Method of AI-Based Precipitation Estimation by Using FY-4B Satellite Data

Nianqing Liu , Jianying Jiang * , Dongyan Mao , Meng Fang , Yun Li , Bowei Han , Suling Ren

Posted Date: 19 August 2024

doi: 10.20944/preprints202408.1289.v1

Keywords: FY-4B; artificial intelligence; multi-temporal; precipitation estimation; evaluation

Article

Method of AI-Based Precipitation Estimation by Using FY-4B Satellite Data

Nianqing Liu [1,2,3], Jianying Jiang [1,2,3,*], Dongyan Mao [1,2,3], Meng Fang [1,2,3], Yun Li [1,2,3], Bowei Han [1,2,3] and Suling Ren [1,2,3]

[1] National Satellite Meteorological Center (National Centre for Space Weather), Beijing 100081, China
[2] Innovation Center for FengYun Meteorological Satellite (FYSIC)
[3] Key Laboratory of Radiometric Calibration and Validation for Environmental Satellites/Key Laboratory of Space Weather, CMA
* Correspondence: jiangjy@cma.gov.cn

Abstract: A smart precipitation estimation method based on FY-4B meteorological satellite data (FY-4B_AI) is proposed in this paper. This method spatiotemporally matches 125 features derived from multi-temporal and multi-channels of FY-4B satellite data with the precipitation at stations. Then a precipitation model using the Light Gradient Boosting Machine algorithm is constructed. Comparative results between FY-4B_AI and GPM/IMERG-L products for over 450 million station cases throughout 2023 shows that: 1) FY-4B_AI is superior to GPM IMERG-L in the average absolute error, root mean square error, relative error, correlation coefficient, probability of detection and critical success index. While in the mean error and false alarm rate, FY-4B_AI is slightly inferior to GPM/IMERG-L. 2) Evaluation of strong weather event applications reveals that both FY-4B_AI and GPM/IMERG-L can accurately represent the spatial distribution characteristics of precipitation, no matter in the southeast humid region or the northwest dry region. Notably, FY-4B_AI, due to its higher spatiotemporal resolution, provides a more detailed distribution of precipitation.

Keywords: FY-4B; artificial intelligence; multi-temporal; precipitation estimation; evaluation

1. Introduction

Precipitation is an important component of the global water cycle, playing a crucial role in global water and energy balance [1,2]. Accurate precipitation estimation is essential for the development of effective environmental policies and disaster mitigation measures. Traditional precipitation observation methods rely on direct measurements from ground-based rain gauges. Although ground observations provide high-precision precipitation data, their spatial distribution is constrained by topography, which makes it difficult to accurately reflect the spatio-temporal distribution patterns of precipitation in data-sparse regions (mountainous areas, oceans, etc.) [3,4]. Traditional precipitation estimation methods primarily rely on ground observations and radar data. However, weather radars, due to the terrain blockage, the uncertainty of the relationship between rainfall radar reflectivity and rainfall intensity, and the limitation of telemetry range and other factors lead to insufficient precision of precipitation observation in complex terrain areas[5]. As one of the important means of precipitation monitoring, satellite observation has the advantages of all-weather, continuous observation, and high spatial and temporal resolution, which effectively makes up for the shortcomings of the ground observation methods [6].

Over the past decades, scientists have been exploring how to utilize satellite data for precipitation estimation. Early studies primarily focused on the use of visible and infrared window channels, while recent studies have increasingly focused on the application of infrared split-window channels and other multi-channel infrared data. Infrared split-window channels have been widely used for precipitation estimation due to their ability to provide information about cloud top temperature and atmospheric vertical structure. This information is crucial for identifying precipitation areas and estimating precipitation intensity. For example, lower cloud top temperatures

 doi:10.20944/preprints202408.1289.v1

are often associated with stronger precipitation events. Besides infrared split-window channels, other infrared channels, such as infrared water vapor channels, have also been used for precipitation estimation. These channels can provide information about the water vapor content in the atmosphere, which is essential for understanding the dynamics of precipitation systems.

Compared to infrared and visible channels, microwave channels possess stronger penetration capabilities, allowing for the acquisition of structural information within precipitation cloud systems[7]. Grody pioneered the development of a method for estimating land surface precipitation based on microwaves[8]. The Tropical Rainfall Measuring Mission (TRMM) satellite, a joint mission by the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA), carried the Precipitation Radar (PR), the world's first spaceborne precipitation radar, capable of providing a three-dimensional structure of heavy rainfall[9]. TRMM also carried microwave instruments capable of providing high-quality microwave-based precipitation estimates. On the base of scattering theory, using the dynamic clustering method and according to the brightness temperature of nine microwave channels of TRMM/TMI (TRMM Microwave Imager) and their combinations, Li et al.[10] established a rainfall intensity retrieval algorithm with scattering index and polarization corrected temperature (PCT) as parameters. Based on the MicroWave Radiation Imager (MWRI) data of TRMM/TMI and FY-3B, Li et al.[12] generated a cloud-radiation dataset that can describe MWRI observation features to retrieve the total rainfall intensity. The Global Precipitation Measurement(GPM) mission, succeeding TRMM, is a new generation global satellite precipitation product, providing strong data support for research on global climate change, drought and flood monitoring, etc. [12,13].

Recent studies have shown that combining satellite data from multiple frequency bands can enhance precipitation estimation accuracy. This integration not only encompasses maximizing the extraction of information from individual frequency bands but also includes the synergistic effects between different frequency bands, known as "multi-band synergy"(MBS). This approach can better capture the complexity of precipitation-related phenomena and thus improve the accuracy and reliability of precipitation estimates.

Despite significant progress, precipitation estimation based on satellite data still faces challenges. For example, how to identify precipitation regions more accurately, how to improve the spatial and temporal resolution of precipitation estimation, and how to deal with the opacity of clouds are all topics that need to be further investigated.

In recent years, machine learning (ML) techniques have been increasingly applied in meteorology [14–16], especially in precipitation estimation using satellite remote sensing data. Machine learning methods, such as deep learning and attention mechanisms, have proven to provide higher accuracy and efficiency in handling complex weather systems and precipitation patterns. Deep learning techniques [17,18], particularly Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have achieved significant advancements in the field of precipitation forecasting. For instance, the U-Net architecture has been utilized for near-real-time precipitation estimation, demonstrating its powerful capabilities in processing image data, which is particularly important for satellite remote sensing data. Attention mechanisms [19] are methods that allow models to focus on the most relevant information in input data, showing significant potential in improving precipitation estimation accuracy. By applying attention mechanisms to satellite remote sensing data, researchers can better identify and predict precipitation events, leading to improved precipitation estimation precision.

With the advancements of meteorological satellites, spaceborne sensors, and artificial intelligence, precipitation estimation based on satellite data mainly includes precipitation inversion based on channel radiation characteristics and satellite precipitation inversion based on artificial intelligence machine learning [20,21]. Liu et al. [22] analyzed a shear line event over the Tibetan Plateau by using a precipitation estimation product combining the FY-4A meteorological satellite and artificial intelligence. They compared single-temporal and multi-temporal geostationary satellite precipitation estimation products, concluding that multi-temporal geostationary satellite precipitation estimation products improved precipitation inversion efficiency. The algorithms

 doi:10.20944/preprints202408.1289.v1

continuously integrate the latest available ground precipitation data to update the model iteratively, without relying on simultaneous ground-based precipitation observations. This significantly enhances the effectiveness of the inversion product, and it has great significance to the catastrophic weather warning.

