1) to compute precipitation estimates from all microwave sensors, which is an upgrade version compared with GPROF2014 used in IMERG V03. Furthermore, the Advanced Temperature and Moisture Sounder (ATMS), a multi-channel passive microwave radiometer is added as a new source of precipitation estimates. Additional changes from V03 to V04 include calibrating infrared (IR) estimates dynamically to the IMERG high quality (HQ) data, calibrating 2BCMB to GPCP V2.3 over ocean at middle and high latitudes and over land globally, extending IMERG HQ data coverage from 60°N-S to 90°N-S, correcting the bug that mistakenly assigned 1/3 Kalman weighting to IR in certain cases and so on (Huffman et al., 2017c).

On the basis of IMERG V04, the latest V05 applies a more advanced version of GPROF (V05) to compute microwave-based precipitation estimates. In addition, more fields and components are added into V05 products. For example, GPROF estimates from all available microwave sensors are provided in the “HQprecipitation” field. The “Quality Index” is also added for all half-hour and monthly V05 products. When combining the satellite estimates with gauge observations, gauge error estimates are defined to provide feasible weights for correction. Meanwhile, some procedures of IMERG algorithms which are associated with potential issues are removed for safety. For example, the TMI estimates are not computed in GPROF, and the MHS and ATMS estimates are also restricted to specific footprints at each swath edge, which is treated to reduce the amount of microwave-based

precipitation estimates to some extent (Huffman et al., 2017e).

The transition from V04 to V05 only takes eight months. In addition, to our best knowledge, currently there is no study analyzing the quality of IMERG V05 products and comprehensively intercomparing IMERG V03, V04 and V05 to clarify their similarities, differences and improvements. Huffman et al. (2017e) conducted initial validation for IMERG Final Run products compared to the Multi-Radar MultiSensor (MRMS) based ground observations over conterminous US. Wei et al. (2017) evaluated IMERG V03, V04 and other satellite-based precipitation products with gauge observations in three climatologic zones in China and found V04 didn't show an advantage over V03 in the Tibetan Plateau region and the Weihe River Basin. Zhao et al. (2017) investigated the successive IMERG V03 and V04 Final Run products with a comparison of the China daily Precipitation Analysis Products with a resolution of $0.25^\circ \times 0.25^\circ$ over Mainland China and also found that V04 didn't exhibit the anticipated improvement for China compared to V03. Su et al. (2018) evaluated the daily precipitation product from IMERG V04 Final Run for monitoring three heavy rain events mainly over eastern China in 2016 and proved that the daily IMERG V04 Final Run products can basically capture the spatiotemporal patterns of heavy rain events. Therefore, it is urgent to address the following question which is crucial for hydrology applications: How do the three versions of IMERG products perform both at global and regional scales? The objectives of this study are twofold: 1) intercomparing IMERG V03, V04 and V05 Final run products over the globe at $0.1^\circ \times 0.1^\circ$, daily resolutions; 2) evaluating and comparing IMERG V03, V04 and V05 Early, Late and Final run products over Mainland China at $0.1^\circ \times 0.1^\circ$ spatial resolution and multiple temporal resolutions (hourly and daily) against ground-based observations from the China Meteorological Administration (CMA). The remainder of this paper is organized as follows. Section 2 mainly describes the study areas and data sets. Section 3 introduces the methodology, i.e., calculation of evaluation metrics. Section 4 intercompares and evaluates IMERG V03, V04 and V05. Finally, the conclusions, discussions and recommended future works are introduced in Section 5.

2. Study area and data sets

2.1. Study area

In global analysis, the study area covers 60°N-S. In regional analysis, the study area is Mainland China (Fig. 1(a)). Eight sub-regions are defined in Fig. 1(b) according to Chen et al. (2013) and Guo et al. (2016) to comprehensively investigate the error characteristics of IMERG products depending on topography (Tang et al., 2006), and climatic conditions (Guo et al., 2007; Tang et al., 2016a; Qian and Lin, 2005). The sub-regions are: (1) the Xinjiang (XJ) region in central Asia, which is dominated by an arid or semiarid climate with little precipitation over the whole year; (2) the Qinghai-Tibetan plateau (TP) dominated

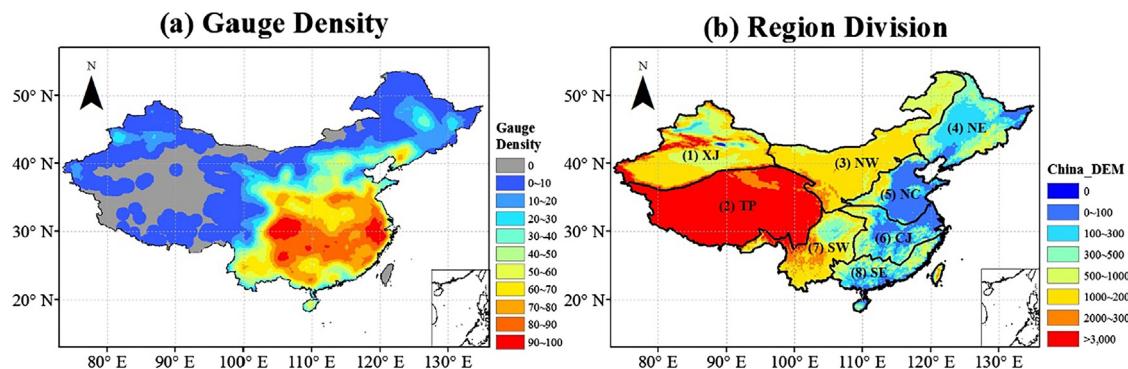


Fig. 1. (a) Spatial distribution of rain gauges shown as gauge density; (b) division of sub-regions in China. Eight sub-regions are outlined with solid black lines and marked with abbreviations for regional evaluations.

by the Plateau mountain climate with high altitude, complex topography and significant precipitation variability; (3) Northwestern China (NW) bounded by the 400 mm annual precipitation isohyet; (4) Northeastern China (NE) located in the north of the Yan Mountains; (5) Northern China (NC) located in the north of the Qinling Mountains and the Huai River. The climate of NW, NE and NC is primarily controlled by the temperate continental climate with a hot and wet summer and a cold and dry winter. (6) the plain region of Changjiang (Yangtze) River (CJ) located in the Middle-Lower reaches of Yangtze River; (7) the Yungui Plateau in southwestern China (SW) is bounded by the Dabashan Mountains to the north and the Wulingshan Mountains to the east; (8) Southeast China (SE) bounded by the Nanling Mountains to the north and the Wuyishan Mountains to the northwest. The climate of CJ and SE is primarily controlled by the subtropical or tropical monsoon from the western Pacific Ocean, characterized by a hot and wet summer and a mild and dry winter, while the major factor affecting the climate of SW is the monsoon between the TP and the Indian Ocean.

2.2. Data sets

2.2.1. Satellite precipitation estimates data sets

The IMERG system merges all available satellite microwave precipitation estimates, microwave-calibrated IR estimates, ground gauge analyses, and potentially other estimates at $0.1^\circ \times 0.1^\circ$ spatial and half-hour temporal resolutions (Huffman et al., 2017f, g). IMERG provides three kinds of products, including Early and Late run about 4 h and 12 h respectively after observation, and Final run about 2 months after the observation month (Huffman et al., 2017a, b). Early and Late run products are produced in near real time and calibrated with climatological coefficients varying by month and location. In contrast, the Final run product is produced post real time after gauge adjustments to be consistent with the Global Precipitation Climatology Centre (GPCC) precipitation data (Huffman et al., 2017f, g). Therefore, the Final run product is characterized by higher accuracy particularly over land, while the Early and Late run products have better timeliness which is appealing to flood forecast and monitoring.

In this study, the Final run products of IMERG V03 (V03_F), V04 (V04_F) and V05 (V05_F), and the Early and Late run products of IMERG V03 (V03_E and V03_L, respectively) and V04 (V04_E and V04_L, respectively) are used. As IMERG V05 Early and Late run products are not available yet, they are absent in the study (Huffman et al., 2017d). For the Final run, IMERG V03_F, V04_F and V05_F are intercompared from June 1, 2014 to May 31, 2015, because IMERG V03_F stopped updating in February 2016 and a full two-year overlapping dataset is not available. For the Early and Late run, IMERG V03_E/V03_L and V04_E/V04_L are compared from June 1, 2015 to May 31, 2016, because IMERG V03 Early and Late run products were available since March 2015. The raw resolutions of these products are $0.1^\circ \times 0.1^\circ$ and half hour. Hourly and daily data sets used in this study are accumulated from half-hourly data sets. The IMERG data could be downloaded from the Precipitation Measurement Missions (PMM) website (Huffman et al., 2017h; <https://pmm.nasa.gov/data-access/downloads/gpm>).

The absence of accurate and independent surface-based observations across the globe makes it impossible to directly evaluate the quality of IMERG products globally. As an alternative strategy, two other global precipitation products are used as the reference in the intercomparison. One is the Global Precipitation Climatology Project (GPCP) Version 2.3 monthly precipitation data set at $2.5^\circ \times 2.5^\circ$ spatial resolution (Adler et al., 2003, 2017). GPCP combines various satellite precipitation products over ocean and land and a gauge analysis over land to provide a consistent estimation of global precipitation (Huffman et al., 1997). GPCP has become a robust and widely used global precipitation product after several iterations of improvements. Another product is Multi-Source Weighted-Ensemble Precipitation (MSWEP) Version 2.1 which is a new global precipitation data set at 3-hourly

temporal and $0.1^\circ \times 0.1^\circ$ spatial resolution (Beck et al., 2017a). MSWEP utilizes various gauge-, satellite- and reanalysis-based data to produce an integrated precipitation estimates over the globe (Beck et al., 2017a), and it performed best among 22 precipitation datasets in a comprehensive global evaluation (Beck et al., 2017b). The potential issue for the intercomparison using other global precipitation datasets such as GPCP and MSWEP is that they may have some overlapping precipitation data sources as IMERG products. For example, both GPCP and IEMRG contain GPCC for bias correction and share some of input datasets, e.g., IR and Special Sensor Microwave Imager/Sounder (SSMIS) data (Adler et al., 2018). However, the three global precipitation estimates (IMERG, GPCP and MSWEP) are produced by different retrieval algorithms, and the relative difference of IMERG V03, V04 and V05 that the paper focuses on are mainly due to the update of IMERG algorithms but not the input datasets. Moreover, the employment of two benchmarks (GPCP and MSWEP) would also contribute to the confidence of the intercomparison not just over land but more importantly over ocean where there are no such anchors like GPCC for bias correction. The GPCP V2.3 could be accessed from Physical Science Division (PSD) in NOAA (<https://www.esrl.noaa.gov/psd/data/gridded/data.gpcp.html>). And MSWEP could be accessed in the website of Toward Locally Relevant Global Hydrological Simulations (GLOH2O) (<http://www.gloh2o.org/>).

