



An Improved Assessment Method and Its Application to the Latest IMERG Rainfall Product in Mainland China

Xinran Xia ¹, Disong Fu ² , Ye Fei ³, Wei Shao ¹ and Xiangao Xia ^{2,4,*} 

¹ School of Hydrology and Water Resources, Nanjing University of Information Science and Technology, Nanjing 210044, China; 20211236003@nuist.edu.cn (X.X.); Shao@nuist.edu.cn (W.S.)

² LAGEO, Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing 100029, China; fudisong@mail.iap.ac.cn

³ National Meteorological Information Center, Chinese Meteorological Administration, Beijing 100089, China; feiy@cma.gov.cn

⁴ National Institute of Natural Hazards, Ministry of Emergency Management of China, Beijing 100085, China

* Correspondence: xxa@mail.iap.ac.cn

Abstract: Quantification of uncertainties associated with satellite precipitation products is a prior requirement for their better applications in earth science studies. An improved scheme is developed in this study to decompose mean bias error (MBE) and mean square error (MSE) into three components, i.e., MBE and MSE associated hits, missed precipitation, and false alarms, respectively, which are weighted by their relative frequencies of occurrence (RFO). The trend of total MBE or MSE is then naturally decomposed into six components according to the chain rule for derivatives. Quantitative estimation of individual contributions to total MBE and MSE is finally derived. The method is applied to validation of Integrated Multisatellite Retrievals for GPM (IMERG) in Mainland China. MBE associated with false alarms is an important driver for total MBE, while MSE associated with hits accounts for more than 85% of MSE, except in inland semi-arid area. The RFO of false alarms increases, whereas the RFO of missed precipitation decreases. Both factors lead in part to a growing trend for total MBE. Detection of precipitation should be improved in the IMERG algorithm. More specifically, the priority should be to reduce false alarms.

Keywords: precipitation; IMERG; error decomposition; gauge rainfall rate; validation



Citation: Xia, X.; Fu, D.; Fei, Y.; Shao, W.; Xia, X. An Improved Assessment Method and Its Application to the Latest IMERG Rainfall Product in Mainland China. *Remote Sens.* **2021**, *13*, 5107. <https://doi.org/10.3390/rs13245107>

Academic Editors: Marios Anagnostou, Petros Katsafados, Yagmur Derin and John Kalogiros

Received: 2 November 2021

Accepted: 13 December 2021

Published: 16 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Satellite observation is the most important measure to retrieve precipitation at a global scale, which provides a growing legacy of high-quality precipitation data for earth system sciences [1–3]. In particular, the launches of the Tropical Rainfall Measurement Mission satellite (TRMM) in 1997 and the Global Precipitation Measurement (GPM) core satellite in 2004 are two milestones in this aspect. Precipitation Radar (PR) onboard TRMM is the first spaceborne instrument to actively measure precipitation profiles. The dual-frequency precipitation radar (DPR) on the GPM core satellite improves the detection capability of light rainfall (0.2 mm h^{-1}) and snowfall. It also expands coverage over latitude band 65°N – 65°S [4]. These active instruments provide not only new insights into precipitation but also a vital source for calibrating retrievals from other precipitation sensors, e.g., passive microwave (PMW) sensors. PMW is sensitive to atmospheric column precipitation particles, whereas infrared (IR) is only sensitive to the cloud top layer [5]. More reliable rainfall estimations are then expected from PMW relative to IR. The TRMM microwave imager (TMI) and GPM microwave imager (GPI) provide a good reference for the cross-calibration of PMW imagers and sounders onboard other platforms, which spans the temporal and spatial coverage of PMW precipitation retrievals [5,6]. PMW observation has wider space coverage relative to PR and DPR. Specifically, a synergy of multiple PMW sensors on board different polar-orbiting satellites can effectively improve the detection efficiency,

although it is less than that of IR radiometers onboard geostationary satellites. Most global precipitation products are mainly produced by PMW measurements but utilize IR data for spatiotemporal completeness, such as TMPA (TRMM Multisatellite Precipitation Analysis) [7] and IMERG (Integrated Multisatellite Retrievals for GPM) [5].

There are dozens of high-resolution satellite precipitation products, among which IMERG is the latest. The highest temporal and spatial resolutions of IMERG are half an hour and 10 km, respectively [6]. The IMERG algorithm incorporated multiyear experience from international partners in its PMW calibration technique [8], IR [9] and PMW [10] precipitation retrievals, as well as PMW and IR morphing, and the Kalman filter algorithm [11,12]. It has been widely used to improve the accuracy of climate modeling in terms of extreme rain and snowfall to strengthen the applications for current and future disasters, disease, resource management, energy production, and food security [6].

Precipitation is a complex phenomenon showing dramatic temporal and spatial variations, which makes it a challenging task to estimate it accurately. Mathematically, the estimation is an ill-posed inversion problem. Satellite precipitation products are unavoidably associated with substantial large systematic and random errors, which vary with regions, climate, and precipitation types [10]. Therefore, it is necessary to evaluate the retrievals at multitemporal and spatial scales. This can not only provide suggestions for data users to select appropriate products, but also on how to apply and consider the impact of data quality in actual applications. More importantly, it can recommend the direction to improve the retrievals.

Many validation studies have been reported, which concern verification methods and demonstrate the performance of multiple satellite precipitation products at different temporal and spatial scales [13–18]. It is widely suggested that IMERG V06 outperforms other satellite rainfall products as a result of several major improvements [17]. The performance of the IMERG rainfall product depends on climate and topography. Underestimation of rainfall by IMERG was found in subtropical regions; for example, in the Indian subcontinent and Ganges Brahmaputra Meghna basin [19,20]. Although IMERG outperforms the reanalysis rainfall products in many aspects, the triple collocation analysis implies that IMERG snowfall needs further development [17], which may account for the fact poor performance of IMERG over north China [21]. On the contrary, good performance is generally found in the warm season, although systematic and random biases should still be adjusted before driving hydrological models [22–24]. Satellite-based precipitation retrieval techniques perform poorly in complex terrain, because low-level, topographically forced updrafts may trigger heavy rainfall that shows great spatiotemporal variability [23], which is also not captured by poor coverage of gauge measurements.

Uncertainties mainly come from three distinct situations, namely, hit (H): both gauged and satellite rainfall exceed a threshold; missed (M): satellite rainfall is below the threshold while gauged rainfall passes the threshold; false alarm (F): satellite rainfall exceeds threshold but gauged rainfall does not. Uncertainty associated with corrective negatives (C): both satellite and gauged rainfall are below the threshold, or are negligible. Knowledge of dominant contribution to total uncertainty is essential for the algorithm developers. Tian et al. (2009) proposed a novel error decomposition scheme of mean bias error (MBE), pointing out that evaluating total MBE may lead to misunderstanding. This is because missed precipitation is, to a certain extent, offset by false alarms, likely resulting in a negligible overall error. Assessment of long-term performance of satellite precipitation products is vital for climate study, which is addressed by a very few precious studies, for example, Tang et al. (2020) [17].