This paper is an updated iteration and further refinement of the reference [22]. Firstly, FY-4B is used to replace FY-4A. The FY-4B meteorological satellite was successfully launched on June 3, 2021. It features an additional 7.24-7.6 μm low-level water vapor channel compared to the FY-4A meteorological satellite. The rapid changes in low-level water vapor play a significant role in indicating the development of local sudden convective systems. Secondly, since multi-temporal satellite precipitation estimation products are more practical in real-time applications (refer to as [22]), this paper only adopts the multi-temporal FY-4B satellite data precipitation estimation algorithm (FY-4B_AI), without considering the single-frame satellite precipitation estimation products. The research scope from [22] is expanded from the Tibetan Plateau to the entire mainland China, followed by product accuracy evaluation, aiming to establish a disaster weather FY-4B satellite precipitation inversion model suitable for mainland China.

This paper is structured as follows: Section 2 introduces the study area and data; Section 3 presents the research methods; Section 4 examines the accuracy of the precipitation estimation product based on artificial intelligence; Section 5 evaluates its application in strong weather events; finally, Section 6 provides conclusions and discussions.

2. Verification Area and Data

2.1. Verification Area

The validation area encompasses the entire mainland China, ranging from 70°E to 135°E longitude and 15°N to 55°N latitude, as shown in Figure 1.

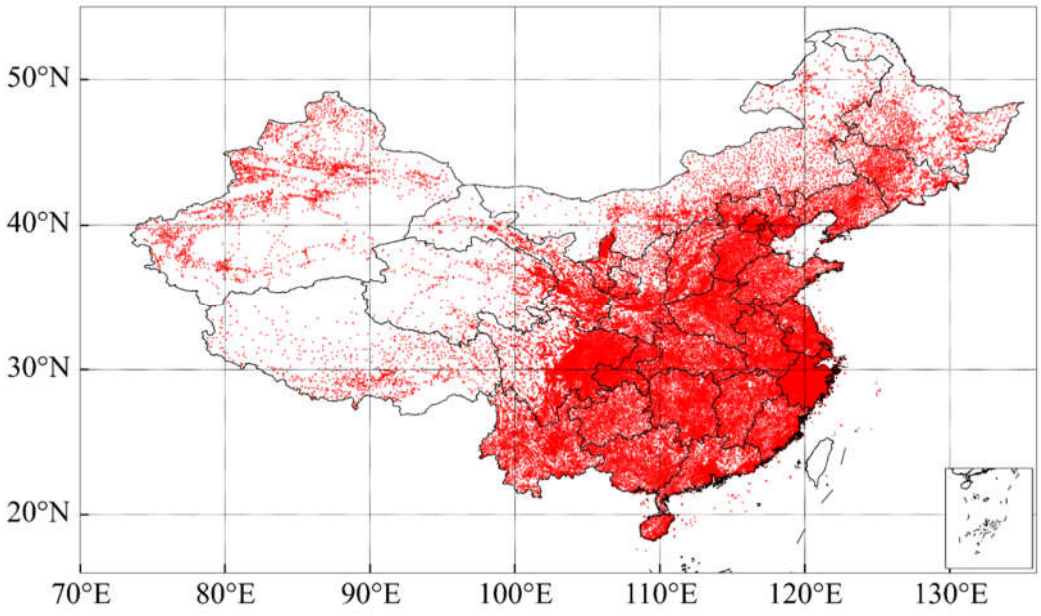


Figure 1. The distribution of ground-based precipitation observation stations in mainland China (red dots). (cited from Figure 1a of reference [22]).

2.2. FY-4B/AGRI Data

The Fengyun-4B (FY-4B) satellite, which launched on June 3, 2021, is the first operational satellite in China's new generation of geostationary meteorological satellites, the Fengyun-4 series [23]. It carries four instruments, three of which are weather-related, namely, the Advanced Geostationary Radiation Imager(AGRI), the Geostationary High-speed Imager (GHI), and the Geostationary

 doi:10.20944/preprints202408.1289.v1

Interferometric Infrared Sounder(GIIRS). The FY-4B/AGRI has 15 channels (Table 1), with a resolution of 0.5-1 km for visible and near-infrared channels, 2 km for shortwave infrared channels, 2-4 km for mid-wave infrared channels, and 4 km for far-infrared channels. This study utilizes the radiance brightness temperature values and derived variables from channels 07-15 (a total of 9 channels) in the mid-wave infrared to far-infrared bands of FY-4B/AGRI as input. The original horizontal resolution of channel 7 is 2 km. The input data used here are reprocessed with a downscaled resolution of 4 km, ensuring consistent spatial resolution across all channels.

Table 1. FY-4B/AGRI Performance Parameters.

Band	Central Wavelength (μm)	Bandwidth (μm)	Spatial Resolution (km)	Main purpose
1	0.47	0.45～0.49	1	Aerosol
2	0.65	0.55～0.75	0.5	Fog, cloud
3	0.825	0.75～0.90	1	Vegetation
4	1.379	1.371～1.386	2	Cirrus
5	1.61	1.58～1.64	2	Cloud, snow
6	2.25	2.10～2.35	2	Cirrus, aerosol
7	3.75	3.50～4.0（high）	2	Fire
8	3.75	3.50～4.0（low）	4	Land surface
9	6.25	5.80～6.70	4	Upper-level water vapor
10	6.95	6.75～7.15	4	Mid-level water vapor
11	7.42	7.24～7.60	4	Low-level water vapor
12	8.55	8.3～8.8	4	Cloud
13	10.80	10.30～11.30	4	Surface temperature
14	12.00	11.50～12.50	4	Surface temperature
15	13.3	13.00～13.60	4	Clouds and water vapor

This study uses data from the Northern Hemisphere region with a 15-minute observation frequency, specifically data observed at the start of every 00, 15, 30, and 45 minutes of each hour.

2.3. Ground-Based Rain Gauge Precipitation Data in Mainland China

The ground-based rain gauge precipitation data used in this study are hourly observations from China's ground meteorological stations, including both national-level ground meteorological stations (more than 10,000) and regional meteorological automatic stations (more than 50,000), totaling approximately 70,000 stations. The distribution of rain gauges (red dots in the figure) is shown in Figure 1. As can be seen from the figure, ground rain gauges are densely distributed in the low-altitude areas of eastern and central China, while they are relatively sparse in the high-altitude areas of western China and northern regions like Inner Mongolia.