2.2.2. Rain gauge data sets

More than 30,000 gauges over Mainland China are adopted to provide hourly ground-based precipitation observations as independent reference from June 1, 2014 to May 31, 2016 (Shen et al., 2014). All gauges are categorized into IMERG $0.1^\circ \times 0.1^\circ$ grid cells. Only grid cells with at least one gauge are chosen to conduct evaluation, and if more than one gauge in one grid cell, averaged value is employed. This processing method has been widely used in many other similar works (Cohen Liechti et al., 2012; Demaria et al., 2011; Liebmann and Allured, 2005; Ma et al., 2018; Salio et al., 2015; Tang et al., 2018a; Xu et al., 2017). The hourly gauge observations go through several steps of quality control, including (1) eliminating extreme values, (2) examining internal consistency, and (3) examining spatial consistency (Shen et al., 2010). The spatial distribution of gauge density is displayed in Fig. 1(a). The majority of gauges are located in southern and eastern China, while fewer gauges are located in northern and western China. Note that the limited gauges, especially in northern and western China, could contribute to the uncertainties in the evaluation of IMERG products. As said before, GPCC networks are used as reference to adjust IMERG Final run products. However, GPCC networks only use data from 194 China's stations, which means that the gauges used in this paper (more than 30,000) are mostly (> 99%) independent of GPCC networks (Shen et al., 2013) guaranteeing the reliability of evaluation results. Besides, the gauge adjustments for IMERG Final run products are conducted at the monthly scale, while the evaluation of this paper are carried out at hourly and daily scales.

3. Methodology

To quantify the accuracy of IMERG products of different versions, eight statistic metrics are used in this study (Table 1). The Pearson correlation coefficient (CC) describing the linear correlation level between satellite precipitation estimates and gauge observations is able to measure the capability of reproducing the temporal variations. The bias ratio (β) describing the error and bias is able to measure the overestimation ($\beta > 1$) or underestimation ($\beta < 1$) of satellite precipitation products. The variability ratio (γ) describes the difference of the distributions of precipitation products and gauge observations. The modified Kling-Gupta efficiency (KGE') is an integrated statistic consisting of CC, β and γ (Kling et al., 2012; Zambrano-Bigiarini et al., 2017). Another indice describing errors is the root mean square error (RMSE). Contingency table indices, which include the probability of

Table 1

List of statistic metrics used in the evaluation and comparison of IMERG V03/V04.^a

Statistic metrics	Equation	Perfect value
Correlation Coefficient (CC)	$CC = \frac{\frac{1}{N} \sum_{n=1}^N (S_n - \bar{S})(G_n - \bar{G})}{\sigma_S \sigma_G}$	1
Bias Ratio (β)	$\beta = \frac{\bar{S}}{\bar{G}}$	1
Variability Ratio (γ)	$\gamma = \frac{\sigma_S / \bar{S}}{\sigma_G / \bar{G}}$	1
Modified KGE (KGE')	$KGE' = 1 - \sqrt{(CC)^2 + (\beta)^2 + (\gamma)^2}$	1
Root Mean Squared Error (RMSE)	$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (S_n - G_n)^2}$	0
Probability of Detection (POD)	$POD = \frac{n_{11}}{n_{11} + n_{01}}$	1
False Alarm Ratio (FAR)	$FAR = \frac{n_{10}}{n_{11} + n_{10}}$	0
Critical Success Index (CSI)	$CSI = \frac{n_{11}}{n_{11} + n_{01} + n_{10}}$	1

^aNotation: N is the number of samples; S_n is the satellite precipitation estimate; \bar{S} is the averaged satellite precipitation estimate; G_n is the gauge based precipitation observation; \bar{G} is the averaged gauge based precipitation observation; σ_S is the standard deviations of satellite precipitation series; σ_G is the standard deviations of gauge based precipitation observation; n_{11} is the number of precipitation events observed by satellites and gauges at the same time; n_{01} is the number of precipitation events observed by gauges but not by satellites; n_{10} is the opposite to n_{01} ; n_{00} is the number of precipitation events observed neither by satellites nor gauges.

detection (POD), the false alarm ratio (FAR), and the critical success index (CSI) are included to describe the consistency between estimated and observed rain event occurrence. The rain/no rain threshold is set to 0.1 mm for both hourly and daily scales (Tang et al., 2016a).

4. Results

4.1. Global intercomparison between IMERG V03, V04 and V05 products

4.1.1. Global view of the spatial difference

The spatial distribution of global mean daily precipitation of IMERG V05_F is shown in Fig. 2(a). In addition, by subtracting V03_F, V04_F and V05_F precipitation rates from those of GPCP, the spatial differences between the three products and GPCP are displayed in Fig. 2(b)–(d), respectively. The resolution of V03_F, V04_F, V05_F and MSWEP is averaged to $2.5^\circ \times 2.5^\circ$ to be consistent with GPCP. Over the global ocean, the mean precipitation rate of V04_F and V05_F is ~ 3.10 and ~ 3.04 mm/day, which is about 30% higher than that of V03_F (~ 2.36 mm/day) and much closer to that of the globally reanalyzed benchmark (GPCP, ~ 2.94 mm/day; MSWEP, ~ 2.93 mm/day). This phenomenon could mainly be attributed to the upgraded schemes both in V04 and V05 that climatologically calibrates the GMI-DPR Combined Level-2 product (2BCMB), to GPCP V2.3 over ocean at middle and high latitudes (Huffman et al., 2017a, b). 2BCMB is the calibrator for all of the passive microwave radiometers in the GPM constellation. Besides, the upgraded GPROF algorithms, GPROF2017 version 1 for V04 and GPROF V05 for V05, are also contributed to the significant differences. According to the study by Behrangi et al. (2014), the near-global (60°N – S) oceanic precipitation rate for 3-yr (2007–09) by the merged precipitation estimate from the CloudSat, TRMM, and Aqua platforms (MCTA), Climate Prediction Center Merged Analysis of Precipitation (CMAP) and GPCP version 2.2 is 3.13, 2.96 and 2.97 mm/day, respectively, which could further demonstrate the consistency between oceanic precipitation estimates from V04_F/V05_F and other satellite precipitation products. While over the global land, different products have relatively comparable mean daily precipitation rates: about 2.30 mm/day for V03_F, V04_F, V05_F and GPCP and about 2.22 mm/day for MSWEP, which is mainly due to the gauge-correction

composition (GPCC) in most precipitation products over land.

Over ocean, as shown in Fig. 2(b–d), V04_F and V05_F precipitation is much larger and closer to GPCP than V03_F's, e.g., northeastern and mid-southern Pacific Ocean, northwestern and southwestern Atlantic Ocean and the Equatorial Zones. While over land, the three versions have similar spatial distributions of the differences with GPCP. And the differences between V03_F and V04_F/V05_F are relatively small, with negative and positive differences distributing alternatively. This result can be partially attributed to the rough spatial average for IMERG products from $0.1^\circ \times 0.1^\circ$ to $2.5^\circ \times 2.5^\circ$ grids, which makes slight differences invisible. Thus, more specific evaluation at IMERG's raw spatial resolution should be carried out over land, e.g., the Mainland China, which is executed as below.

Seasonal differences between IMERG V03_F and V05_F are also analyzed as representation (Fig. 3). There are notable positive differences over ocean in all seasons particularly in Equatorial and high latitudes, and the difference in MAM is relatively small. In contrast, over land, most regions have relatively low differences and apparent differences only occur in equatorial regions, such as northern South America and central Africa. In addition, the difference exhibits strong seasonality over some land regions, e.g., the distinctive difference in India and southern and western China mainly occurs in JJA when the monsoon is dominant in those regions. The seasonality of differences can also be found in Northwest Pacific Ocean, where there are large positive adjustments in JJA but negative adjustments in DJF. However, along both sides of the equator in Middle and East Pacific Ocean and Atlantic Ocean, the difference remains similar in all four seasons.

4.1.2. Analysis on zonal mean

Fig. 4 presents the curves of the latitudinal distribution of the annual mean precipitation for IMERG V03_F, V04_F, V05_F, GPCP and MSWEP over both ocean and land. Over ocean, the profile curves of five products have similar latitudinal pattern and they are stable and smooth. The rainfall peak is located at 6.5°N . This is mainly due to the Intertropical Convergence Zone (ITCZ), where a large number of precipitation events move forward north of the Equator in boreal summer (Wu et al., 2003). On the contrary, the minimum rainfall amount is located near the Tropics of Cancer and Capricorn (23.5°N/S) and 60°S , as the prevalence of downdraft under the control of high pressure in these areas (Hendon et al., 2014). Compared with V03_F, the curves of V04_F and V05_F are much closer to that of GPCP and MSWEP, especially between 0 and 20°N and between 10°S and 20°S . V04_F and V05_F show slight overestimation than benchmarks overall, while V03_F has significant underestimation, mostly occurring in high latitude zones and Equatorial zones (see Fig. 2(b)).

Over land, it is noted that the variation of precipitation profiles is much larger and complicated than that over ocean, which is mainly due to the widespread variability of continental precipitation related to topography at local scale (Yong et al., 2015). The maximum rainfall over land is located at about 1.5°S . Furthermore, the curves of V03_F, V04_F, V05_F and GPCP are quite close to each other, especially between 20°N/S . This is mainly because of the rain gauge adjustments using the GPCC full analysis for IMERG Final run products over land. MSWEP shows smaller magnitudes than other four products for most latitude ranges, while relatively large values appear near the 50°S .

4.2. Evaluation of IMERG V03, V04 and V05 in China

In this section, the products including IMERG V03 (V03_E, V03_L and V03_F), V04 (V04_E, V04_L and V04_F) and V05 (V05_F) are evaluated in China using independent high-density ground gauges. The evaluation of IMERG products in China is executed in two different perspectives, i.e., national-based and grid-based, in order to enhance the reliability of the evaluation conclusions. In national-based evaluation, the different versions of IMERG products are compared using averaged values over the whole country. In grid-based evaluation, the