The purpose of this study is to further the decomposition scheme but focusing on the decomposition of MBE and mean square error (MSE) simultaneously. The outstanding feature of this improved method is to produce quantitative relationships between total MBE/MSE and their individual contributions. More importantly, long-term changes in MBE (MSE) associated with H, M, and F, as well as their relative frequencies of occurrence (RFO), are investigated. The contributions of these conditional errors to the unconditional

errors are quantified and compared. The method is applied to the IMERG assessment in Mainland China. Interesting spatial and temporal distributions of MBE and MSE decomposition are presented. Spatial and temporal variability of IMERG performance in mainland China is presented. Potential improvements in IMERG during the past two decades are discussed.

2. Data and Method

2.1. Data

The latest IMERG precipitation product is derived by the V06 algorithm, which includes several major improvements. Previous studies showed that the IMERG product exhibited better performance than other satellite precipitation datasets, which were mostly based on short-term observations (after 2014 when the GPM Core Observatory was launched) [17]. Here we evaluate daily IMERG V06 rainfall covering the period beginning from June 2000 since the retrospective processing of PMW data before GPM Core Observatory was finalized. Daily rainfall, the accumulated rainfall from 00:00 to 24:00 UTC with half-hourly resolution, is validated.

We use daily rain-gauge data from 2395 stations over Mainland China as the reference. Note that monthly rainfall data at some of these stations are used as the reference to calibrate the IMERG V06 product. Therefore, strictly speaking, this is not an independent validation. The gauged data are quality controlled by the Chinese Meteorological Administration, which include gross error checks, time stamp checks, etc. The data are widely used for the validation studies [17,19–21,23]. Given that the ultimate goal of our research is to drive our hydrological model by using the IMERG product after we carefully evaluate its performance, the validation is made in nine watersheds, separately. Figure 1 shows the gauge map with different colors in these watersheds. The Songliao watershed (SL) is located in northeastern China, where the climate is cold-zone—dry in winter and warm in summer. The Haihe watershed (HH) covers the north China plain, which has a subtropical monsoon climate. Yellow River watershed (YR) has a semi-arid climate. The climate in the inland watershed (IL) is arid-zone cold desert according to the Koppen–Geiger climate classification map. The climate in the Huaihe watershed (HU) is typically a temperate, warm summer and dry winter. The climate in the western part of the Yangtze river watershed (YZ) is temperate-zone dry in the warm season, whereas there is a temperate-zone warm summer and no dry season in the eastern part. The climate in the southeast watershed (SE) and the Pearl delta watershed (PD) is typically temperate-zone warm summer and no dry season. Temperate-zone dry, warm summer is typical in the southwestern watershed (SW).

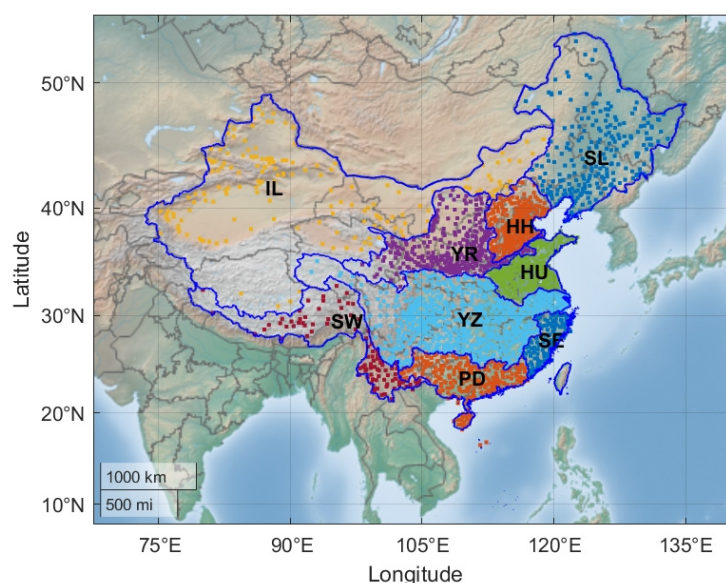


Figure 1. Spatial distribution of gauge stations in Mainland China, assigned into nine catchments.

2.2. Statistical Metrics

We used a simple collocation method to combine the gauge and IMERG grid in the validation. Measurements at gauges with a IMERG grid are averaged to compare IMERG pixel rainfall. Given that daily IMERG is calculated from half-hourly retrievals from 00:00 to 24:00 UTC, we used the daily gauged measurements from 08:00 to next 08:00 Beijing Time that agrees with that of IMERG, since UTC is 8 h behind Beijing Time. In order to evaluate the IMERG performance, the statistical merits were calculated based on the collocated twenty years (2001–2020) of gauged and IMERG rainfall products. The metrics include the Pearson correlation coefficient (CC) to describe the covariance between IMERG and gauged data. The mean bias error (MBE) and root mean square error (RMSE) are used to describe the systematic and random error characteristics of IMERG. We also used the Kling–Gupta efficiency [24,25] that combines the contributions of correlation, bias, and variability term.

$$KGE = 1 - \sqrt{(CC - 1)^2 + (\beta - 1)^2 + (\gamma - 1)^2} \quad (1)$$

where β is the bias ratio, i.e., the ratio of the mean of IMERG to that of the gauged data; γ is the ratio of the coefficient of variation between IMERG and gauged data.

Since missing and false alarm by the IMERG are two critical metrics demonstrating its performance. The probability of detection (POD) and false alarm ratio (FAR) are also calculated based on the number of hit events (H), the number of missed events (M), and the number of false alarms (F). Finally, the critical success index ($CSI = H/(H + M + F)$) is also calculated to demonstrate the capability of IMERG to detect precipitation occurrence.

2.3. Decomposition Scheme

The unconditional (total) MBE_T is calculated from the equation below.

$$MBE_T = \frac{1}{N_T} \sum_{i=1}^{N_T} (Y_i - X_i), \quad (2)$$

where Y_i and X_i represent daily satellite and gauged rainfall rate (mm d^{-1}), respectively. N_T represents annual total collocated samples of satellite and gauged rainfall rate. It is natural to decompose MBE_T into three items as follows.

$$MBE_T = \frac{1}{N_T} \sum_{h=1}^{N_H} (Y_h - X_h) + \frac{1}{N_T} \sum_{m=1}^{N_M} (Y_m - X_m) + \frac{1}{N_T} \sum_{f=1}^{N_F} (Y_f - X_f), \quad (3)$$

where N_H , N_M and N_F represent annual events associated with H, M, and F, respectively. The three items in the bracket of the right-hand side of Equation (3) represent the sum of individual MBEs associated with H, M, and F, from left to right, respectively. Note that marginal biases associated with corrective negatives are neglected. The equation can intuitively be transformed to Equation (4).

$$MBE_T = \frac{N_H}{N_T} MBE_H + \frac{N_M}{N_T} MBE_M + \frac{N_F}{N_T} MBE_F, \quad (4)$$

where MBE_H , MBE_M and MBE_F represent annual mean MBEs associated with H, M, and F. The ratios of N_H , N_M and N_F to N_T mean RFO associated with H, M and F, respectively. MBE_T is actually the weighted average of mean MBEs associated with H, M, and F. The weights are their RFOs. Since all items in Equation (4) vary with time, we can calculate the derivatives using the chain rule and the following equation is then derived.