The ground precipitation observation data used in this study includes observation time, 1-hour accumulated rainfall (unit: mm/h), longitude and latitude of the station, and data quality control code. The minimum observed rainfall value is 0.1 mm/h. The data are near real-time. Quality control was performed on the data, and only data with a quality control code of 0, indicating no data anomalies, were selected for analysis. Within the research time range from 00:00 on January 1, 2023, to 23:00 on December 31, 2023, a total of 70,000 stations/hour × 24 hours/day × 365 days of data were collected. After removing data from satellite orbit control and stations that failed quality control, 453 million station times of data were used for testing.

 doi:10.20944/preprints202408.1289.v1

2.4. GPM/IMERG-L Dataset

The Integrated Multi-satellite Retrievals for GPM (IMERG) is the unified U.S. algorithm that provides the multi-satellite precipitation product for the U.S. GPM team[24]. IMERG late precipitation estimates from the various precipitation-relevant satellite passive microwave (PMW) sensors comprising the G PM constellation are computed using the 2017 version of the Goddard Profiling Algorithm (GPROF2017), then gridded, intercalibrated to the GPM Combined Ku Radar-Radiometer Algorithm (CORRA) product, and merged into half-hourly 0.1°x0.1° fields. "Late" multi-satellite product ~14h after observation time, using both forward and backward morphing. The hourly precipitation data applied in the paper is from January 1 to December 31, 2023. Unit: mm/h.

3. Research Methods

3.1. Spatiotemporal Matching Method for Testing the Accuracy of Satellite Precipitation Estimation

Previous studies [22,25–27] have shown that PERSIANN-CCS data have unsatisfactory performance in the accuracy test. Therefore, this study only compares the AI-based FY-4B meteorological satellite precipitation estimation (FY-4B_AI) with that of GPM/IMERG-L. The spatial resolution of FY-4B_AI products is 0.04° longitude × 0.04° latitude, with a temporal resolution of 1 hour, and the unit is mm/h. The spatial resolution of GPM/IMERG-L is 0.1° longitude × 0.1° latitude, with a temporal resolution of 0.5 hour, and the unit is mm/h. The precipitation estimation product accuracy verification uses ground-based rain gauge observations as the truth value.

Spatiotemporal matching is applied to the data. The temporal matching method is as follows: FY-4B_AI data from 00:00 to 01:00 is matched with ground rain gauge data from 00:00 to 01:00; GPM/IMERG-L data from 00:00 to 00:30 and 00:30 to 01:00 are averaged and then matched with ground rain gauge data from 00:00 to 01:00. The spatial matching method is to select the nearest grid point of the satellite precipitation product based on the location of the ground rain gauge station.

3.2. Satellite Precipitation Estimation Product Evaluation Methods

In this study, eight evaluation indexes, namely mean error (ME), mean absolute error (MAE), root mean square error (RMSE), relative error (RE), correlation coefficient (CC), probability of detection (POD), false alarm rate (FAR), and critical success index (CSI), are used to evaluate the performances of the two satellite precipitation products, FY-4B_AI and GPM/IMERG-L, in precipitation events (refer to reference [22] for details).

3.3. Intelligent Precipitation Estimation Algorithm Based on Multi-Temporal Satellite Data

Meteorological ground station precipitation represents the accumulated precipitation over the past hour. In conditions of rapidly evolving weather systems, the precipitation rate within a single hour can vary significantly. Therefore, this study utilizes multi-temporal satellite observation information to obtain better precipitation estimation results. The time-frequency of FY-4B/AGRI in the Northern Hemisphere observation is 15 minutes. This study selects data with 15-minute intervals within one hour as input, derives multiple features based on channel information, and trains a model using the Light Gradient Boosting Machine (LGBM) algorithm to achieve 1-hour satellite precipitation inversion. The LGBM is a decision tree-based ensemble algorithm known for its fast training speed, low memory consumption, high accuracy, and support for distributed processing. It is commonly used to address various classification and regression problems.

3.3.1. Feature Construction for FY-4B_AI Satellite Precipitation Estimation

In the FY-4B_AI method, taking the 01:00 satellite precipitation estimation algorithm as an example, the model receives input from five time points: 00:00, 00:15, 00:30, 00:45, and 01:00. Each time point has nine channels of radiance brightness temperature data. And the features derived from these data include the time rate of change of radiation values and the radiation differences between different channels at the same time.

 doi:10.20944/preprints202408.1289.v1

The constructed features include (Figure 2): (1) Maximum and minimum brightness temperatures of the nine channels at the five time points, a total of 18 features; (2) Maximum and minimum brightness temperature differences of the nine channels at 15-minute intervals, a total of 18 features; (3) Maximum and minimum brightness temperature differences of the nine channels at 30-minute intervals, a total of 18 features; (4) Maximum and minimum brightness temperature differences of the nine channels at 45-minute intervals, a total of 18 features; (5) Differences between the nine channels at 60-minute intervals, a total of 9 features; (6) Maximum and minimum differences between different channels at the same time point. Considering that channel 7 (sensitive to the high-temperature end) and channel 8 (sensitive to the low-temperature end) in the mid-infrared contain solar radiation information during the day, the different features between different channels at the same time point is only calculated for channels 7 and 8, and do not participate in the 9-15 channel subtraction, a total of 2 x (1 + 6 + 5 + 4 + 3 + 2 + 1) = 44 features. Therefore, a total of 18 + 18 + 18 + 18 + 9 + 44 = 125 features are derived. These features capture a large amount of satellite data information within one hour, laying the foundation for training a high-precision precipitation estimation model.

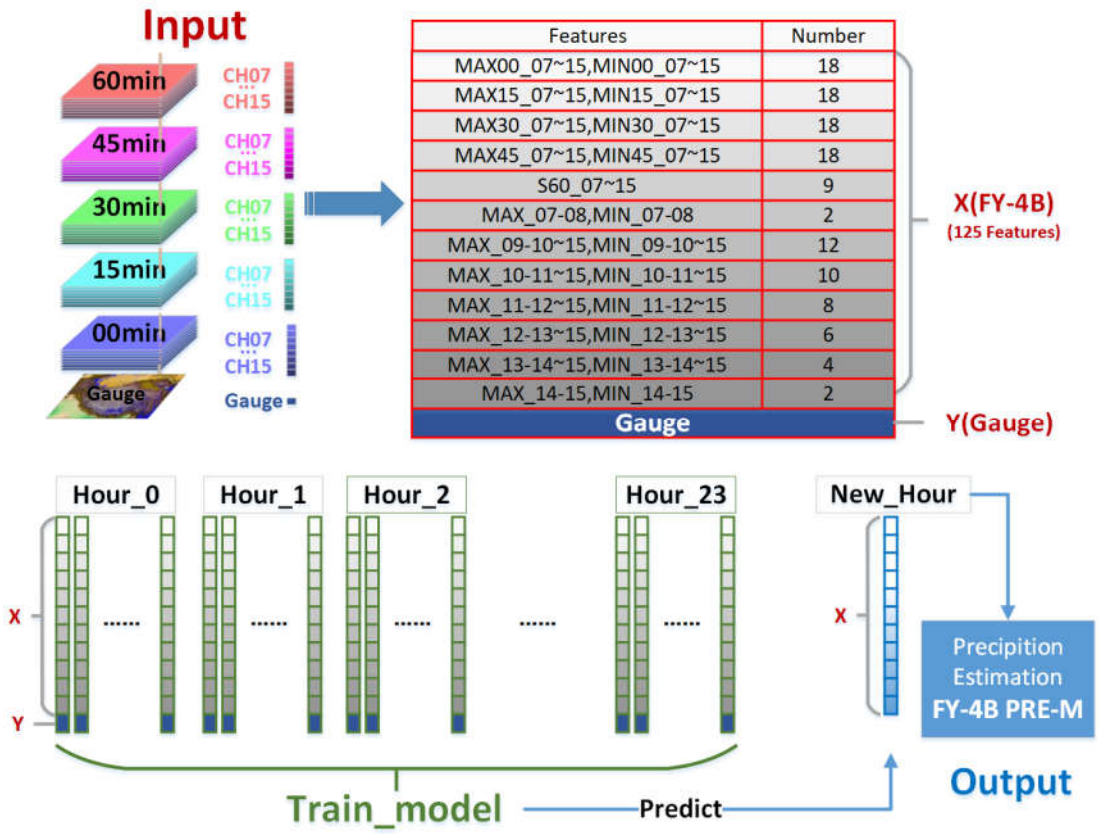


Figure 2. Data Processing Flow and Algorithm Model for multi-temporal FY-4B Meteorological Satellite Precipitation Estimation Product based on Artificial Intelligence.