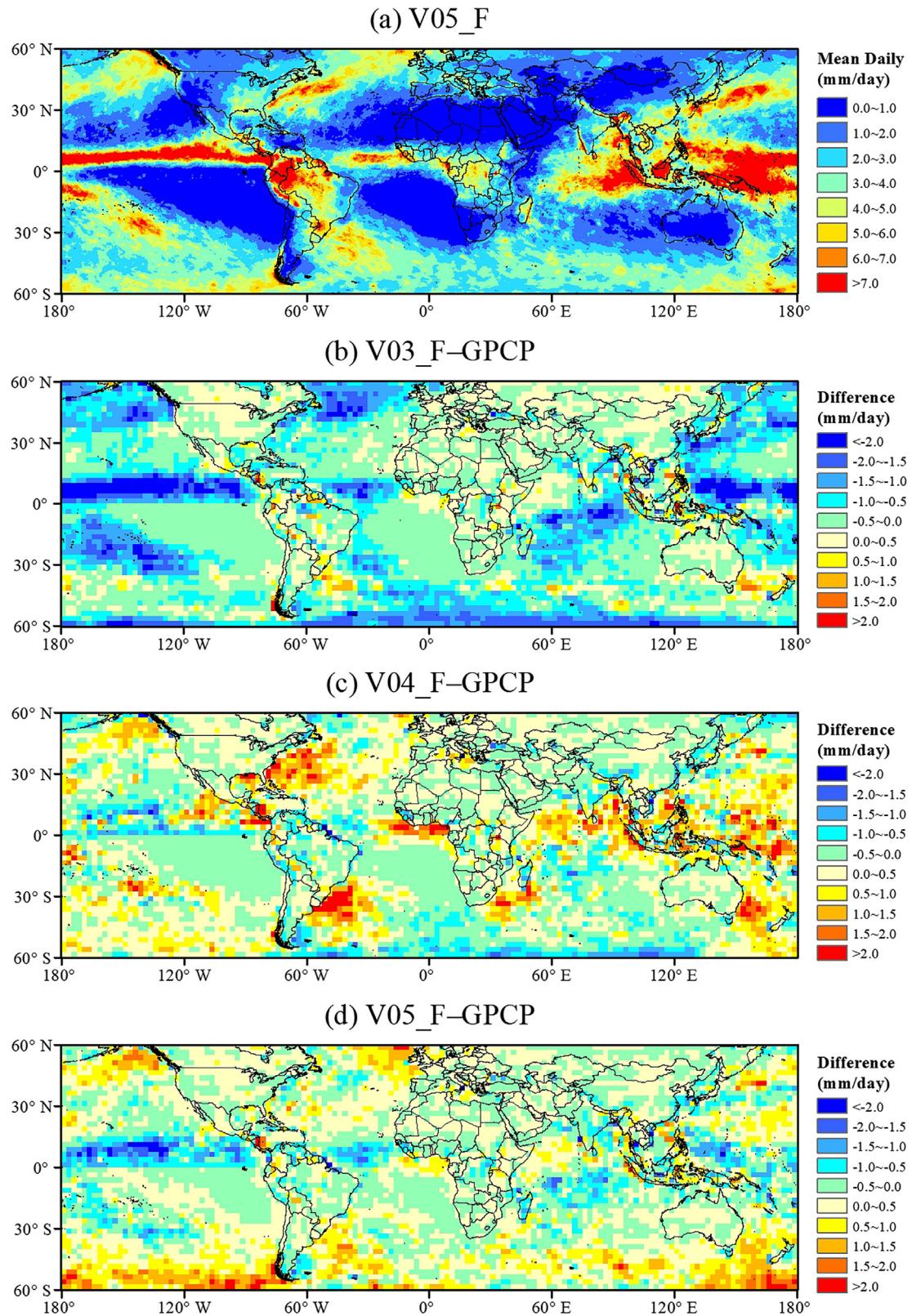


Fig. 2. Spatial distribution of global mean daily precipitation of (a) IMERG V05_F and the difference between (b) V03_F and GPCP (V03_F – GPCP), (c) V04_F and GPCP (V04_F – GPCP) and (d) V05_F and GPCP (V05_F – GPCP) for the 1-year study period (Jun 2014–May 2015).

different versions are compared using the exact value at each grid cell. The period for the Final run products is from June 2014 to May 2015 at $0.1^\circ \times 0.1^\circ$ spatial resolution and hourly and daily temporal resolutions. While the Early and Late run products are from June 2015 to May 2016 at $0.1^\circ \times 0.1^\circ$ spatial resolution and hourly temporal resolution.

4.2.1. National-based evaluation

The spatial distribution of gauge-based mean daily precipitation over China is shown in Fig. 5(a), indicating that most rain events occur in southern and eastern China. Benchmarked by gauge observations, V03_F, V04_F and V05_F all exhibit good capacity in capturing the

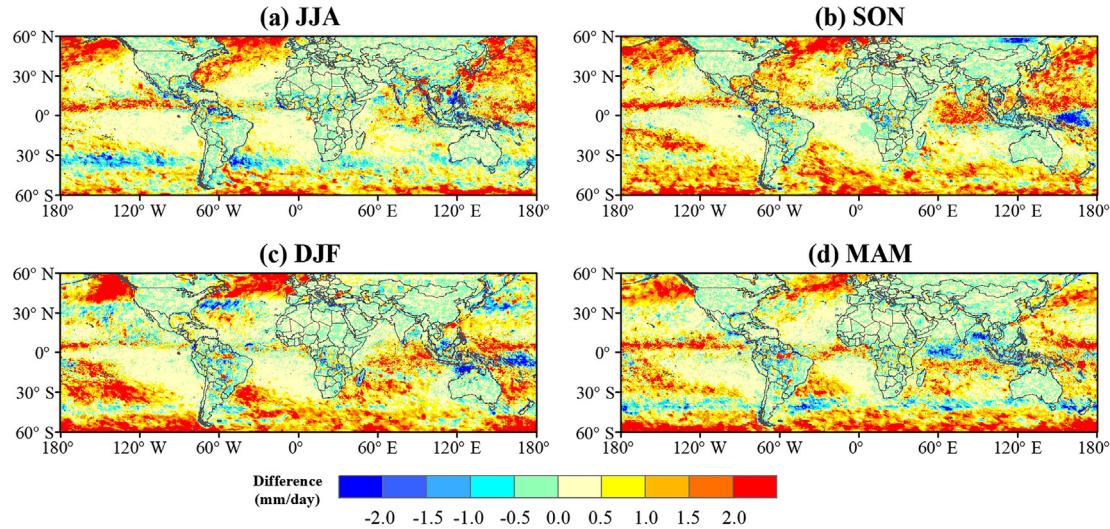


Fig. 3. Seasonal variation of mean daily precipitation difference between IMERG V03_F and V05_F (V05_F – V03_F) for the 1-year study period (Jun 2014–May 2015): boreal (a) summer (Jun–Aug), (b) autumn (Sep–Nov), (c) winter (Dec–Feb), (d) spring (Mar–May).

spatial characteristics of daily mean precipitation over Mainland China (Fig. 5(b)–(d)). However, all versions seem to overestimate precipitation in southeastern and northeastern China. Impressively, it is abnormal that V04_F shows banded distribution of precipitation in TP and XJ, which disappears in the latest V05_F. Fig. 5(e)–(f) are snapshots of mean daily precipitation difference between V04_F and V03_F and between V05_F and V03_F. It is obvious that V04_F tends to produce higher precipitation than V03_F in northern and central CJ, northern SW and southwestern SE, where V05_F also shows overestimation but with milder magnitude. In southern TP, V04_F produces much lower precipitation but V05_F shows no apparent difference.

Table 2 shows the summary of evaluation metrics calculated with the mean value over the whole country at hourly and daily scales. It's clear that all indices look satisfying at both scales. The high CC values more than 0.9 indicate strong correlation between IMERG products and gauge based observation, implying the capacity of IMERG to quantify precipitation time series over large regions. The β values larger than 1 indicate overestimation of the satellite precipitation estimates in China. To further investigate this phenomenon, the GPCC Monitoring Product (Version 4) which is used as the calibrator in IMERG Final run products (Schneider et al., 2011; Huffman et al., 2017a, b) is compared with gauge data at the monthly temporal resolution. All gauges are categorized into GPCC $1^\circ \times 1^\circ$ grid cells, and only grid cells with at least one gauge are chosen to conduct evaluation and if there is more than one gauge in one grid cell, averaged value is employed. The β value

calculated by mean coincident GPCC and gauge data is 1.11, which is almost same as the β values by comparing various IMERG products to gauge data shown in Table 2. This finding implies that GPCC calibration for IMERG Final run products tends to overestimate precipitation over Mainland China leading to the overestimation of IMERG. More representative ground stations are required to be included into GPCC networks in China to improve the performance of satellite precipitation estimates. The RMSE values are low at the hourly scale, and they become higher at the daily scale due to error accumulation. Good performance of contingency indices (POD, FAR and CSI) demonstrates the capacity of IMERG products to capture most precipitation events accurately in China. On the whole, the performance of V04_F and V05_F is comparable with V03_F as the average of the whole country.

4.2.2. Grid-based evaluation

4.2.2.1. Probability density function of precipitation intensity. Probability density function (PDF) is widely used for comparing characteristics of different precipitation data sets in terms of precipitation intensity and estimation sensitivity (Kirstetter et al., 2013; Chen et al., 2013). In addition, PDF is good at describing light precipitation rates (Kirstetter et al., 2013), which makes it a good method to investigate the performance of IMERG to detect light precipitation (Tang et al., 2016a). Fig. 6 shows the PDFs of IMERG V03_F, V04_F and V05_F at hourly and daily resolutions with different precipitation intensities over Mainland China, benchmarked by gauge observations. The first bin of

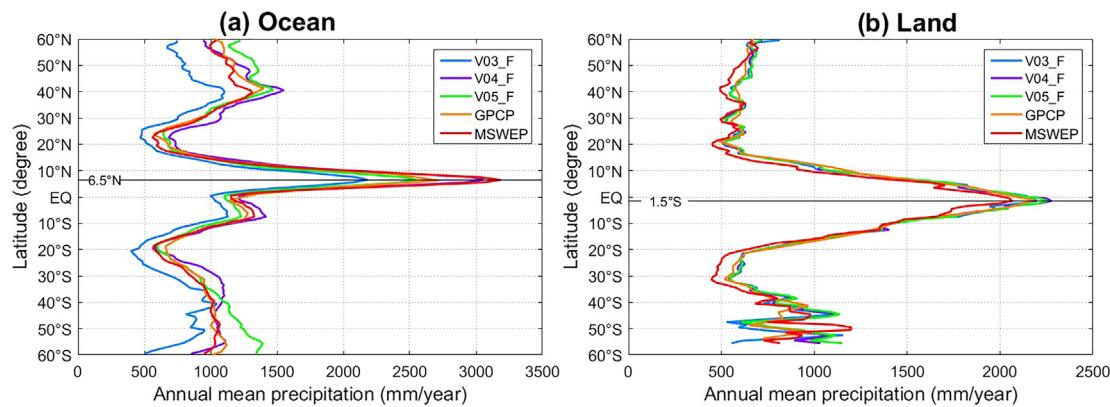


Fig. 4. Latitudinal distribution of the annual mean precipitation of IMERG V03_F, V04_F, V05_F, GPCP and MSWEP over (a) ocean and (b) land for the 1-year study period (Jun 2014–May 2015).