$$\frac{\partial}{\partial t} (MBE_T) = \sum_{I=H,M,F} \left\{ MBE_I \frac{\partial}{\partial t} \left(\frac{N_I}{N_T} \right) + \frac{N_I}{N_T} \frac{\partial}{\partial t} (MBE_I) \right\}, \quad (5)$$

Linear trends of all items in Equation (5) are calculated by the least-square regression. This suggests that the MBE_T trend is attributable not only to changes in annual MBEs

associated with H, M, and F, but also changes in their RFOs. It is straightforward to discriminate between the contributions of individual components to the MBE_T trend using Equation (5). The decomposition scheme is also applicable to unconditional MSE (MSE_T). The investigation of the error sources is then achieved by this decomposition, the results derived from which are presented in the following section. The threshold for the occurrence of rainfall is 0.5 mm d^{-1} , which accounts for the measurement uncertainties of satellite and gauged rainfall products. Note that the method and findings here are weakly dependent on the threshold.

3. Results

Figure 2 presents spatial distribution of four statistical metrics describing the precipitation detection capacity of the IMERG product. POD in the IL watershed is generally smaller than that in other watersheds. Another interesting feature is that POD in the Yungui Plateau is relatively smaller than the surrounding regions, which likely implies that PMW sensor nighttime precipitation should be improved in the algorithm. In addition, this is also likely to be associated with large uncertainties of satellite remote-sensing precipitation over the complex terrain [26–28]. Furthermore, FAR in the IL watershed is larger than that in other watersheds, which suggests that much effort should be paid to the improvement of precipitation retrieval over semi-arid regions [23]. Table 1 presents the statistical metrics derived from seasonal pairs of IMERG and gauged data in nine catchments. Several features merit mention. First, the performance of IMERG shows remarkable spatial and temporal variability. Second, seasonal mean IMERG rainfall tends to be smaller than gauged measurements, especially in the summer rainy season, which likely indicates that IMERG underestimates convective rainfall. Third, precipitation occurrence detection described by POD, FAR, and CSI are highly dependent on rainfall amount. Precipitation occurrence detection is better in rainy seasons or humid regions than that in dry seasons or semi-arid regions. On the contrary, the KGE that combines the contributions of correlation, bias, and variability term shows smaller variation between seasons and regions than POD, FAR, and CSI.

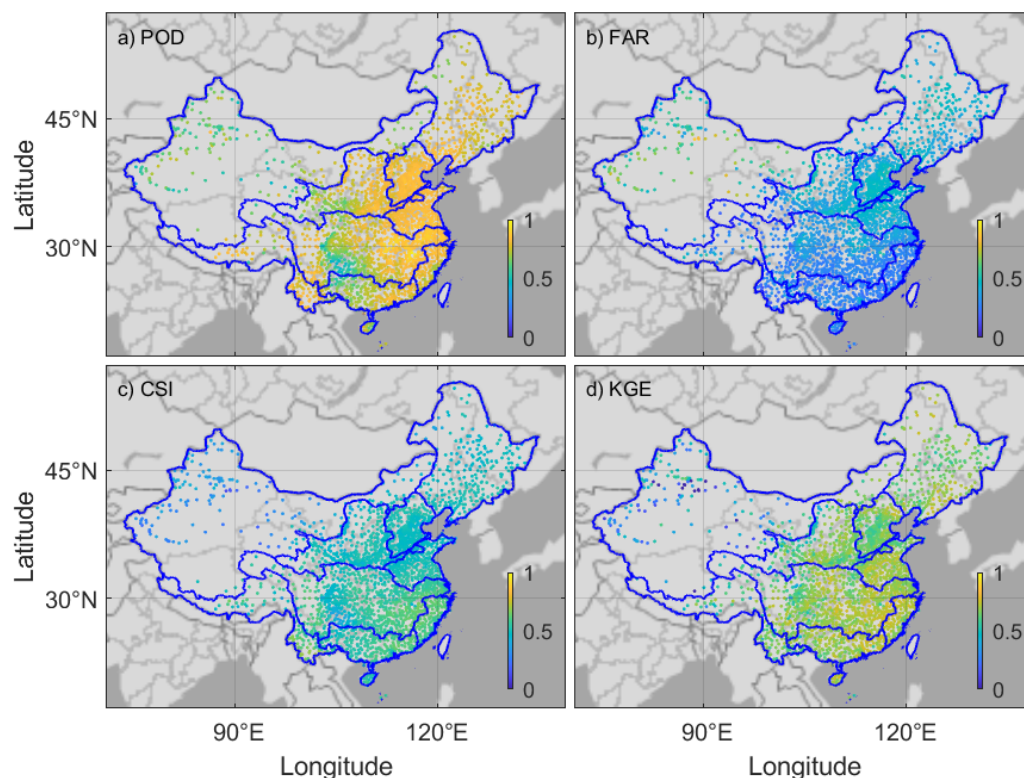


Figure 2. Spatial distribution of POD (a), FAR (b), CSI (c), and KGE (d).

Table 1. Statistical metrics derived from seasonal pairs of IMERG and gauged data in nine catchments. The metrics include the average daily gauged rainfall (Mean_G), average daily IMERG rainfall (Mean_I), correlation coefficient (CC), root mean square error (RMSE), percentage of detection (POD), false alarm ratio (FAR), the critical success index (CSI), and the Kling–Gupta efficiency (KGE).