3.3.2. Precipitation Estimation Algorithm Model

FY-4B_AI utilizes a Light Gradient Boosting Machine-based algorithm. The construction process is illustrated by using the 01:00 FY-4B_AI precipitation estimation algorithm model as an example. Based on the satellite data at 00:00, 00:15, 00:30, 00:45, and 01:00, the 125 features (X) mentioned above are calculated. Selecting the 1-hour rainfall from the ground rain gauge at 01:00 (Y, Gauge) as the truth value, a training data point is established through the spatial matching method between satellite data and ground rain gauge precipitation data. If there are N ground rain gauge observation stations in the study area, N training data points are established. When estimating precipitation from satellite data each time, only the satellite data and precipitation observation data available in the previous 24

 doi:10.20944/preprints202408.1289.v1

hours are used to train the model. For example, to estimate precipitation at 01:00 today, satellite data from yesterday's 01:00 to today's 00:00 and precipitation observation data are used as input for training the model. The model does not require input of the current 1-hour ground precipitation observation data. It can immediately perform precipitation estimation after obtaining the current satellite data, effectively enhancing the timeliness of FY-4B_AI precipitation estimation. This is particularly important for monitoring and warning of rapidly developing and changing strong convective weather. Furthermore, even if there is a small amount of missing rain gauge precipitation data or satellite data in the previous 24 hours, the model can still be trained. Therefore, the model exhibits excellent compatibility. The model establishment and data processing flow are shown in Figure 2.

In the algorithmic model, each input feature is analyzed and scored for its importance. A higher score indicates greater importance. Analysis of the 125 features hour by hour from January 1 to December 31, 2023 (ranking details are omitted), reveals that 13 of the top 20 most important features are brightness temperature differences between different channels at the same time point, 6 are single-channel brightness temperature values, and 1 is the maximum value of a single channel across multiple time points. The results show that the most crucial information for satellite precipitation inversion comes from the differences between different channels at the same time point, with key information contained in the mid-wave infrared to far-infrared bands. Features ranked 21-40 primarily consist of brightness temperature differences of the same channel at different times. Overall, the importance of brightness temperature differences at 15 minutes, 30 minutes, 45 minutes, and 60 minutes decreases sequentially. This is crucial additional information for improving the quality of satellite precipitation estimation products.

4. Accuracy Verification of Precipitation Estimation Products from Meteorological Satellites Based on Artificial Intelligence

This study's process of precipitation estimation inversion using multi-temporal satellite data based on artificial intelligence only requires historical 1-hour ground rain gauge observations as input data for training the model. In other words, when inverting precipitation at 01:00 using satellite data, the rain gauge data at that specific time point is not used. This means the satellite precipitation estimation does not incorporate rain gauge observation data. When evaluating the accuracy of the FY-4B_AI satellite precipitation estimation product, the 1-hour precipitation from ground rain gauge stations within the study area is used as the truth value of the test. The study analyzes and compares the FY-4B_AI and GPM/IMERG-L satellite-derived data, conducting statistical analysis by time (month) and region (along the 400 mm rainfall line), respectively.

4.1. Overall Accuracy Assessment

From January 1 to December 31, 2023, two types of satellite precipitation estimation (FY-4B_AI and GPM/IMERG-L) and ground rain gauge precipitation were carried out to assess the accuracy through the time-space matching method, and the results of the precipitation accuracy assessment are shown in Table 2.

Table 2. Accuracy Assessment Results for Satellite-derived Precipitation.

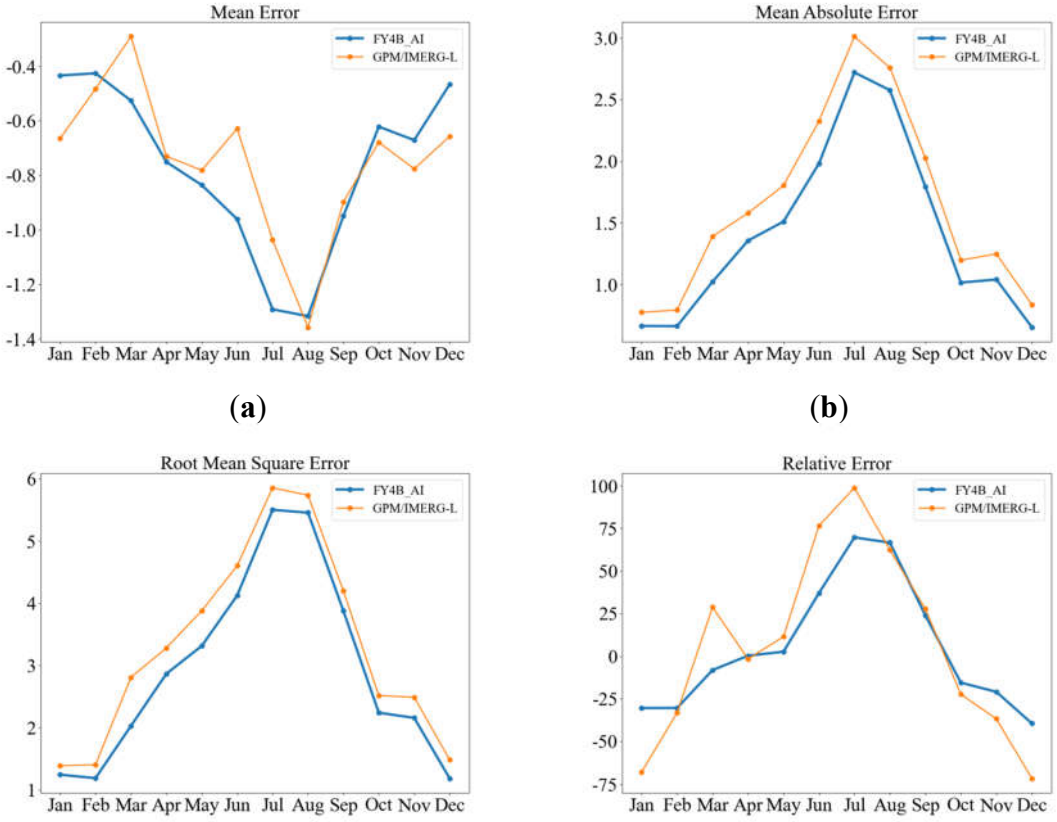
INDEX	FY-4B_AI	GPM/IMERG-L
ME	-0.88	-0.80
MAE	1.67	1.92
RMSE	3.68	4.07
RE	17.72%	26.28%
CC	0.44	0.36
POD	61.84%	47.31%
FAR	62.85%	62.21%
CSI	0.30	0.27

 doi:10.20944/preprints202408.1289.v1

The results show that (Table 2): among the eight evaluation indexes, six indexes of FY-4B_AI are better than those of GPM/IMERG-L, i.e., mean absolute error(MAE), root mean square error(RMSE), relative error(RE), correlation coefficient(CC), probability of detection(POD) and critical success index(CSI). In particular, the RE of FY-4B_AI is 17.72%, significantly lower than that of GPM/IMERG-L (26.28%).