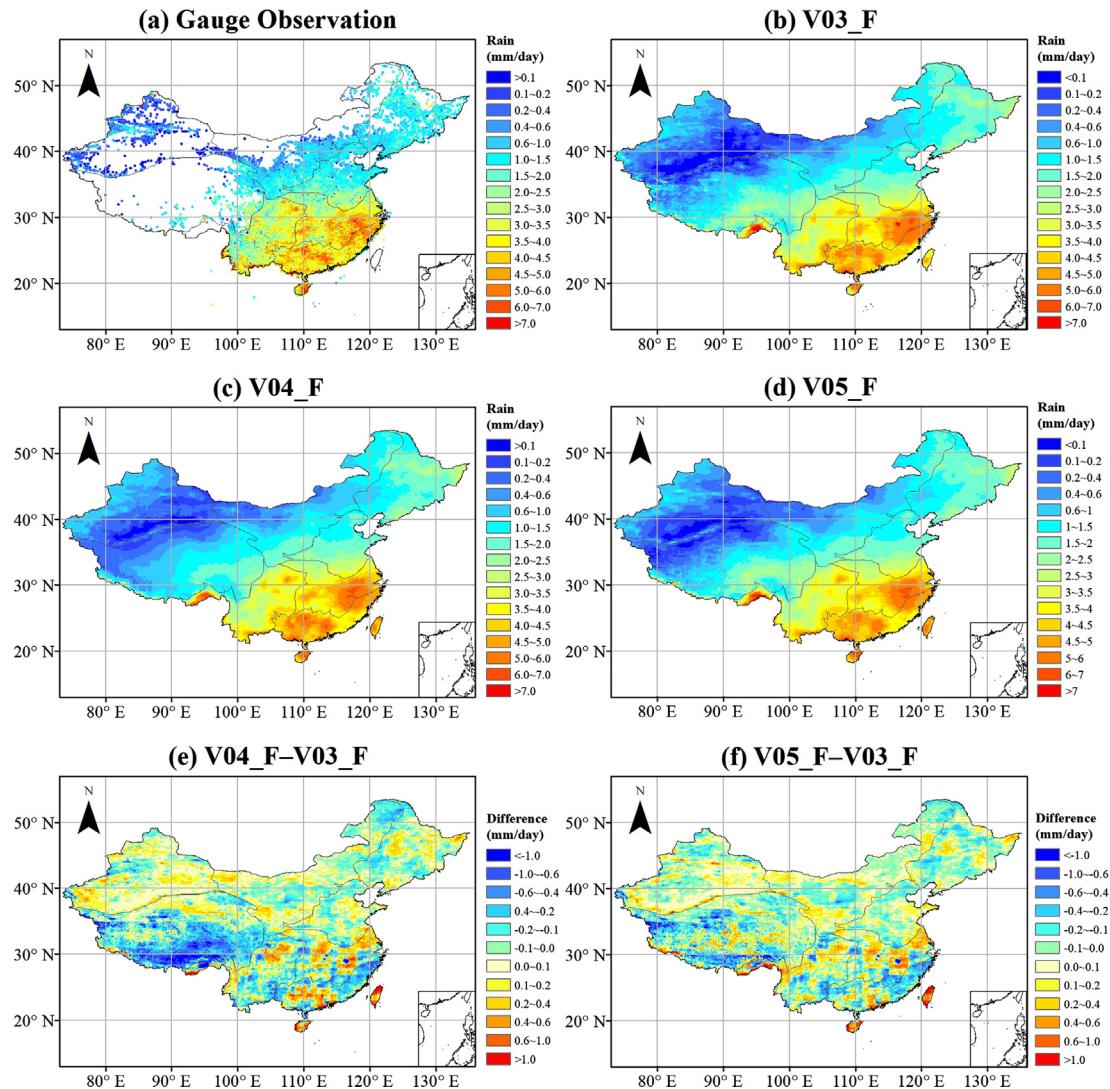


Fig. 5. Spatial distribution of daily mean precipitation of (a) gauge observation, (b) IMERG V03_F, (c) V04_F, (d) V05_F and their daily difference (e) V04_F – V03_F, (f) V05_F – V03_F over Mainland China for the 1-year study period (Jun 2014–May 2015).

Table 2

Summary of evaluation metrics for IMERG V03_F/V04_F/V05_F calculated with the mean value over the whole country at hourly and daily timescales for the 1-year study period (Jun 2014–May 2015).

Timescale	Product	CC	β	γ	KGE'	RMSE (mm)	POD	FAR	CSI
Hourly	V03_F	0.91	1.12	0.89	0.82	0.06	0.91	0.10	0.82
	V04_F	0.92	1.11	0.90	0.83	0.06	0.89	0.10	0.81
	V05_F	0.93	1.13	0.89	0.82	0.06	0.93	0.11	0.84
Daily	V03_F	0.96	1.12	0.89	0.83	1.05	0.93	0.10	0.84
	V04_F	0.96	1.11	0.90	0.84	0.96	0.91	0.11	0.82
	V05_F	0.96	1.13	0.89	0.82	1	0.92	0.12	0.82

both two scenarios is set to 0–0.1 mm/h or mm/day (i.e., no rain) and presented separately due to its large proportion. For the hourly scenario shown in Fig. 6(a), more than 90% precipitation is under 0.1 mm/h. V03_F shows obvious overestimation in this bin while V04_F and V05_F have similar proportions with gauge observations, which demonstrates the improvement of V04_F and V05_F in detecting no rain events. In terms of nonzero rain events, the most evident difference occurs in bins ranging from 0.1 mm/h to 0.4 mm/h. For bins ranging from 0.1 mm/h to 0.2 mm/h, all products underestimate precipitation rates with various degrees while for bins ranging from 0.2 mm/h to 0.4 mm/h,

V03_F and V05_F show underestimation but V04_F shows obvious overestimation. As for moderate and heavy rains (greater than 0.6 mm/h), the three products appear to have more precipitation than gauge observations. This patterns could account for the overall overestimation of IMERG products indicated by bias ratio (greater than 1) in Table 2. For the daily scenario shown in Fig. 6(b), the distribution patterns of three estimates show higher consistence. Nevertheless, V03_F overestimates no rain events and reduces its proportion in light precipitation (0.1–1 mm/day). On the contrary, V04_F and V05_F underestimate no rain events and increase their proportions in nonzero precipitation (> 0.1 mm/day). In general, V03_F, V04_F and V05_F have comparable capacity of reproducing the PDF structure of precipitation at hourly and daily scales.

4.2.2.2. Statistical evaluation of IMERG products. Fig. 7 shows the box plot of grid-scale statistics for IMERG V03_F, V04_F and V05_F at hourly and daily scales over Mainland China. For each box, the central mark is the median, and the edges of the box are the 25th and 75th percentiles. The percentage of points that the whiskers span is set to 9% and 91% plotted by horizontal bands. The purple line represents the perfect value for each indicator. At the hourly scale, V04_F shows similar performance with V03_F concerning CC, β , γ and KGE', while V05_F seems to outperform the other two versions with higher CC and KGE'.

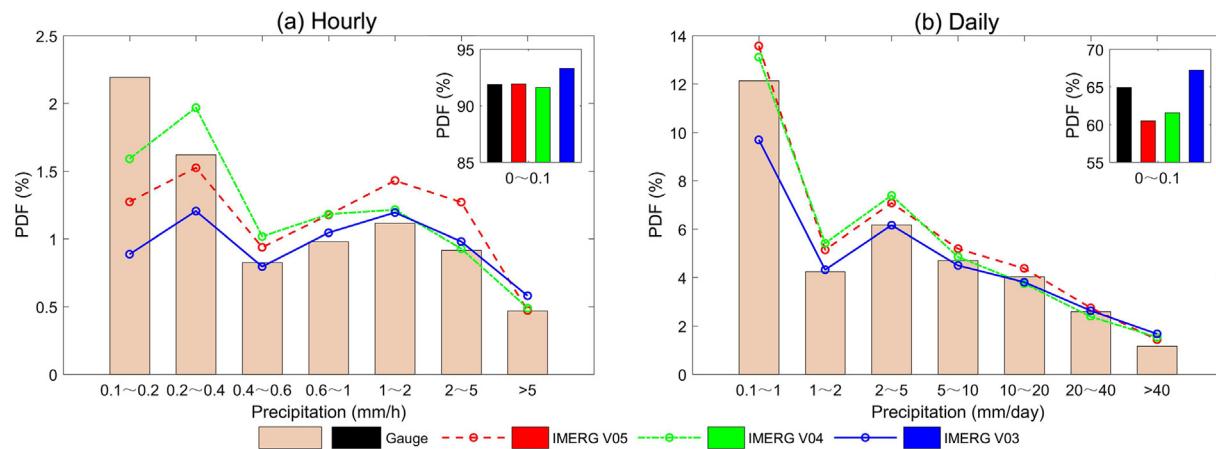


Fig. 6. Probability density function (PDF) of gauge, IMERG V03_F, V04_F and V05_F with different intensities at (a) hourly and (b) daily scales for the 1-year study period (Jun 2014–May 2015).

As for contingency statistics, V04_F and V05_F have much higher POD than V03_F, but at the same time the FAR of them also increases by similar magnitudes, which leads to the same level of CSI. At the daily scale, the difference among the three products is very similar to that at the hourly scenario.

Table 3 summarizes specific values of statistic metrics. The mean hourly CC value increases from ~ 0.31 for V03_F and V04_F to 0.36 for V05_F. Correspondingly, the mean hourly KGE' increases from ~ 0.13 for V03_F and V04_F to 0.18 for V05_F. With respect to the hourly RMSE, V05_F has a lower value (0.93 mm) compared to the value (~ 1.04 mm) for V03_F and V04_F. Overall, V05_F is the best in terms of averaged statistics over Mainland China, while V04_F shows even slightly worse performance than V03_F.

To further reveal the error characteristics of IMERG Final run products over different parts of China, more detailed statistics at the hourly scale for eight sub-regions is also provided in Table 3. All the three

versions of IMERG have relatively better performance in CJ, SW and SE with high CC (larger than 0.3), KGE' (larger than 0.2) and CSI (about 0.3), comparing with low CC (lower than 0.25), KGE' (lower than 0.1) and CSI (lower than 0.20) in XJ and TP. The bad performance in XJ and TP may be attributed to complicated terrain and severe climate conditions. It is worth noting that V03_F, V04_F and V05_F greatly overestimate the precipitation in NE with the β of 1.52, 1.51 and 1.55, respectively. Great overestimation can also be found in NC (1.39, 1.38 and 1.39 for V03_F, V04_F and V05_F, respectively) and NW (1.37, 1.33 and 1.42 for V03_F, V04_F and V05_F, respectively), which are in the south and west to NE, respectively. This phenomenon implies that the version 03, 04 and 05 retrievals of IMERG all have great uncertainty in estimating precipitation in northern China with high latitudes. Despite all this, when considering all eight indices comprehensively, V05_F shows apparent improvement over V03_F and V04_F in eastern China and southern China, while V03_F and V04_F have comparable

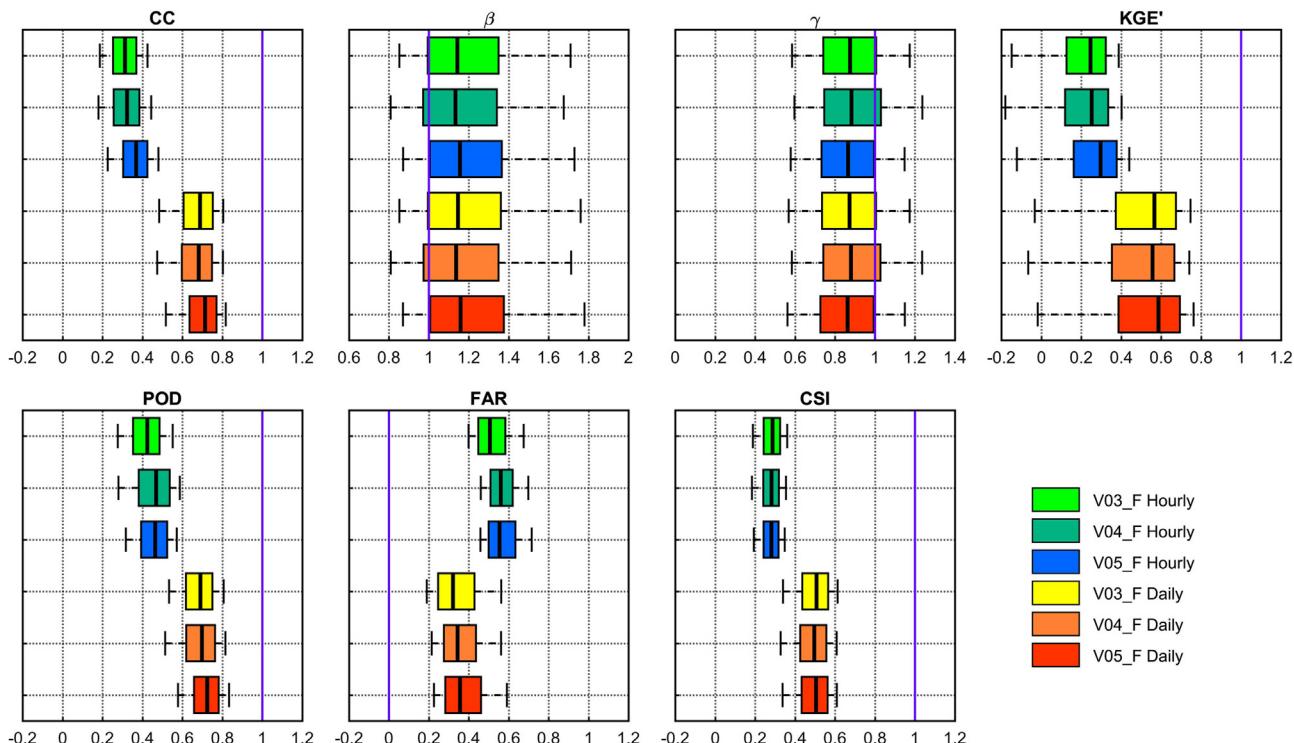


Fig. 7. Box plot of grid based metrics (from left to right and up to bottom): CC, β , γ , KGE', POD, FAR, CSI, calculated by evaluating IMERG V03_F, V04_F and V05_F against gauge observations at hourly and daily scales for the 1-year study period (Jun 2014–May 2015).