Catchment	Season	Mean_G	Mean_I	RMSE	CC	POD	FAR	CSI	KGE
SL	Spring	1.23	1.06	2.67	0.75	0.54	0.24	0.46	0.67
	Summer	4.31	3.90	7.34	0.72	0.65	0.11	0.60	0.64
	Autumn	1.26	1.06	3.05	0.73	0.54	0.22	0.47	0.64
	Winter	0.27	0.21	1.18	0.40	0.26	0.66	0.17	0.35
HH	Spring	0.94	0.84	2.46	0.77	0.52	0.19	0.46	0.64
	Summer	3.90	3.70	8.43	0.69	0.61	0.13	0.56	0.61
	Autumn	1.24	1.11	3.36	0.72	0.52	0.17	0.47	0.59
	Winter	0.18	0.16	0.81	0.69	0.37	0.52	0.26	0.65
IL	Spring	0.42	0.47	1.73	0.55	0.48	0.40	0.36	0.51
	Summer	1.20	1.11	3.05	0.60	0.55	0.19	0.49	0.53
	Autumn	0.47	0.44	1.62	0.63	0.47	0.34	0.38	0.60
	Winter	0.11	0.17	1.36	0.14	0.29	0.80	0.14	−0.10
YR	Spring	1.01	0.95	2.56	0.72	0.55	0.26	0.46	0.68
	Summer	3.19	2.95	6.56	0.68	0.64	0.16	0.57	0.63
	Autumn	1.53	1.41	3.36	0.73	0.60	0.22	0.51	0.70
	Winter	0.19	0.19	0.90	0.59	0.35	0.61	0.23	0.59
HU	Spring	1.81	1.61	3.90	0.78	0.53	0.14	0.49	0.66
	Summer	5.75	5.34	10.96	0.73	0.63	0.14	0.57	0.68
	Autumn	1.94	1.77	4.80	0.73	0.55	0.16	0.50	0.63
	Winter	0.84	0.71	2.18	0.73	0.48	0.30	0.40	0.69
YZ	Spring	3.93	3.66	6.77	0.73	0.68	0.21	0.58	0.72
	Summer	6.21	5.83	10.74	0.71	0.68	0.14	0.61	0.68
	Autumn	2.75	2.59	5.78	0.73	0.67	0.28	0.53	0.72
	Winter	1.40	1.33	4.43	0.63	0.61	0.44	0.41	0.62
SW	Spring	1.98	1.91	4.74	0.67	0.64	0.28	0.51	0.66
	Summer	6.49	5.80	9.04	0.62	0.77	0.13	0.69	0.60
	Autumn	2.45	2.31	5.52	0.67	0.68	0.27	0.54	0.66
	Winter	0.47	0.53	2.11	0.79	0.56	0.54	0.34	0.71
SE	Spring	5.90	5.50	8.67	0.73	0.76	0.15	0.67	0.72
	Summer	7.44	7.25	12.60	0.72	0.68	0.13	0.62	0.68
	Autumn	3.26	3.15	7.92	0.74	0.69	0.28	0.55	0.73
	Winter	2.53	2.40	5.55	0.77	0.76	0.38	0.52	0.67
PD	Spring	5.05	4.81	10.09	0.70	0.72	0.27	0.57	0.70
	Summer	8.60	8.28	14.01	0.71	0.73	0.14	0.65	0.70
	Autumn	3.53	3.37	9.02	0.74	0.67	0.29	0.52	0.74
	Winter	1.35	1.43	4.47	0.74	0.65	0.57	0.35	0.67

Spatial distribution of MBE_T and weighted MBEs of H, M, and F are shown in Figure 3. Table 2 presents their spatial mean biases and one standard deviations in nine watersheds.

MBE_T (Figure 3a) is generally not further away from zero at most stations in northern China (SL, HU, IL, YR, and HU), while in southern China (YZ, SW, SE, and PD), a few stations evidence relatively large MBE_T exceeding 1.5 mm d^{-1} . Regional mean MBE_T are within 0.10 to approximately 0.23 mm d^{-1} in all watersheds except in IL (inland semi-arid area), where it is close to zero (Table 2). Positive MBE_T values imply that IMERG tends to overestimate precipitation in general from the statistical view. Weighted MBE_H (by its RFO) is generally smaller than MBE_T , while a similar spatial distribution is observed (correlation coefficient between them is 0.88). Small weighted MBE_H values are derived in northern China, varying from -0.05 to 0 mm d^{-1} ; while in southern China, weighted MBE_H is relatively large and occasionally comparable to MBE_T , indicating a considerable contribution to MBE_T . For example, weighted MBE_H in PD reaches 0.13, which accounts for 81% of MBE_T . Weighted MBE_F is an important contributor to MBE_T in all watersheds (Table 2), while it is partly offset by weighted MBE_M values that is relatively smaller in magnitude. The reason for larger weighted MBE_F is mostly due to more occurrence of false alarms than missed precipitations. This is more evident in the SL watershed where the RFO associated with missed precipitation are below 5%, while it exceeds 8% for false alarms. Poor precipitation occurrence detection there is partly because detection of snowfall by the PMW remote sensing is still a big challenge. Therefore, it is essential to improve the detection accuracy of IMERG, especially to reduce the probability of false alarms.

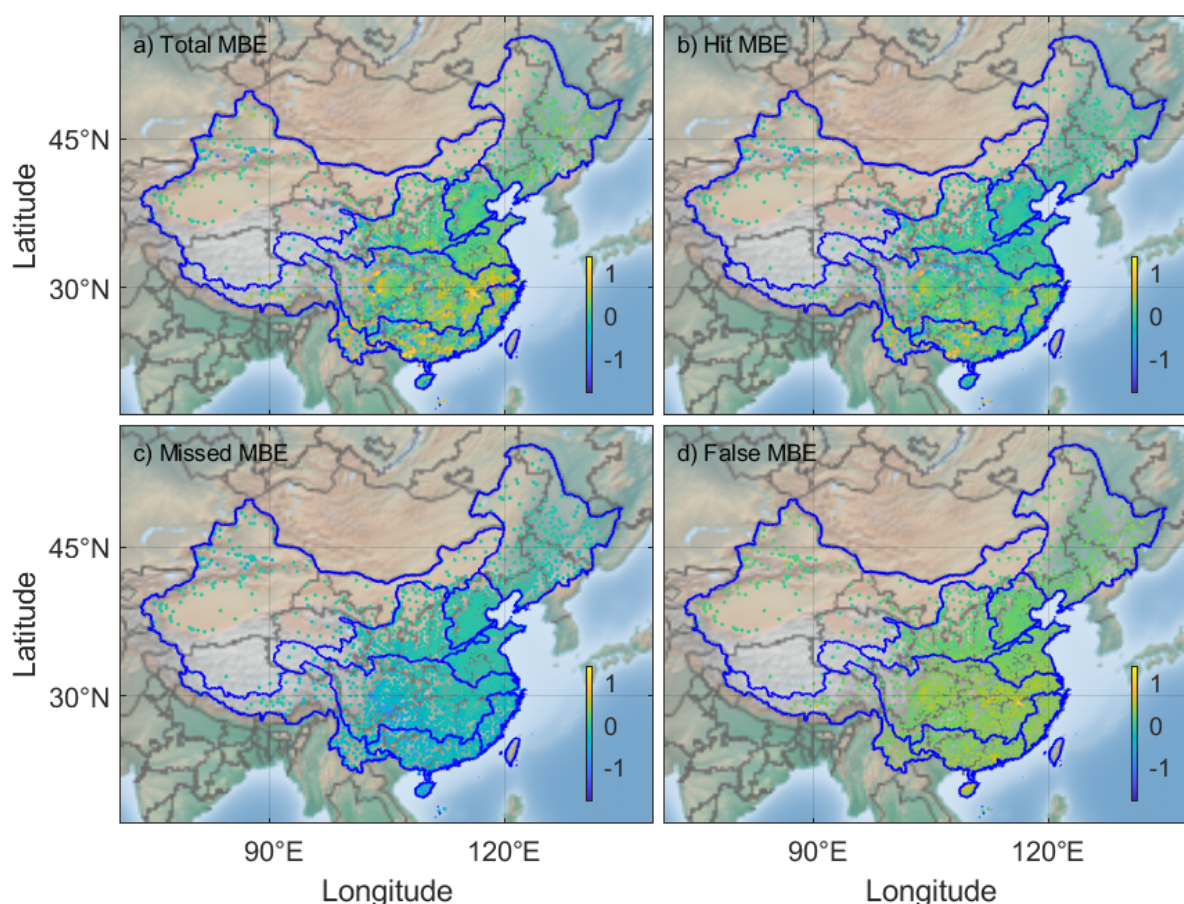


Figure 3. Spatial distribution of the total MBE (a) and its three components, i.e., hit (b), missed (c), and false MBE (d) in China. Note that MBE components are weighted by their relative occurrence frequencies, which makes the sum of MBEs associated with hit, missed, and false alarms equal to the total MBE.