4.2. Monthly Variation Characteristics of Satellite-Derived Precipitation Accuracy

The monthly distribution maps of the average precipitation inversion evaluation for FY-4B_AI and GPM/IMERG-L from January to December 2023 (Figure 3) show that several indicators exhibit distinct annual variation characteristics. The monthly average indicators for the two types of satellite-derived precipitation evaluation generally show that FY-4B_AI is better than GPM/IMERG-L. Specifically, FY-4B_AI outperforms GPM/IMERG-L in mean absolute error, root mean square error, correlation coefficient, and probability of detection. Among them, mean absolute error and root mean square error are larger in the summer half-year (April-September) and smaller in the winter half-year (October-March). The mean absolute error and root mean square error of FY-4B_AI are better than those of GPM/IMERG-L, reaching their minimum in December 2023 and their maximum in July 2023, suggesting a correlation between these features and seasonal variations in precipitation. The correlation coefficient and probability of detection also show FY-4B_AI outperforming GPM/IMERG-L, but their correlation with season is not obvious.



 doi:10.20944/preprints202408.1289.v1

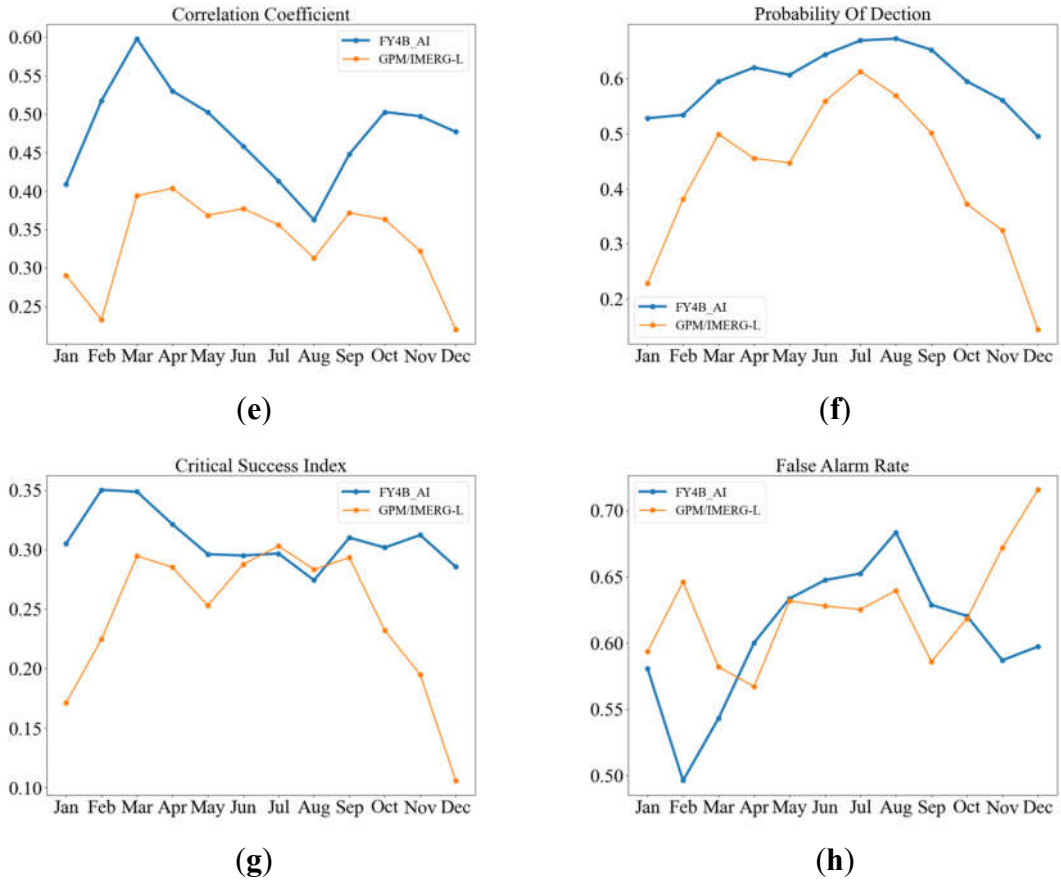


Figure 3. Monthly Distribution of Satellite-derived Precipitation Accuracy Evaluation Indexes from January to December 2023.

The mean error of GPM/IMERG-L is the smallest in March 2023 and the largest in August 2023. The mean error of FY-4B_AI is better than that of GPM/IMERG-L in January, February, October, and December 2023. According to BO ZHAO et al. [28], GPM/IMERG-L is not sensitive enough to detect precipitation events in winter. Therefore, it can be assumed that the mean error of FY-4B_AI is better than that of GPM/IMERG-L in winter precipitation (of course, this is only the result for one year, 2023, and more years of data are needed to verify it). Relative error exhibits a similar characteristic distribution as mean error. In terms of the critical success index, GPM/IMERG-L outperforms FY-4B_AI only in July and August 2023, while FY-4B_AI outperforms GPM/IMERG-L in all other months. The false alarm rate is roughly equivalent for FY-4B_AI and GPM/IMERG-L, consistent with the results shown in Table 2.

4.3. Spatial Variation Characteristics of Satellite-Derived Precipitation Accuracy (Northwest Dry Region/Southeast Humid Region)

As seen in Figure 1, ground rain gauge stations are densely distributed in the low-altitude areas of eastern and central China, while they are relatively sparse in the high-altitude areas of western China and northern regions like Inner Mongolia. Satellite observation of precipitation has the advantages of continuous, all-weather, and high spatial and temporal resolution, and thus has a natural advantage in high-altitude areas like western China regions and northern Inner Mongolia. Therefore, it is necessary to discuss the accuracy of satellite precipitation products in different regions of China (Northwest Dry Region/Southeast Humid Region).

Statistically, the modern 400-mm isoprecipitation line from the Greater Khingan Mountains to the southwest, passing through Zhangjiakou in Hebei Province, Lanzhou in Gansu Province, Lhasa in Tibet, and reaching the eastern Himalayas. It serves as a significant boundary separating monsoon and non-monsoon regions, arid and semi-arid regions in China. It is also a boundary separating

 doi:10.20944/preprints202408.1289.v1

farming and nomadic herding. Furthermore, it reveals a high correlation between climate and population density. That is, one side of this line is the semi-arid region, while the other side is the humid region.