Table 3

Summary of mean evaluation metrics for IMERG V03_F/V04_F/V05_F calculated with all grid cells at hourly and daily scales for the 1-year study period (Jun 2014–May 2015).

Region	Timescale	Product	CC	β	γ	KGE'	RMSE(mm)	POD	FAR	CSI
China	Hourly	V03_F	0.31	1.26	0.89	0.13	1.04	0.42	0.52	0.28
		V04_F	0.32	1.23	0.94	0.12	1.05	0.45	0.57	0.28
		V05_F	0.36	1.28	0.87	0.18	0.93	0.45	0.57	0.28
	Daily	V03_F	0.66	1.29	0.89	0.39	7.17	0.68	0.35	0.49
		V04_F	0.66	1.26	0.93	0.38	7.27	0.68	0.37	0.48
		V05_F	0.69	1.31	0.87	0.41	6.51	0.71	0.39	0.49
XJ	Hourly	V03_F	0.17	1.48	1.21	-0.57	0.32	0.28	0.76	0.14
		V04_F	0.19	1.06	1.91	-0.84	0.30	0.24	0.71	0.14
		V05_F	0.23	1.52	0.99	-0.34	0.27	0.39	0.78	0.16
TP	Hourly	V03_F	0.24	1.25	0.94	0.03	0.62	0.27	0.59	0.19
		V04_F	0.22	0.7	1.84	-0.38	0.53	0.22	0.55	0.17
		V05_F	0.27	1.24	0.93	0.08	0.54	0.34	0.63	0.21
NW	Hourly	V03_F	0.28	1.37	0.83	0.04	0.54	0.42	0.57	0.27
		V04_F	0.26	1.33	0.86	0.05	0.59	0.40	0.57	0.26
		V05_F	0.31	1.42	0.80	0.06	0.49	0.45	0.62	0.25
NE	Hourly	V03_F	0.26	1.52	0.73	-0.03	0.75	0.45	0.63	0.25
		V04_F	0.25	1.51	0.74	-0.04	0.82	0.42	0.65	0.23
		V05_F	0.29	1.55	0.72	-0.03	0.68	0.49	0.68	0.24
NC	Hourly	V03_F	0.32	1.39	0.81	0.09	0.79	0.52	0.55	0.32
		V04_F	0.32	1.38	0.81	0.09	0.82	0.52	0.60	0.29
		V05_F	0.36	1.39	0.80	0.12	0.73	0.52	0.62	0.28
CJ	Hourly	V03_F	0.32	1.18	0.91	0.21	1.28	0.43	0.48	0.30
		V04_F	0.34	1.21	0.89	0.22	1.26	0.52	0.55	0.32
		V05_F	0.38	1.21	0.88	0.26	1.14	0.47	0.52	0.31
SW	Hourly	V03_F	0.33	1.18	0.95	0.18	1.11	0.37	0.47	0.27
		V04_F	0.34	1.16	0.98	0.18	1.13	0.39	0.52	0.27
		V05_F	0.39	1.19	0.94	0.24	0.99	0.40	0.52	0.27
SE	Hourly	V03_F	0.33	1.13	0.93	0.25	1.54	0.40	0.50	0.28
		V04_F	0.36	1.13	0.93	0.28	1.53	0.49	0.57	0.29
		V05_F	0.40	1.11	0.94	0.32	1.38	0.45	0.53	0.30

performance with each other. While over western regions, V03_F and V05_F show overestimation in TP with the β of ~ 1.25 , while V04_F exhibits severe underestimation with the β of 0.7. The severe underestimation shown by V04_F over TP is also proved in XJ, where V04_F exhibits much lower (approximately 30%) and better (much close to 1) β than the other two versions. In previous studies, V03_F has shown a significant overestimation in TP (Tang et al., 2016a; Ma et al., 2017), so estimation of lower precipitation of V04_F in that areas seems reasonable to some extent (Fig. 5(e)). However, the severe underestimation of V04_F demonstrates that V04_F actually estimate too few precipitation amounts in that regions. In XJ, the good performance of V04_F in terms of β is offset by its extremely bad γ values, which makes V04_F the worst precipitation products as the integrated index, KGE', indicates. It suggests that V04_F could produce accurate total precipitation amounts but the poor estimates of the distribution of precipitation rates. Instead, V05_F shows the best performance in the complex XJ and TP areas, indicated by the highest CC, KGE' and CSI values. Nevertheless, the performance of all three version products at the hourly scale still has much room to be improved.

4.2.2.3. Taylor diagrams for evaluation metrics. Similar to Table 3, detailed statistics about the error characteristics of IMERG Early and Late run products over Mainland China and eight sub-regions are provided in Table 4. In addition, Taylor diagrams (Taylor, 2001) are plotted to graphically summarize how well the seven different precipitation estimates (V03_E, V03_L, V03_F, V04_E, V04_L, V04_F and V05_F) match the gauge observations in terms of CC, standard deviation and centered RMSE (Fig. 8). In the Taylor diagrams, the uneven angular coordinate is CC represented by blue dash-and-dot lines, and the radial coordinate is standard deviation normalized by

the standard deviation value of gauge observation and represented by black dotted lines. In addition, the green concentric semi-circles centered by observation represents centered RMSE values. The gauge observation is marked by a red star, and the closer the marker of precipitation estimates is to the gauge observation, the better performance the product has. The Taylor diagrams suggest that Final run of V03 and V04 don't show expected improvements compared with their Early and Late run algorithms in most areas at the hourly scale over China. As for real-time precipitation products, the Late run shows very similar performance to the Early run of the same version. While V04 Early/Late run has better performance than V03 products in most regions except XJ and TP, implying that the great uncertainty in V04_F over western China could be sourced into the algorithms of Early run. As for post-real-time precipitation products, V05_F improves the quality of precipitation estimates over almost all regions compared to V04_F and V03_F. In western and northern China, it is evident that different precipitation estimates have large variations, which may be due to the huge challenge brought by snow and ice and complex terrain in precipitation retrieval.

4.2.2.4. Spatial differences among the different IMERG versions. Fig. 9 displays the spatial distributions of representative evaluation metrics (CC, β , γ , KGE', POD, FAR and CSI) of IMERG V05_F and the differences between V05_F and V03_F and between V04_F and V03_F over Mainland China. The spatial distribution of V05_F statistics is plotted to represent the performance of IMERG products as their similarities. The β difference is calculated by measuring the difference between the absolute error of V05_F/V04_F and V03_F to 1.0. Therefore, negative β difference means V05_F/V04_F tends to produce more accurate estimates than V03_F. The γ difference is calculated through the same way, and the other differences are calculated by simply subtracting the

Table 4

Summary of evaluation metrics for IMERG V03_E/V04_E/V03_L/V04_L calculated with all grid cells at hourly and daily scales for the 1-year study period (Jun 2015–May 2016).

Region	Timescale	Product	CC	β	γ	KGE'	RMSE (mm)	POD	FAR	CSI
China	Hourly	V03_E	0.27	0.98	1.22	0.02	1.21	0.39	0.51	0.28
		V04_E	0.31	0.97	1.31	-0.01	1.21	0.44	0.56	0.28
		V03_L	0.29	1.00	1.21	0.03	1.19	0.43	0.51	0.30
		V04_L	0.32	0.98	1.29	0.00	1.20	0.48	0.56	0.29
XJ	Hourly	V03_E	0.17	1.21	1.41	-0.49	0.54	0.28	0.68	0.17
		V04_E	0.22	0.40	4.09	-2.40	0.38	0.23	0.63	0.16
		V03_L	0.17	1.13	1.57	-0.57	0.53	0.29	0.67	0.17
		V04_L	0.23	0.42	3.86	-2.19	0.38	0.25	0.62	0.17
TP	Hourly	V03_E	0.21	0.80	1.44	-0.07	0.60	0.24	0.61	0.18
		V04_E	0.18	0.28	4.68	-2.90	0.51	0.18	0.59	0.15
		V03_L	0.21	0.70	1.63	-0.17	0.57	0.26	0.60	0.19
		V04_L	0.19	0.28	4.73	-2.96	0.51	0.20	0.57	0.16
NW	Hourly	V03_E	0.20	1.22	1.00	-0.04	0.74	0.30	0.63	0.19
		V04_E	0.21	0.83	1.48	-0.14	0.66	0.30	0.64	0.19
		V03_L	0.20	1.24	0.99	-0.06	0.75	0.33	0.62	0.21
		V04_L	0.22	0.87	1.41	-0.10	0.66	0.33	0.62	0.21
NE	Hourly	V03_E	0.25	1.77	0.63	-0.22	1.14	0.46	0.61	0.26
		V04_E	0.24	1.24	0.92	0.06	1.00	0.41	0.65	0.23
		V03_L	0.25	1.80	0.63	-0.24	1.14	0.48	0.62	0.27
		V04_L	0.24	1.26	0.90	0.04	0.99	0.44	0.64	0.24
NC	Hourly	V03_E	0.26	1.23	0.95	0.04	1.08	0.41	0.57	0.27
		V04_E	0.30	1.29	0.91	0.05	1.11	0.46	0.65	0.25
		V03_L	0.26	1.29	0.91	0.02	1.09	0.45	0.57	0.28
		V04_L	0.30	1.29	0.91	0.06	1.11	0.49	0.63	0.26
CJ	Hourly	V03_E	0.27	0.83	1.32	0.09	1.31	0.42	0.46	0.31
		V04_E	0.32	1.07	1.02	0.20	1.39	0.53	0.54	0.33
		V03_L	0.30	0.83	1.32	0.11	1.26	0.47	0.46	0.33
		V04_L	0.33	1.09	1.01	0.21	1.36	0.57	0.53	0.34
SW	Hourly	V03_E	0.30	0.94	123	0.10	1.21	0.37	0.47	0.27
		V04_E	0.34	0.83	1.44	0.01	1.17	0.38	0.52	0.26
		V03_L	0.32	0.97	1.20	0.12	1.19	0.42	0.48	0.30
		V04_L	0.35	0.85	1.42	0.02	1.16	0.41	0.52	0.28
SE	Hourly	V03_E	0.31	0.71	1.55	-0.02	1.55	0.42	0.46	0.31
		V04_E	0.35	0.86	1.26	0.18	1.60	0.51	0.51	0.33
		V03_L	0.32	0.72	1.56	-0.02	1.52	0.46	0.47	0.33
		V04_L	0.35	0.86	1.27	0.18	1.57	0.55	0.52	0.34

indices of V03_F from that of V05_F/V04_F.