Table 2. The regional average and one standard deviation of total MBE (MBE_T) and its three components (H, M, and F) weighted by their relative occurrence frequency in nine catchments. Note that MBE_T is equal to the sum of its three components.

Catchment	MBE_T (mm d ⁻¹)	MBE_H (mm d ⁻¹)	MBE_M (mm d ⁻¹)	MBE_F (mm d ⁻¹)
SL	0.21 ± 0.13	0.03 ± 0.09	−0.10 ± 0.04	0.26 ± 0.06
HH	0.12 ± 0.11	−0.05 ± 0.09	−0.09 ± 0.02	0.24 ± 0.03
IL	0.00 ± 0.23	−0.05 ± 0.15	−0.10 ± 0.08	0.14 ± 0.07
YR	0.10 ± 0.16	−0.01 ± 0.11	−0.12 ± 0.06	0.22 ± 0.07
HU	0.23 ± 0.16	0.00 ± 0.13	−0.12 ± 0.03	0.32 ± 0.05
YZ	0.22 ± 0.35	0.10 ± 0.25	−0.25 ± 0.13	0.36 ± 0.09
SW	0.22 ± 0.53	0.13 ± 0.39	−0.22 ± 0.11	0.31 ± 0.12
SE	0.21 ± 0.38	0.14 ± 0.29	−0.30 ± 0.10	0.37 ± 0.04
PD	0.16 ± 0.47	0.13 ± 0.37	−0.35 ± 0.11	0.38 ± 0.10

The spatial distribution of MSE_T (Figure 4a) generally follows that of a rainfall rate that decreases from north to south and from west to east. As shown in Table 3, the minimum regional mean MSE_T is observed in IL, being 4.4 mm d⁻¹, which is smaller by a factor of 100 than the maximum in PD (104.9 mm d⁻¹). The reason for this phenomenon is that the precipitation intensity and RFO in PD, the humid region, are far larger than that in IL, which is located in the semi-arid region. The decomposition of MSE_T suggests that it is overwhelmingly contributed to (>85%) by weighted MSE_H in all watersheds, even in IL where the smallest weighted MSE_H can even account for 68% of MSE_T . Weighted MSE_M is also smaller than weighted MSE_F , which is mostly because of larger RFO associated with false alarms relative to missed precipitation. Furthermore, occasional false alarms of relatively large IMERG rainfall rates also contribute to a larger MSE_F . This feature is more outstanding in HU and HH where MSE_F is more than twice larger than MSE_M . Weighted MSE_M and MSE_F in IL contributed to MSE_T by 14% and 18%, respectively, indicating that there is much room to improve IMERG rainfall detection in this inland semi-arid region. Specifically, detection of light rainfall events should be improved and false alarms of relatively larger rainfall events should be avoided as far as possible.

Table 3. The regional average and one standard deviation of total MSE (MSE_T) and the percentages accounted for by MSE associated with H, M, and F in nine catchments.

Catchment	MSE_T (mm d ⁻¹)	MSE_H (mm d ⁻¹)	MSE_M (mm d ⁻¹)	MSE_F (mm d ⁻¹)
SL	19.0 ± 8.3	87.3 ± 5.6	4.3 ± 5.2	8.4 ± 2.8
HH	22.7 ± 6.7	89.1 ± 3.1	3.9 ± 1.9	6.9 ± 2.1
IL	4.4 ± 5.5	68.0 ± 13.4	14.2 ± 10.5	17.7 ± 11.6
YR	16.1 ± 7.7	85.8 ± 4.0	5.6 ± 2.8	8.5 ± 3.0
HU	41.4 ± 9.4	91.5 ± 2.0	3.0 ± 1.4	5.4 ± 1.5
YZ	54.4 ± 19.7	90.1 ± 3.1	4.0 ± 2.2	5.9 ± 2.0
SW	36.3 ± 24.3	85.1 ± 7.3	5.4 ± 2.5	9.5 ± 7.4
SE	85.6 ± 22.1	91.8 ± 1.8	3.7 ± 1.4	4.4 ± 1.3
PD	104.9 ± 44.9	91.6 ± 3.5	4.0 ± 3.1	4.3 ± 1.6