After analyzing the overall and monthly accuracy characteristics of the two precipitation products, FY-4B_AI and GPM/IMERG-L, in 2023, this subsection uses the 400 mm average rainfall line in 2023 as a boundary (Figure 4) to distinguish between the Northwest Dry Region and the Southeast Humid Region. It investigates the precipitation distribution characteristics on both sides of this boundary line and compares the accuracy of two different precipitation products in these two regions. The results indicate (Tables 3 and 4) that FY-4B_AI is better than GPM/IMERG-L in 7 out of 8 evaluation indexes in both the Northwest Dry Region and the Southeast Humid Region. Especially, the relative error of FY-4B_AI is significantly lower than that of GPM/IMERG-L.

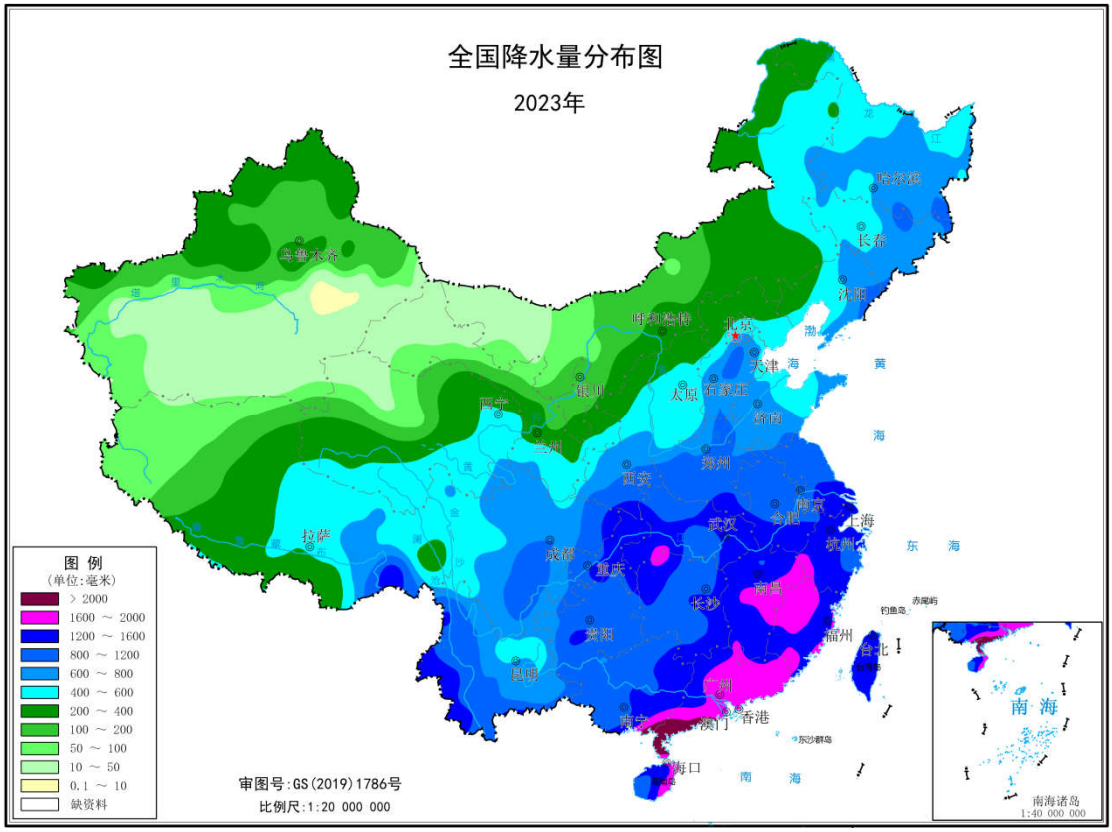


Figure 4. Precipitation Distribution Map of China in 2023 (Unit: mm).**(Cited from Figure 1.1**5 of the "China Climate Bulletin (2023)").

Table 3. Accuracy Assessment Results for Satellite-derived Precipitation in Northwest Dry Region.

INDEX	FY-4B_AI	GPM/IMERG-L
ME	-0.73	-0.79
MAE	1.24	1.37
RMSE	2.64	2.83
RE	-3.78%	-20.73%
CC	0.35	0.32
POD	59.05%	38.25%
FAR	72.89%	65.45%
CSI	0.23	0.22

 doi:10.20944/preprints202408.1289.v1

Table 4. Accuracy Assessment Results for Satellite-derived Precipitation in Southeast Humid Region.

INDEX	FY-4B_AI	GPM/IMERG-L
ME	-0.91	-0.80
MAE	1.76	2.03
RMSE	3.85	4.27
RE	21.95%	35.53%
CC	0.45	0.36
POD	62.39%	49.09%
FAR	60.10%	61.66%
CSI	0.32	0.27

5. Application Evaluation of Strong Weather Events

Based on the above analysis, the accuracy evaluation of the two types of satellite precipitation inversion indicates that FY-4B_AI is generally better than GPM/IMERG-L in 6 out of 8 evaluation indexes. To analyze the application capabilities of these two satellite precipitation inversion products in typical strong weather events over mainland China, this study evaluates their application capabilities in two strong weather events that occurred in the Southeast Humid Region in June 2023 and the Northwest Dry Region in July 2023.

5.1. Application Evaluation of the Strong Weather Event in Guizhou on June 18, 2023 (Southeast Humid Region)

From 00:00 to 00:00 (UTC, the same below) on June 18-19, 2023, Guizhou experienced a significant heavy rainfall process. Notably, the Puding station in Anshun, Guizhou, recorded a 24-hour rainfall of 194.4 mm (heavy rainstorm), ranking first in the 24-hour rainfall list (the second strongest rainfall in June locally). The heavy rainfall at the Puding station mainly occurred from 13:00 to 14:00 on June 18, with a short-duration heavy rainfall of 105.1 mm/h. This subsection uses the 1-hour rainfall at the Puding station from 13:00 to 14:00 on June 18 as an example to evaluate application capabilities.

The overlay of infrared cloud images and rainfall at five consecutive time points (13:00, 13:15, 13:30, 13:45, and 14:00) for FY-4B_AI, ground station truth values, and GPM/IMERG-L are shown in Figure 5. The results indicate: 1) Compared to the ground station truth values, both FY-4B_AI and GPM/IMERG-L accurately reproduce the spatial distribution characteristics of precipitation. However, due to the spatial resolution of FY-4B_AI being 0.04° and that of GPM IMERG-L being 0.1°, FY-4B_AI provides a more detailed distribution of precipitation. 2) Meanwhile, due to the temporal resolution of FY-4B_AI being 15 minutes and that of GPM IMERG-L being 30 minutes, FY-4B_AI updates its results five times (15 minutes/time) within the one-hour period from 13:00 to 14:00, while GPM/IMERG-L (30 minutes/time) updates three times, and the station only updates twice (60 minutes/time). This clearly demonstrates the high temporal resolution characteristics of FY-4B_AI. 3) Both FY-4B_AI and GPM/IMERG-L can capture the heavy precipitation zone located in western Guizhou (Puding station), but the center of the precipitation zone in GPM/IMERG-L is shifted. Similar conclusions can be made for the distribution of heavy precipitation zones in western Hunan.