As for CC, V05_F are in better agreement with gauge data in southern and eastern China with higher CC than other areas, which is consistent with previous studies (Tang et al., 2016a; Chen and Li, 2016). The CC is much worse in northern and western China with values lower than 0.2 in some areas. In addition, the KGE' and CSI values of V05_F share very similar distribution patterns with CC. The underlying reasons for the low performance over western and northern areas include uneven topography, complex climatic and weather conditions posing great challenge to satellite precipitation retrieval (Dinku et al., 2007; Tang et al., 2018b). It should be noted that gauges in such regions are also subject to more severe observation errors such as wind and snowfall induced error measurement, which contributes to the uncertainty in evaluation (Yang et al., 2001; Adam and Lettenmaier, 2003). Considering the comparisons between new versions and the old one, V05_F demonstrates better estimations than V04_F and V03_F across the whole country with higher CC, KGE' and comparable CSI. In particular, V04_F shows worse quality than V03_F in TP indicated by negative differences of CC, KGE' and CSI, while the latest V05_F gets rid of the retrogressive performance of V04_F and evidently improves the performance in those high-altitude and complicated regions.

The spatial distributions of contingency statistics difference are distinct from that of continuous statistics. Concerning POD, V05_F shows better performance than V03_F in most regions except NC and northern SW, while V04_F is much better than V03_F in southeastern China, i.e., CJ and SE, but can't reach the level of V03_F in other

regions. Meanwhile, the higher POD of V05_F and V04_F are accompanied by worse FAR (Fig. 9(q–r)) in those corresponding regions. As a consequence, the distributions of CSI difference exhibit large variability. Over most parts of CJ and SE, V05_F and V04_F has larger CSI than V03_F due to the improvement of POD surpasses the deterioration of FAR. While over most parts of NC and northern SW, the CSI of V05_F and V04_F is far inferior to that of V03_F caused by poorer POD and FAR.

4.2.2.5. Spatial differences within IMERG Early and Late run. Four IMERG near-real-time estimates (V03_E, V04_E, V03_L, and V04_L) are also included into spatial analysis at the hourly scale over Mainland China. Fig. 10 shows the spatial distributions of metric differences between IMERG V03_E and V04_E and between IMERG V03_L and V04_L.

As for IMERG Early run, V04_E is quite different from V03_E. Firstly, V04_E has larger CC over most parts of central and southern China (i.e., NC, CJ, SW and SE) and XJ, and the improvement is more significant and widely distributed compared with the Final run case in Fig. 9(c). The average CC increases from 0.27 to 0.31 compared V04_E to V03_E across the whole Mainland China. Secondly, the performance difference indicated by KGE' is extremely significant (Fig. 10(b)). Improvements of V04 in KGE' can be found over NE, CJ and SE, but degeneration is apparent over TP, XJ and central NC. Thirdly, the RMSE difference is obviously negative over the western and northern China except central NC where RMSE difference is extremely positive, and the difference is

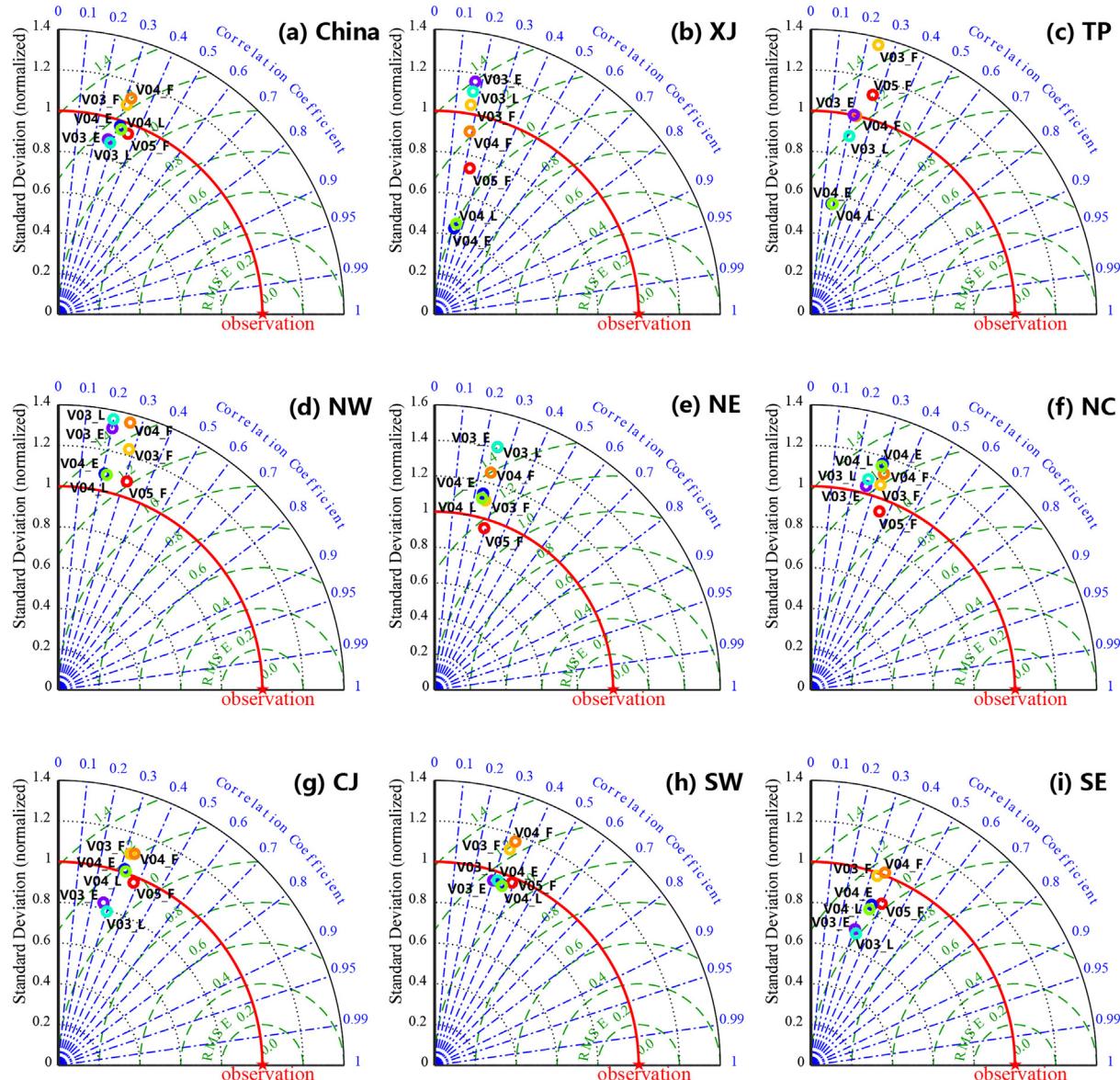


Fig. 8. Taylor diagrams consisting of CC, standard deviation (normalized) and RMSE for hourly precipitation estimates from IMERG V03 (V03_E, V03_L, V03_F), V04 (V04_E, V04_L, V04_F) and V05 (V05_F) over (a) China and eight sub-regions: (b) XJ, (c) TP, (d) NW, (e) NE, (f) NC, (g) CJ, (h) SW, (i) SE, for the 1-year study period (Jun 2014–May 2015).

mostly positive over southern parts. Finally, the CSI of V04_E is higher in southern NW, central and southern CJ and western SE, but notably lower in NE, northern SW and TP compared with that of V03_E. The average CSI difference is almost zero, which could be attributed to the offset between better POD and worse FAR of V04_E. As for IMERG Late run products shown in Fig. 10(e)–(h), the spatial distribution characteristics of the four indices are almost the same as the Early run comparison. In general, IMERG V04 Early and Late run products are only more effective than V03 over CJ and SE, while over other regions, the new versions show similar or even worse performance than the old versions.

5. Summary and conclusion

In this study, the newly IMERG V05 Final run products and its previous versions, IMERG V03 and V04, are firstly intercompared over the globe at daily temporal resolution, also cross-evaluated against the globally reanalyzed benchmark GPCP 2.3 and MSWEP 2.1 with analysis on spatial distributions and latitudinal mean profiles. Then the Early,

Late and Final run products of IMERG V03, V04 and V05 are carefully evaluated over Mainland China against more than 30,000 gauges at hourly and daily temporal resolutions and $0.1^\circ \times 0.1^\circ$ spatial resolution. The conclusions of this study mainly contain:

- (1) IMERG V04_F and V05_F products show significant difference and improvements compared to V03_F over ocean. According to statistics, mean precipitation of V04_F and V05_F increases by +31.36% and +28.81% respectively compared with that of V03_F over ocean. In general, the significant increase of V04_F and V05_F over ocean is reasonable, as the average oceanic precipitation of V04_F and V05_F is ~ 3.10 mm/day and ~ 3.04 mm/day, which is much closer to that of the benchmark GPCP (~ 2.94 mm/day) and MSWEP (~ 2.93 mm/day) than that of V03_F (~ 2.36 mm/day). The ocean regions with much improved values are northeastern and mid-southern Pacific Ocean, northwestern and southwestern Atlantic Ocean and the Equatorial Zones. However, the difference over land is less significant.
- (2) At the national averaged scale over Mainland China, high CC that is

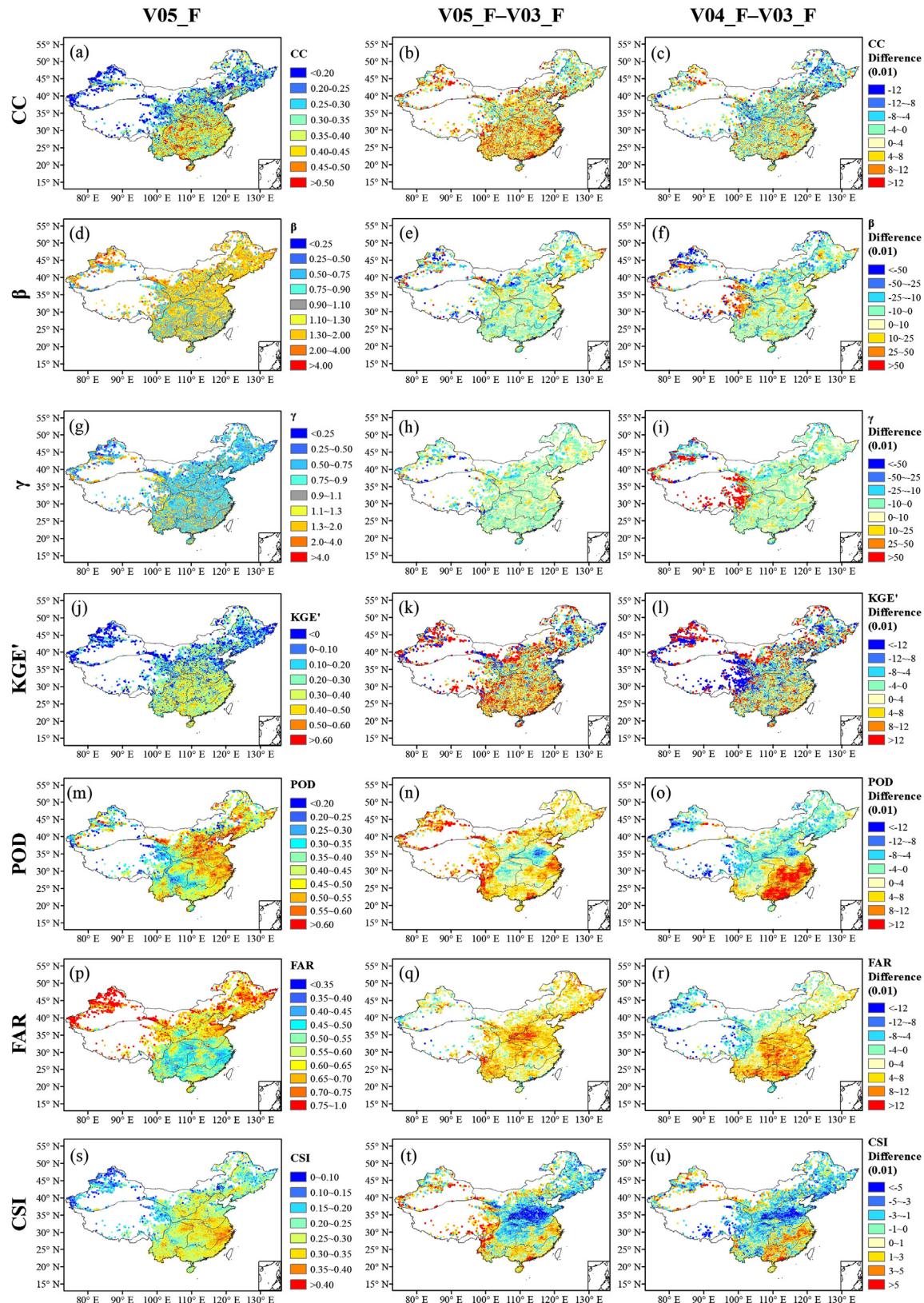


Fig. 9. Spatial distributions of CC (a-b), β (d-e), γ (g-h), KGE' (j-k), POD (m-n), FAR (p-q), CSI (s-u) for V05_F (the first column), the differences between V05_F and V03_F (the second column) and between V04_F and V03_F (the third column) at the hourly scale over Mainland China for the 1-year study period (Jun 2014–May 2015).