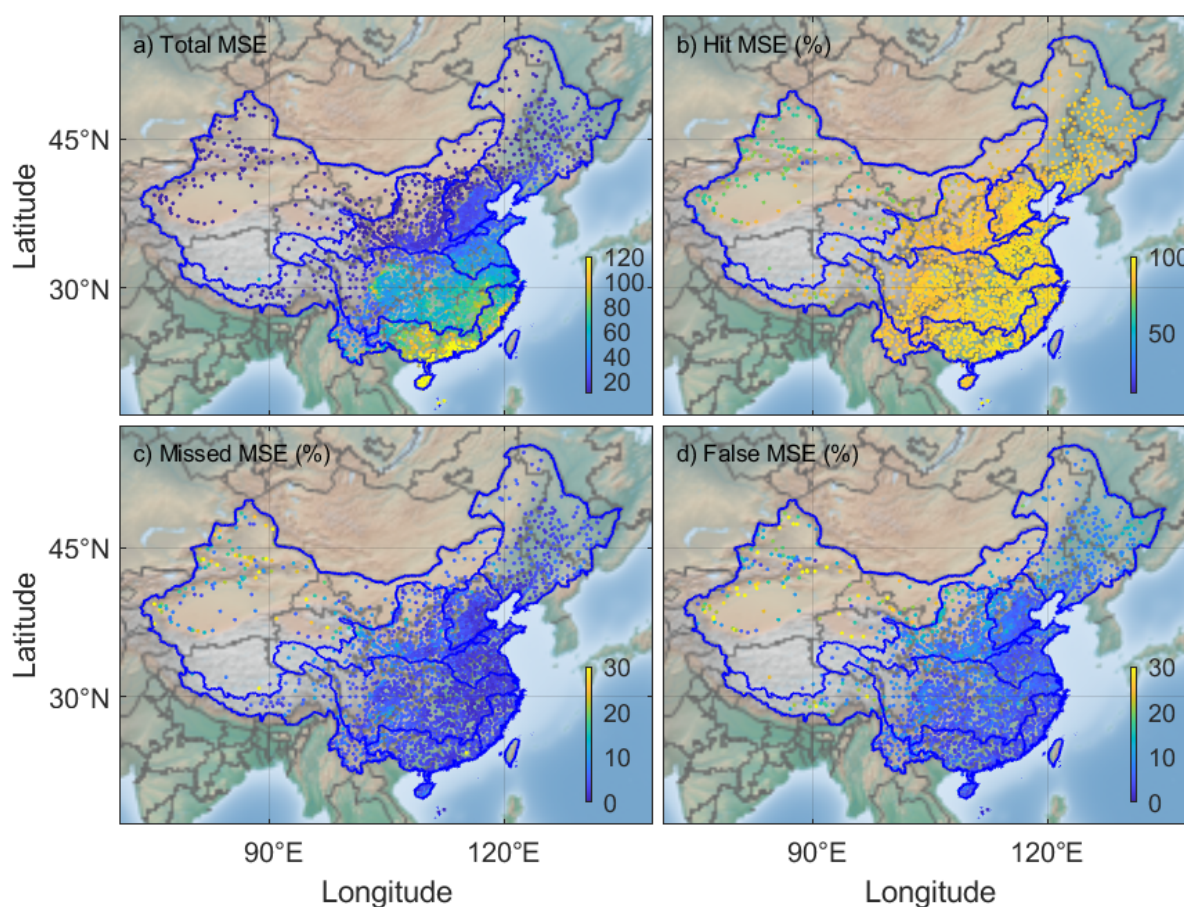


Figure 4. Spatial distribution of the total MSE (a) and percentage of contribution by its three components, i.e., hit (b), miss (c), and false (d) in China.

Figure 5 presents the long-term trend of MBE_T in nine watersheds, which is composed of six components according to Equation (4). All nine watersheds saw an increase in MBE_T during the past twenty years, which suggests a slight deterioration of the IMERG product from this point of view. The growing trend of MBE_T is generally supported by the trends of MBE_H , MBE_M , and MBE_F except in SW and PD. MBE_M has shown a relatively stable trend in the past 20 years. MBE_F decreased from 4.2 to 3.4 in PD and decreased from 3.8 to 3.4 in SE, which would support a decreased MSE_T to some extent. With the help of Equation (5), we can see that the MBE_T trend is not only determined by trends in MBE_H , MBE_M , and MBE_F that are weighted by their mean RFOs, but it is also attributable to changes in the RFOs associated with H, M, and F that are weighted by their average MBE values. The major feature of Figure 6, the time series of RFOs associated with H, M, F, and C, is that IMERG tends to miss fewer precipitation events but produces more false alarms. This feature can be found in nine watersheds but is most prominent in PD, where RFO_F increases from about 7% to 13%, whereas RFO_M decreases from about 10% to 8%. Both would be expected to produce a growing MBE_T as suggested by Equation (4). Regarding MSE_T , the trends in nine watersheds are marginal (Figure 7). Although MSE_H , MSE_M , and MSE_F have improved more or less in SE, SW, and PD (especially in PD, MSE_H , MSE_M and MSE_F has decreased by a quarter in the past 20 years), the MSE_T trend is still negligible. This is because these improvements are partially or fully offset by the negative effect of the large increasing tendency of false alarms on the MSE_T trend.

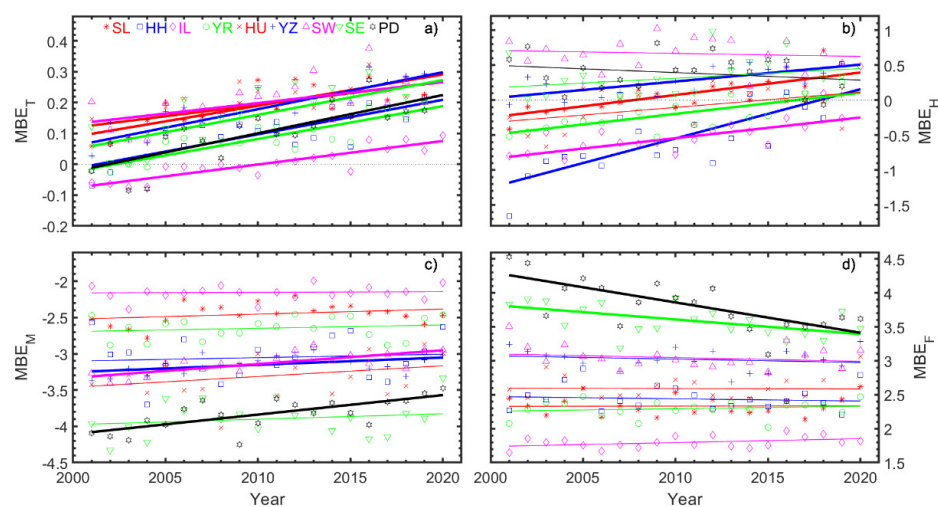


Figure 5. Time series of MBE and its linear trends associated with hit (a), missed (b), false alarm (c), and correct negative (d). A thick line indicates that the trend is statistically significant (p -value < 0.05).

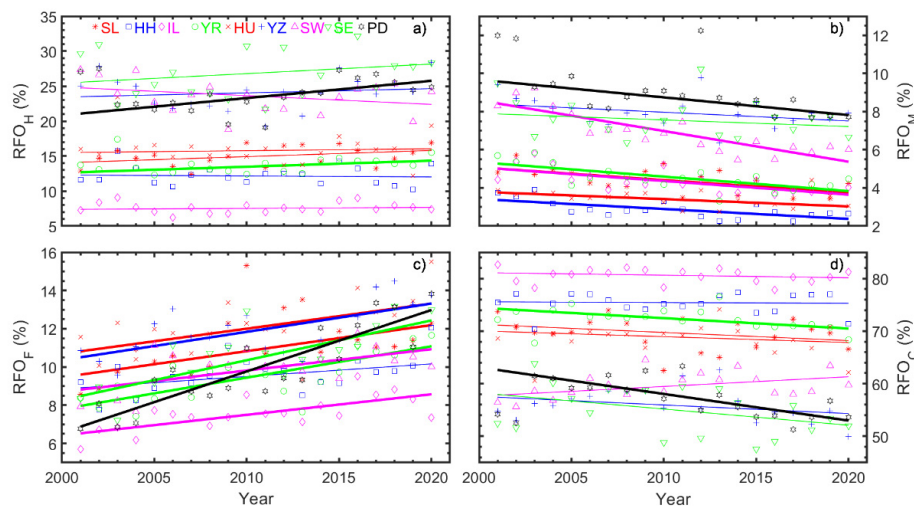


Figure 6. Time series of the relative frequency of occurrence (RFO) and its linear trends associated with hit (a), missed (b), false alarm (c), and correct negative (d). A thick line indicates that the trend is statistically significant (p -value < 0.05).