 doi:10.20944/preprints202408.1289.v1

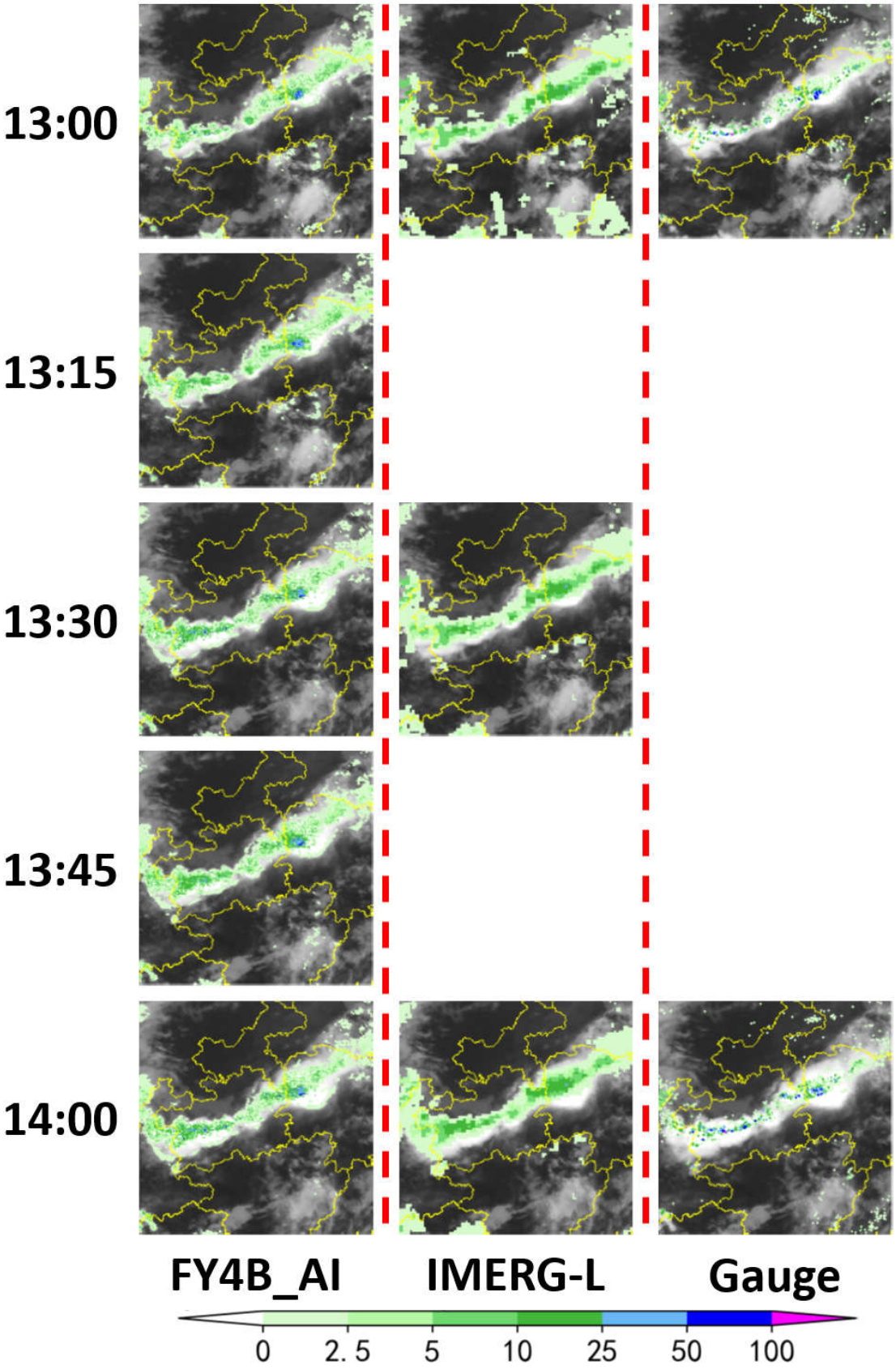


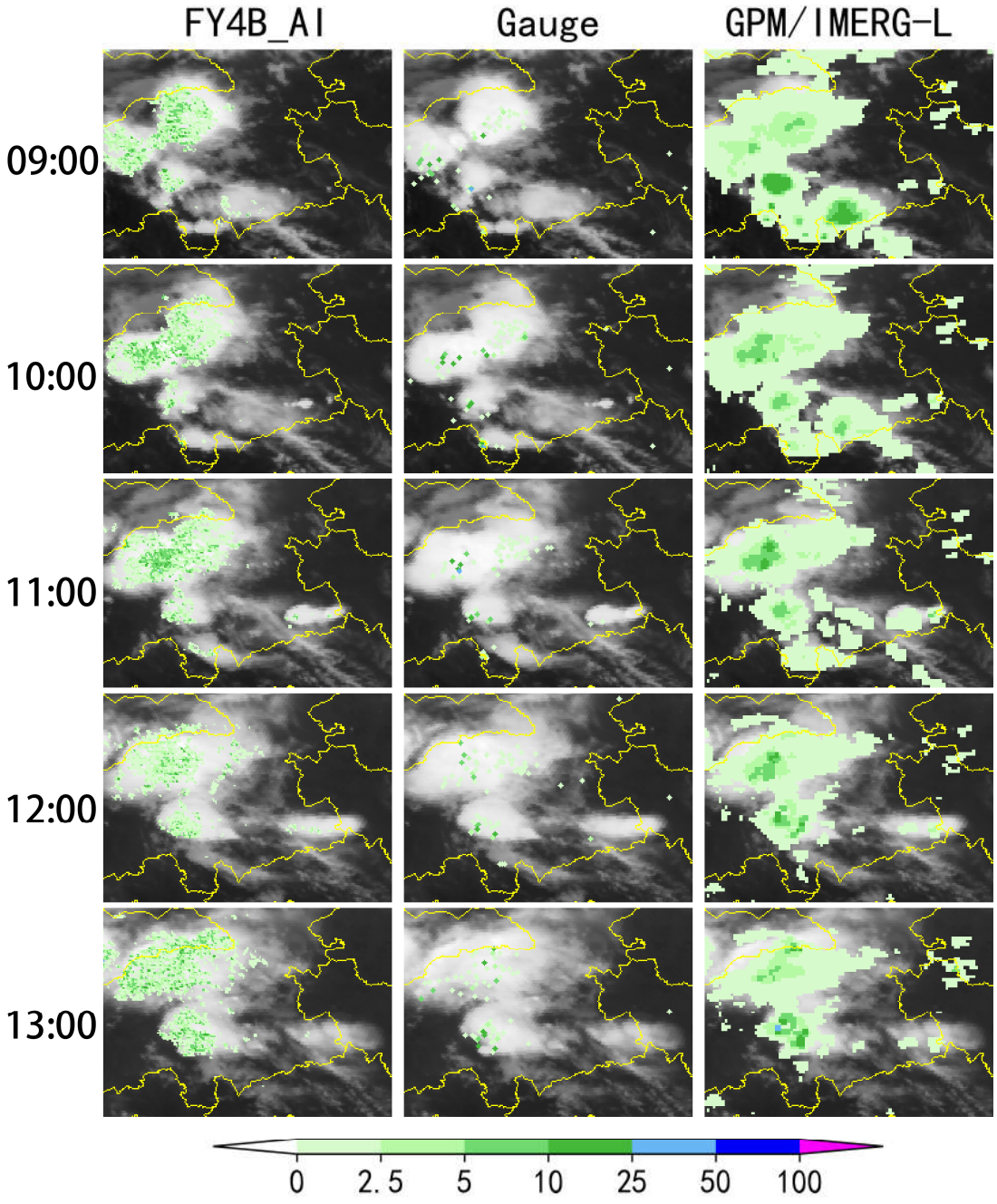
Figure 5. Overlay of FY-4B_AI, GPM/IMERG-L, Ground Station Truth Values, and Infrared Cloud Images from 13:00 to 14:00 (UTC) on June 18, 2023.