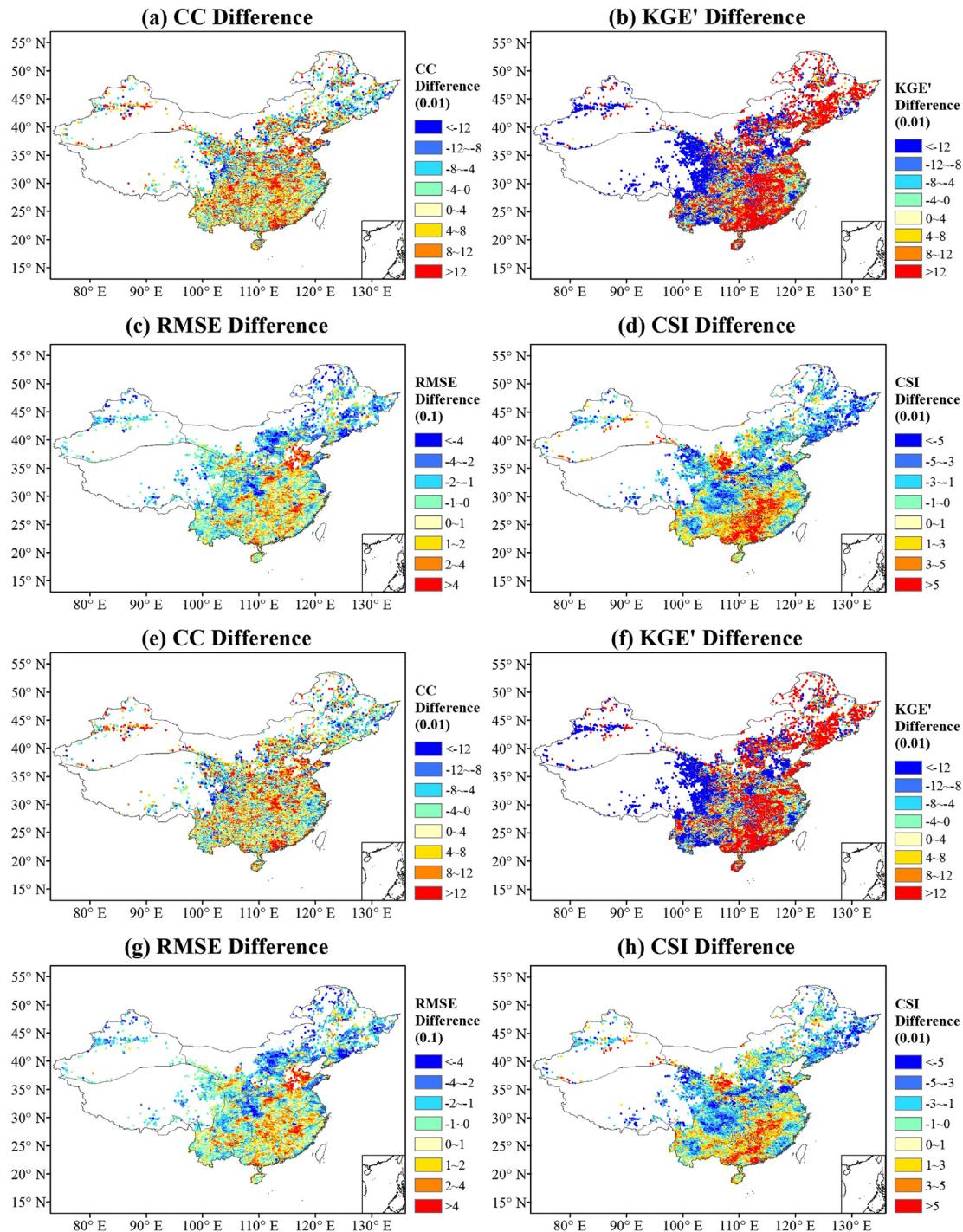


Fig. 10. Spatial distributions of the difference between IMERG V03_E and V04_E ($V_{04_E} - V_{03_E}$) ((a)-(d)) and between IMERG V03_L and V04_L ($V_{04_L} - V_{03_L}$) ((e)-(h)) for CC, KGE', RMSE and CSI successively at the hourly scale over China for the 1-year study period (Jun 2015–May 2016).

- more than 0.9 at the hourly scale and even larger at the daily scale indicates that all IMERG versions have the ability to represent the precipitation time series over large regions, even though all of them overestimate precipitation values with β of about 1.12 at all resolutions. The performance of V03_F, V04_F and V05_F are about the same at the national averaged scale.
- (3) At grid-based scale, V03_F, V04_F and V05_F all could basically describe the PDF structure of precipitation over Mainland China. In particular, V04_F and V05_F outperforms V03_F for zero or little precipitation (0–0.4 mm/h) at the hourly scale, while V03_F has better capacity for moderate and heavy precipitation both at hourly

($> 0.4 \text{ mm/h}$) and daily scales ($> 1 \text{ mm/day}$). Overall, IMERG V05_F precipitation products generally improves upon its last two predecessors, V04_F and V03_F, at the hourly scale across the whole country, which could demonstrate the advancement of the latest IMERG V05 algorithms, while V04_F only shows comparable performance with V03_F.

- (4) The quality of V03, V04 and V05 is heterogeneous in different areas of Mainland China. As a whole, the improvement from IMERG V03_F to V05_F is more concentrated in southeastern China (e.g., Lower Reaches of Changjiang River and the Pearl River) and western China (e.g., TP and XJ). Both V04_F and V05_F show worse

performance in precipitation occurrence identification in central and northern China than V03_F. In addition, compared with its processor V03_F, V04_F exhibits retrogressive performance in northern and western China which are subject to complex terrain or snow and ice during winter. When comparing all seven IMERG precipitation products, it is evident that V05_F gives a relatively better performance than other estimates in China, especially NW, NE, NC and CJ. Besides, V04 real-time estimates have better performance in most regions than V03 except XJ and TP, while the Early and Late run of the same version generally show similar performance.

In summary, IMERG V04 and V05 show significant improvements from their predecessors V03 basically over oceans. While over lands represented by Mainland China, V04 and V03 are about the same level while V05 is identified to be the best performed precipitation estimates. In addition, there should call on more localized studies to investigate into specific regions with ground gauge/radar combined networks as well as reanalysis. Further investigations should be also carried out to assess the underlying insights from IMERG Level-2 retrieval algorithms for error and uncertainty and how the uncertainty propagating to the IMERG Level-3 precipitation products.

Acknowledgements

This study was jointly supported by the National Key Research and Development Program of China (Grant No. 2016YFE0102400) and the National Natural Science Foundation of China (Grant No. 7146101701 and No. 91437214). The authors' great gratitude is extended to the China Meteorological Administration for providing the ground-based precipitation data. The IMERG data are provided by the NASA/Goddard Space Flight Center's PMM Science Team and Precipitation Processing System (PPS), which develop and compute the IMERG data as a contribution to GPM, and archived at the NASA Goddard Earth Science Data and Information System Center (GES DISC). While the GPCP and GPCC data are provided by the NOAA/OAR/ESRL PSD, Boulder, Colorado, USA. MSWEP is developed by Hylke Beck (Princeton University, Princeton, NJ, USA). We also thank Dr. Sheng Chen from Sun Yat-sen University for providing the shape file of China division.