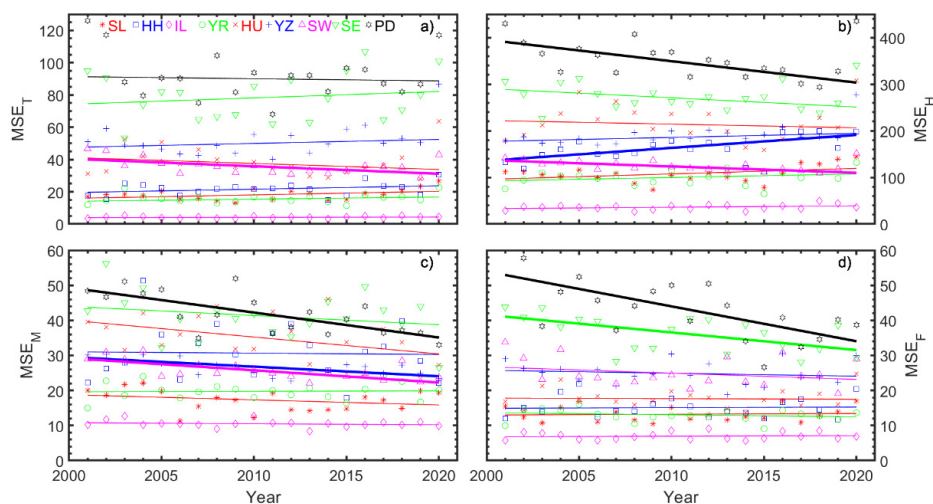


Figure 7. (a–d) Similar to Figure 4, but for MSE.

4. Discussion

Satellite remote-sensing precipitation is still a challenging task even though much progress has been made during recent decades. With the release of new satellite precipitation products, evaluation of the quality of these new products is still urgently required. The fact that most satellite precipitation products provide more than two decades' data is worthy of record. Detailed assessment of long-term performance of satellite precipitation products is critical for its usage in sciences and applications, which has been overlooked in most previous validation studies. Validation must go beyond reporting the overall error metrics. In this study, we use rainfall measurements at 2395 gauge in mainland China to evaluate the latest IMERG rainfall product during the last two decades. To that end, time-series features are used here to study the correspondence between the satellite-derived and measured precipitation transients. We developed an improved validation method that is capable of decomposing MBE and MSE into independent components. A simple but effective method to decompose the trends of MBE and MSE to six individual components are presented, which would be expected to reveal detection capability and retrieval uncertainties, as well as their quantitative contributions to total MBE and MSE. It was suggested that the accuracy of IMERG precipitation retrievals increases over the time period from 2001 to 2018 by using such metrics as CC, CSI, and KGE. This is attributed to more PMW sensors, higher resolutions and more frequency channels [17]. However, it is shown here that the false alarm ratio increases, whereas missing rainfall rate decreases, which leads to an increasing MBE of IMERG precipitation. There are dozens of statistical metrics used in the validation, but no consensus is reached in the community with regard to how many statistical metrics should be included. Generally, these metrics may suffer from three weaknesses, i.e., interdependence, underdetermination, and incompleteness [15]. It is suggested that the error characteristics can be fully described by the joint probability distribution, namely, $p(Y,X)$, which may provide more insight of satellite precipitation retrieval uncertainties. Gauged measurements are always taken as the reference; however, it should be kept in mind that they also suffer from many measurement uncertainties. Furthermore, the spatial scale mismatch between satellite gridded precipitation products and in situ gauged measurements is perhaps the least-understood topic, which needs further discussion. Dense gauge networks are absolutely required for inquiring into the spatial representativeness of gauge stations, which has been shown by very few previous studies [18]. Given the fact that satellite and gauge-based rainfall products are complementary, it is of significance that they may be blended together to produce a more accurate product, a prerequisite for which is certainly to fully understand their uncertainties via comprehensive validation [29,30].

5. Conclusions

Based on the collocated daily IMERG and gauged data during 2000–2020, a comprehensive evaluation of IMERG rainfall products has been performed. A decomposition scheme was developed to separate total MBE and MSE to their components, i.e., the corresponding values associated with hit, missed, and false alarms. The trends of MBE and MSE are partly accounted for by weighted trends in MBEs associated with hit, missed, and false alarms that are weighted by their relative occurrence; furthermore, trends in the relative occurrence frequency of hit, missed, and false alarms weighted by their average MBE or MSE values are also important drivers for changes in total MBE and MSE. Major conclusions are as follows.

IMERG tends to underestimate multiannual mean precipitation in mainland China, which is more evident in rainy seasons. The precipitation-occurrence-detection capability of IMERG in rainy seasons and humid regions is better than that in dry seasons and semi-arid regions.

Weighted MBE_F is larger than weighted MBE_M in magnitude partly as a result of more occurrence of false alarms than missed precipitations. Total MSE is dominantly contributed by weighted hit MSE.

A downward tendency of missed precipitation rate was observed in nine watersheds, while the false-alarm rate increased, thereby leading to a growing trend of total MBE. Improvement in IMERG rainfall detection still needs improvement, especially in reducing the impact of false alarms of precipitation.

Author Contributions: Conceptualization, X.X. (Xiangao Xia); Data curation, D.F. and Y.F.; Formal analysis, X.X. (Xinran Xia); Methodology, X.X. (Xiangao Xia); Writing—original draft, X.X. (Xinran Xia) and W.S.; Writing—review & editing, D.F., W.S. and X.X. (Xiangao Xia). All authors have read and agreed to the published version of the manuscript.

Funding: The research is supported by National Natural Science of Foundation of China (NSFC No. 42030608 and 42075079).

Data Availability Statement: The IMERG data is obtained from <https://disc.gsfc.nasa.gov/datasets/>, accessed on 1 November 2021. The gauged data are available from <http://data.cma.cn>, accessed on 1 November 2021.

Acknowledgments: The authors appreciate the China Meteorological Data Service Center and NASA for providing gauged and IMERG data, respectively.