References

- Adam, J., Lettemaier, D., 2003. Adjustment of global gridded precipitation for systematic bias. *J. Geophys. Res. Atmos.* 108 (D9).
- Adler, R.F., Huffman, G.J., Chang, A., Ferraro, R., Xie, P., Janowiak, J., Rudolf, B., Schneider, U., Curtis, S., Bolvin, D., Gruber, A., Susskind, J., Arkin, P., 2003. The Version 2 global precipitation climatology project (GPCP) monthly precipitation analysis (1979–Present). *J. Hydrometeorol.* 4, 1147–1167.
- Adler, R.F., Gu, G., Sapiano, M., Wang, J.J., Huffman, G.J., 2017. Global precipitation: means, variations and trends during the satellite era (1979–2014). *Surv Geophys.* 38 (4), 679–699.
- Adler, R., Sapiano, M., Huffman, G., Wang, J., Gu, G., Bolvin, D., Chiu, L., Schneider, U., Becker, A., Nelkin, E., Xie, P., Ferraro, R., Shin, D., 2018. The global precipitation climatology project (GPCP) monthly analysis (New Version 2.3) and a review of 2017 global precipitation. *Atmos* 9 (4), 138.
- 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.
- Beck, H., van Dijk, A., Levizzani, V., Schellekens, J., Miralles, D., Martens, B., de Roo, A., 2017a. MSWEP: 3-hourly 0.25° global gridded precipitation (1979–2015) by merging gauge, satellite, and reanalysis data. *Hydrol. Earth Syst. Sci.* 21 (1), 589–615.
- Beck, H., Vergopolan, N., Pan, M., Levizzani, V., van Dijk, A., Weedon, G., Brocca, L., Pappenberger, F., Huffman, G., Wood, E., 2017b. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.* 21 (12), 6201–6217.
- Behrangi, A., Stephens, G., Adler, R.F., Huffman, G.J., Lambrightsen, B., Lebsack, M., 2014. An update on the oceanic precipitation rate and its zonal distribution in light of advanced observations from space. *J. Clim.* 27 (11), 3957–3965.
- Chen, F., Li, X., 2016. Evaluation of IMERG and TRMM 3B43 monthly precipitation products over mainland China. *Remote Sens.* 8 (6), 472.
- Chen, S., et al., 2013. Similarity and difference of the two successive V6 and V7 TRMM multisatellite precipitation analysis performance over China. *J. Geophys. Res. Atmos.* 118, 13060–13074.
- Cohen Liechti, T., Matos, J., Boillat, J., Schleiss, A., 2012. Comparison and evaluation of satellite derived precipitation products for hydrological modeling of the Zambezi River Basin. *Hydrol. Earth Syst. Sci.* 16 (2), 489–500.
- Demaria, E., Rodriguez, D., Ebert, E., Salio, P., Su, F., Valdes, J., 2011. Evaluation of mesoscale convective systems in South America using multiple satellite products and an object-based approach. *J. Geophys. Res. Atmos.* 116 (D8).
- Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S.J., Ropelewski, C.F., 2007. Validation of satellite rainfall products over East Africa's complex topography. *Int. J. Remote Sens.* 28, 1503–1526.
- Gaona, M.R., Overeem, A., Leijnse, H., Uijlenhoet, R., 2016. First-year evaluation of GPM rainfall over the Netherlands: IMERG day 1 final run (V03D). *J. Hydrometeorol.* 17 (11), 2799–2814.
- Guo, H., Chen, S., Bao, A., Behrangi, A., Hong, Y., Ndayisaba, F., et al., 2016. Early assessment of integrated multi-satellite retrievals for global precipitation measurement over China. *Atmos. Res.* 176, 121–133.
- Guo, Z., Liu, X., Xiao, W., Wang, J., Meng, C., 2007. Regionalization and integrated assessment of climate resource in China based on GIS (in Chinese). *Resour. Sci.* 29, 2–9.
- Hendon, H.H., Lim, E.P., Nguyen, H., 2014. Seasonal variations of subtropical precipitation associated with the southern annular mode. *J. Clim.* 27 (9), 3446–3460.
- Hong, Y., Chen, S., Xue, X., Hodges, G., 2012. Global precipitation estimation and applications. In: *Multiscale Hydrologic Remote Sensing: Perspectives and Applications*. CRC Press, pp. 371–386.
- Hong, Y., Hsu, K., Sorooshian, S., Gao, X., 2004. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteorol.* 43, 1834–1853.
- Hou, A.Y., et al., 2014. The global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* 95, 701–722.
- Huffman, G., Adler, R., Arkin, P., Chang, A., Ferraro, R., Gruber, A., Janowiak, J., McNab, A., Rudolf, B., Schneider, U., 1997. The global precipitation climatology project (GPCP) combined precipitation dataset. *Bull. Amer. Meteorol. Soc.* 78 (1), 5–20.
- Huffman, G.J., Adler, R.F., Bolvin, D.T., Gu, G.J., Nelkin, E.J., Bowman, K.P., Hong, Y., Stocker, E.F., Wolff, D.B., 2007. The TRMM multi-satellite precipitation analysis (TMPA): quasi-global, multi-year, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* 8, 38–55.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., 2017a. Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation. NASA/GSFC, pp. 54.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., 2017b. Integrated Multi-satellitE Retrievals for GPM (IMERG) Technical Documentation. NASA/GSFC, pp. 59.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Stocker, E.F., 2017c. V04 IMERG Final Run Release Notes. NASA/GSFC, pp. 2.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Stocker, E.F., 2017d. V05 IMERG Early and Late Run Release Notes. NASA/GSFC, pp. 2.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., Stocker, E.F., 2017e. V05 IMERG Final Run Release Notes. NASA/GSFC, pp. 8.
- Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J., Xie, P., 2017f. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) Version 4.6. PPS. NASA/GSFC, pp. 32.
- Huffman, G.J., Bolvin, D.T., Braithwaite, D., Hsu, K., Joyce, R., Kidd, C., Nelkin, E.J., Sorooshian, S., Tan, J., Xie, P., 2017g. NASA Global Precipitation Measurement (GPM) Integrated Multi-Satellite Retrievals for GPM (IMERG) Algorithm Theoretical Basis Document (ATBD) Version 5.1. PPS. NASA/GSFC, pp. 34.
- Huffman, G.J., Bolvin, D.T., Nelkin, E.J., 2017h. Integrated Multi-satellitE Retrievals for GPM (IMERG) technical documentation, version 5.1. NASA's Precipitation Processing Center. <<https://pmm.nasa.gov/data-access/downloads/gpm>> (accessed 17.11.2017).
- Joyce, R.J., Janowiak, J.E., Arkin, P.A., Xie, P., 2004. CMORPH: a method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.* 5, 487–503.
- Kidd, C., Huffman, G., 2011. Global precipitation measurement. *Meteorol. Appl.* 18, 334–353.
- Kirstetter, P.-E., Hong, Y., Gourley, J.J., Schwaller, M., Petersen, W., Zhang, J., 2013. Comparison of TRMM 2A25 Products, Version 6 and Version 7, with NOAA/NSSL ground radar-based national mosaic QPE. *J. Hydrometeorol.* 14, 661–669.
- Kling, H., Fuchs, M., Paulin, M., 2012. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* 424–425, 264–277.
- Kubota, T., et al., 2007. Global precipitation map using satellite-borne microwave radiometers by the GSMAp Project: production and validation. *IEEE Trans. Geosci. Remote Sens.* 45, 2259–2275.
- Li, N., Tang, G., Zhao, P., Hong, Y., Gou, Y., Yang, K., 2017. Statistical assessment and hydrological utility of the latest multi-satellite precipitation analysis IMERG in Ganjiang River basin. *Atmos. Res.* 183, 212–223.
- Liebmann, B., Allured, D., 2005. Daily precipitation grids for South America. *Bull. Am. Meteorol. Soc.* 86 (11), 1567–1570.
- Ma, Y., Tang, G., Long, D., Yong, B., Zhong, L., Wan, W., Hong, Y., 2016. Similarity and error intercomparison of the GPM and its predecessor-TRMM Multisatellite Precipitation Analysis using the best available hourly gauge network over the Tibetan Plateau. *Remote Sens.* 8 (7), 569.
- Ma, Y., Yang, Y., Han, Z., Tang, G., Maguire, L., Chu, Z., Hong, Y., 2018. Comprehensive evaluation of ensemble multi-satellite precipitation dataset using the dynamic bayesian model averaging scheme over the Tibetan plateau. *J. Hydrol.* 556, 634–644.
- Ma, Z., Zhou, Y., Hu, B., Liang, Z., Shi, Z., 2017. Downscaling annual precipitation with TMPA and land surface characteristics in China. *Int. J. Climatol.* 15.
- Mayor, Y.G., Tereshchenko, I., Fonseca-Hernández, M., Pantoja, D.A., Montes, J.M., 2017. Evaluation of error in IMERG precipitation estimates under different topographic

- conditions and temporal scales over Mexico. *Remote Sens.* 9 (5), 503.
- Qian, W., Lin, X., 2005. Regional trends in recent precipitation indices in China. *Meteorol. Atmos. Phys.* 90 (3), 193–207.
- Salio, P., Hobouchean, M., García Skabar, Y., Vila, D., 2015. Evaluation of high-resolution satellite precipitation estimates over southern South America using a dense rain gauge network. *Atmos. Res.* 163, 146–161.
- Schneider, U., Becker, A., Finger, P., Meyer-Christoffer, A., Rudolf, B., Ziese, M., 2011. GPCC monitoring product: Near real-time monthly land-surface precipitation from rain-gauges based on SYNOP and CLIMAT data. Global Precipitation Climatology Centre (GPCC) 10.5676/DWD_GPCC/MP_M_V4_100.
- Shen, Y., Pan, Y., Yu, J., Zhao, P., Zhou, Z., 2013. Quality assessment of hourly merged precipitation product over China. *Trans. Atmos. Sci.* 36 (1), 37–46 (in Chinese).
- Shen, Y., Xiong, A., Wang, Y., Xie, P., 2010. Performance of high resolution satellite precipitation products over China. *J. Geophys. Res. Atmos.* 115.
- Shen, Y., Zhao, P., Pan, Y., Yu, J., 2014. A high spatiotemporal gauge-satellite merged precipitation analysis over China. *J. Geophys. Res. Atmos.* 119 (6), 3063–3075.
- Sorooshian, S., Hsu, K.-L., Gao, X., Gupta, H.V., Imam, B., Braithwaite, D., 2000. Evaluation of PERSIANN system satellite-based estimates of tropical rainfall. *Bull. Am. Meteorol. Soc.* 81, 2035–2046.
- Su, X., Shum, C., Luo, Z., 2018. Evaluating IMERG V04 final run for monitoring three heavy rain events over Mainland China in 2016. *IEEE Geosci. and Res. Lett.* 15 (3), 444–448.
- Tan, J., Petersen, W.A., Kirstetter, P.E., Tian, Y., 2017. Performance of IMERG as a function of spatiotemporal scale. *J. Hydrometeorol.* 18 (2), 307–319.
- 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.
- Tang, G., Ma, Y., Long, D., Zhong, L., Hong, Y., 2016a. Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatio-temporal scales. *J. Hydrol.* 533, 152–167.
- Tang, G., Zeng, Z., Long, D., Guo, X., Yong, B., Zhang, W., Hong, Y., 2016b. Statistical and hydrological comparisons between TRMM and GPM level-3 products over a mid-latitude basin: is day-1 IMERG a good successor for TMPA 3B42V7? *J. Hydrometeorol.* 17 (1), 121–137.
- Tang, G., Behrangi, A., Long, D., Li, C., Hong, Y., 2018a. Accounting for spatiotemporal errors of gauges: a critical step to evaluate gridded precipitation products. *J. Hydrol.* 559, 294–306.
- Tang, G., Long, D., Hong, Y., Gao, J., Wan, W., 2018b. Documentation of multifactorial relationships between precipitation and topography of the Tibetan Plateau using spaceborne precipitation radars. *Remote Sens. Environ.* 208, 82–96.
- Tang, G., Wen, Y., Gao, J., Long, D., Ma, Y., Wan, W., Hong, Y., 2017. Similarities and differences between three coexisting spaceborne radars in global rainfall and snowfall estimation. *Water Resour. Res.* 53 (5), 3835–3853.
- Tang, Z., Wang, Z., Zheng, C., Fang, J., 2006. Biodiversity in China's mountains. *Front. Ecol. Environ.* 4 (7), 347–352.
- Tapiador, F.J., et al., 2012. Global precipitation measurement: methods, datasets and applications. *Atmos. Res.* 104, 70–97.
- Taylor, K.E., 2001. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res. Atmos.* 106, 7183–7192.
- Wei, G., Lü, H., Crow, W.T., Zhu, Y., Wang, J., Su, J., 2017. Evaluation of satellite-based precipitation products from IMERG V04A and V03D, CMORPH and TMPA with gauged rainfall in three climatologic Zones in China. *Remote Sens.* 10 (2), 30.
- Wu, X., Liang, X.Z., Zhang, G.J., 2003. Seasonal migration of ITCZ precipitation across the equator: why can't gcms simulate it? *Geophys. Res. Lett.* 30 (15), 1824.
- Xu, R., Tian, F., Yang, L., Hu, H., Lu, H., Hou, A., 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 (2), 910–924.
- Yang, D., Goodison, B., Metcalfe, J., Louie, P., Elomaa, E., Hanson, C., Golubev, V., Gunther, T., Milkovic, J., Lapin, M., 2001. Compatibility evaluation of national precipitation gage measurements. *J. Geophys. Res. Atmos.* 106 (D2), 1481–1491.
- Yong, B., Liu, D., Gourley, J.J., Tian, Y., Huffman, G.J., Ren, L., Hong, Y., 2015. Global view of real-time TRMM multisatellite precipitation analysis: implications for its successor global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* 96, 283–296.
- Zambrano-Bigiarini, M., Nauditt, A., Birkel, C., Verbist, K., Ribbe, L., 2017. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile. *Hydrol. Earth Syst. Sci.* 21 (2), 1295–1320.
- Zhao, H., Yang, S., You, S., Huang, Y., Wang, Q., Zhou, Q., 2017. Comprehensive evaluation of two successive V3 and V4 IMERG final run precipitation products over mainland China. *Remote Sens.* 10 (2), 34.