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

References

1. Kucera, P.A.; Ebert, E.E.; Turk, F.J.; Levizzani, V.; Kirschbaum, D.; Tapiador, F.J.; Loew, A.; Borsche, M. Precipitation from space: Advancing earth system science. *Bull. Am. Meteorol. Soc.* **2013**, *94*, 365–375. [\[CrossRef\]](#)
2. Foufoula-Georgiou, E.; Guilloteau, C.; Nguyen, P.; Aghakouchak, A.; Hsu, K.L.; Busalacchi, A.; Joseph Turk, F.; Peters-Lidard, C.; Oki, T.; Duan, Q.; et al. Advancing precipitation estimation, prediction, and impact studies. *Bull. Am. Meteorol. Soc.* **2020**, *101*, E1584–E1592. [\[CrossRef\]](#) [\[PubMed\]](#)
3. 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.; et al. The Global Precipitation Measurement (GPM) Mission for Science and Society. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1679–1695. [\[CrossRef\]](#)
4. Hou, A.Y.; Kakar, R.K.; Neeck, S.; Azarbarzin, A.A.; Kummerow, C.D.; Kojima, M.; Oki, R.; Nakamura, K.; Iguchi, T. The global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 701–722. [\[CrossRef\]](#)
5. Huffman, G.J.; Bolvin, D.T.; Braithwaite, D.; Hsu, K.; Joyce, R.; Kidd, C.; Nelkin, E.J.; Sorooshian, S.; Tan, J.; Xie, P. NASA Global Precipitation Measurement (GPM) Integrated Multi-satellite Retrievals for GPM (IMERG). *Algorithm Theor. Basis Doc.* **2020**, *4*, 26.
6. Kirschbaum, D.B.; Huffman, G.J.; Adler, R.F.; Braun, S.; Garrett, K.; Jones, E.; McNally, A.; Skofronick-Jackson, G.; Stocker, E.; Wu, H.; et al. NASA'S remotely sensed precipitation: A reservoir for applications users. *Bull. Am. Meteorol. Soc.* **2017**, *98*, 1169–1184. [\[CrossRef\]](#)
7. Huffman, G.J.; Adler, R.F.; Bolvin, D.T.; Gu, G.; Nelkin, E.J.; Bowman, K.P.; Hong, Y.; Stocker, E.F.; Wolff, D.B. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. *J. Hydrometeorol.* **2007**, *8*, 38–55. [\[CrossRef\]](#)
8. Berg, W.; Bilanow, S.; Chen, R.; Datta, S.; Draper, D.; Ebrahimi, H.; Farrar, S.; Jones, W.L.; Kroodsma, R.; McKague, D.; et al. Intercalibration of the GPM Microwave Radiometer Constellation. *J. Atmos. Ocean. Technol.* **2016**, *33*, 2639–2654. [\[CrossRef\]](#)
9. Hong, Y.; Hsu, K.L.; Sorooshian, S.; Gao, X. Precipitation estimation from remotely sensed imagery using an artificial neural network cloud classification system. *J. Appl. Meteorol.* **2004**, *43*, 1834–1852. [\[CrossRef\]](#)
10. Guilloteau, C.; Foufoula-Georgiou, E.; Kummerow, C.D. Global multiscale evaluation of satellite passive microwave retrieval of precipitation during the TRMM and GPM eras: Effective resolution and regional diagnostics for future algorithm development. *J. Hydrometeorol.* **2017**, *18*, 3051–3070. [\[CrossRef\]](#)
11. Joyce, R.J.; Xie, P. Kalman filter-based CMORPH. *J. Hydrometeorol.* **2011**, *12*, 1547–1563. [\[CrossRef\]](#)
12. Tan, J.; Huffman, G.J.; Bolvin, D.T.; Nelkin, E.J. IMERG V06: Changes to the morphing algorithm. *J. Atmos. Ocean. Technol.* **2019**, *36*, 2471–2482. [\[CrossRef\]](#)
13. Li, R.; Wang, K.; Qi, D. 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.* **2018**, *123*, 423–442. [\[CrossRef\]](#)
14. Tian, Y.; Peters-Lidard, C.D.; Eylander, J.B.; Joyce, R.J.; Huffman, G.J.; Adler, R.F.; Hsu, K.L.; Turk, F.J.; Garcia, M.; Zeng, J. Component analysis of errors in Satellite-based precipitation estimates. *J. Geophys. Res. Atmos.* **2009**, *114*, 1–15. [\[CrossRef\]](#)
15. Tian, Y.; Nearing, G.S.; Peters-Lidard, C.D.; Harrison, K.W.; Tang, L. Performance metrics, error modeling, and uncertainty quantification. *Mon. Weather. Rev.* **2016**, *144*, 607–613. [\[CrossRef\]](#)
16. Tang, G.; Ma, Y.; Long, D.; Zhong, L.; Hong, Y. Evaluation of GPM Day-1 IMERG and TMPA Version-7 legacy products over Mainland China at multiple spatiotemporal scales. *J. Hydrol.* **2016**, *533*, 152–167. [\[CrossRef\]](#)

17. Tang, G.; Clark, M.P.; Papalexiou, S.M.; Ma, Z.; Hong, Y. Have satellite precipitation products improved over last two decades? A comprehensive comparison of GPM IMERG with nine satellite and reanalysis datasets. *Remote Sens. Environ.* **2020**, *240*, 111697. [\[CrossRef\]](#)
18. Tan, J.; Petersen, W.A.; Kirchengast, G.; Goodrich, D.C.; Wolff, D.B. Evaluation of global precipitation measurement rainfall estimates against three dense gauge networks. *J. Hydrometeorol.* **2018**, *19*, 517–532. [\[CrossRef\]](#)
19. Li, X.; Sungmin, O.; Wang, N.; Liu, L.; Huang, Y. Evaluation of the GPM IMERG V06 products for light rain over Mainland China. *Atmos. Res.* **2021**, *253*, 105510. [\[CrossRef\]](#)
20. Wang, Q.; Xia, J.; She, D.; Zhang, X.; Liu, J.; Zhang, Y. Assessment of four latest long-term satellite-based precipitation products in capturing the extreme precipitation and streamflow across a humid region of southern China. *Atmos. Res.* **2021**, *257*, 105554. [\[CrossRef\]](#)
21. Chen, F.; Li, X. Evaluation of IMERG and TRMM 3B43 Monthly Precipitation Products over Mainland China. *Remote Sens.* **2016**, *8*, 472. [\[CrossRef\]](#)
22. Murali Krishna, U.V.; Das, S.K.; Deshpande, S.M.; Doiphode, S.L.; Pandithurai, G. The assessment of Global Precipitation Measurement estimates over the Indian subcontinent. *Earth Space Sci.* **2017**, *4*, 540–553. [\[CrossRef\]](#)
23. Anjum, M.N.; Ahmad, I.; Ding, Y.; Shangguan, D.; Zaman, M.; Ijaz, M.W.; Sarwar, K.; Han, H.; Yang, M. Assessment of IMERG-V06 precipitation product over different hydro-climatic regimes in the Tianshan Mountains, North-Western China. *Remote Sens.* **2019**, *11*, 2314. [\[CrossRef\]](#)
24. Gupta, H.V.; Kling, H.; Yilmaz, K.K.; Martinez, G.F. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* **2009**, *377*, 80–91. [\[CrossRef\]](#)
25. Kling, H.; Fuchs, M.; Paulin, M. Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. *J. Hydrol.* **2012**, *424*, 264–277. [\[CrossRef\]](#)
26. Moazam, S.; Najafi, M.R. A Comprehensive Evaluation of GPM-IMERG V06 and MRMS with Hourly Ground-Based Precipitation Observations across Canada. *J. Hydrol.* **2020**, *29*, 125929. [\[CrossRef\]](#)
27. Derin, Y.; Yilmaz, K.K. Evaluation of multiple satellite-based precipitation products over complex topography. *J. Hydrometeorol.* **2016**, *154*, 1498–1516. [\[CrossRef\]](#)
28. Mei, Y.; Nikolopoulos, E.I.; Anagnostou, E.N.; Borga, M. Evaluating satellite precipitation error propagation in runoff simulations of mountainous basins. *J. Hydrometeorol.* **2016**, *17*, 1407–1423. [\[CrossRef\]](#)
29. Ehsan Bhuiyan, M.A.; Nikolopoulos, E.I.; Anagnostou, E.N. Machine Learning-Based Blending of Satellite and Reanalysis Precipitation Datasets: A Multiregional Tropical Complex Terrain Evaluation. *J. Hydrometeorol.* **2012**, *20*, 2147–2161. [\[CrossRef\]](#)
30. Bhuiyan, M.A.E.; Yang, F.; Biswas, N.K.; Rahat, S.H.; Neelam, T.J. Machine Learning-Based Error Modeling to Improve GPM IMERG Precipitation Product over the Brahmaputra River Basin. *Forecasting* **2020**, *2*, 14. [\[CrossRef\]](#)