 doi:10.20944/preprints202408.1289.v1

5.2. Application Evaluation of the Strong Weather Event in Inner Mongolia on July 20, 2023 (Northwest Dry Region)

Starting from the late afternoon of July 20, 2023, several strong convective cloud clusters developed in central and southeastern Inner Mongolia. As they moved eastward, these clusters merged and intensified, bringing strong convective weather, such as short-duration heavy rainfall and 8-10 magnitude thunderstorm winds along the route.

The overlay of infrared cloud images and rainfall at five consecutive hours (09:00, 10:00, 11:00, 12:00, and 13:00) on July 20, 2023, for FY-4B_AI, ground station truth values, and GPM/IMERG-L are shown in Figure 6. The results indicate that in the sparsely populated eastern and central Inner Mongolia, FY-4B_AI again demonstrates its high spatial resolution characteristics. Both FY-4B_AI and GPM/IMERG-L accurately reproduce the spatial distribution characteristics of precipitation, but GPM/IMERG-L product has a bias towards larger range and magnitude compared to ground truth, and has some deviation in capturing the center of precipitation. The FY-4B_AI precipitation estimations are closer to the actual situation.



 doi:10.20944/preprints202408.1289.v1

Figure 6. Overlay of FY-4B_AI, Ground Station Truth Values, and GPM/IMERG-L with Infrared Cloud Images from 09:00 to 13:00 on July 20, 2023.

The evaluation results of the two strong weather events indicate that both FY-4B_AI and GPM/IMERG-L can accurately reproduce the spatial distribution characteristics of precipitation, no matter in the Southeast Humid Region or the Northwest Dry Region. However, due to its higher spatiotemporal resolution, FY-4B_AI not only provides more detailed distribution characteristics of precipitation but also updates more frequently, making it a valuable indicator of strong precipitation.

6. Conclusions and Discussions

This study investigated a multi-temporal FY-4B meteorological satellite precipitation estimation method based on artificial intelligence. It detailed the method of deriving features from satellite channel brightness temperature data and the selection method for training model data. This method was applied to estimate FY-4B meteorological satellite precipitation from January 1 to December 31, 2023, and the accuracy of the precipitation products was evaluated. Using ground-based 1-hour precipitation observation data within the study area as the verification truth, the study compared and analyzed two types of satellite-derived precipitation products, FY-4B_AI and GPM/IMERG-L, and conducted application assessments of two strong precipitation events. The main conclusions are as follows:

(1) In the multi-temporal FY-4B precipitation estimation algorithm based on artificial intelligence (FY-4B_AI), a total of 125 features were derived from five time points. This allows the precipitation estimation product to include not only satellite observation information at the current time point but also the changes in precipitation cloud radiation characteristics over the past hour, making the precipitation estimation results more consistent with the accumulated ground observation precipitation for one hour.

(2) Overall accuracy assessment revealed that among the eight evaluation indexes, six indexes of FY-4B_AI are better than those of GPM/IMERG-L, i.e., MAE, RMSE, RE, CC, POD, and CSI. Notably, the RE of FY-4B_AI is significantly lower than that of GPM/IMERG-L.

(3) The monthly distribution of precipitation accuracy showed that several indicators exhibit distinct annual variation characteristics. FY-4B_AI outperforms GPM/IMERG-L in MAE, RMSE, CC, and POD. Among these, MAE and RMSE are larger in the summer half-year and smaller in the winter half-year, suggesting a correlation between these features and seasonal variations in precipitation. The CC and POD also show FY-4B_AI outperforming GPM/IMERG-L, but their correlation with the season is not obvious. The ME and RE of FY-4B_AI are better than those of GPM/IMERG-L in winter precipitation.

(4) The spatial distribution of precipitation accuracy shows that FY-4B_AI is better than GPM/IMERG-L in 7 out of 8 evaluation indexes in both the Northwest Dry Region and the Southeast Humid Region. Notably, the relative error of FY-4B_AI is significantly lower than that of GPM/IMERG-L.

(5) Strong weather event application assessment revealed that both FY-4B_AI and GPM/IMERG-L accurately reproduce the spatial distribution characteristics of precipitation in both the Southeast Humid Region and the Northwest Dry Region. However, FY-4B_AI provides a more detailed distribution of precipitation. And since FY-4B_AI updates more frequently, it has a certain indicative significance for strong precipitation. Of course, we need more individual cases for verification.

This study examined a multi-temporal FY-4B meteorological satellite precipitation estimation method based on artificial intelligence (FY-4B_AI) and trained a precipitation inversion model for mainland China. This satellite precipitation inversion model is based on research results concerning precipitation inversion satellite channel radiation characteristics. By analyzing the channel characteristics of the multi-channel radiometer and its temporal changes regarding potential contributions to precipitation inversion, the study first constructed multiple derived features based on satellite channel brightness temperature. Then, it trained a regional precipitation estimation model

using the Light Gradient Boosting Machine (LGBM) algorithm, improving precipitation inversion efficiency. This algorithm continuously integrates the latest available ground precipitation data to update the model iteratively, without relying on simultaneous ground-based precipitation observations. This significantly enhances the effectiveness of the inversion product, holding crucial importance for disaster weather warnings.

Acknowledgments: We are grateful to Dr. Dapeng Huang of the National Climate Center for his help.

References

1. Xu R, Tian F Q, Yang L, et al. Ground validation of GPM IMERG and TRMM 3B42V7 rainfall products over the southern Tibetan Plateau based on a high-density rain gauge network[J]. Journal of Geophysical Research: Atmospheres, 2017, 122: 1-15.
2. Zambrano-bigiarini M, Nauditt A, Birkel C. Temporal and spatial evaluation of satellite-based rainfall estimates across the complex topographical and climatic gradients of Chile[J]. Hydrology and Earth System Science, 2017, 21: 1295-1320.
3. Baroentti A, Acquaotta F, Fratianni S. Rainfall variability from a dense rain gauge network in north-west Italy[J]. Climate Research, 2018, 75(3): 201-213.
4. Tang G Q, Behrangi A, Long D, et al. Accounting for spatiotemporal errors of gauges: A critical step to evaluate gridded precipitation products[J]. Journal of Hydrology, 2018, 559: 294-306.
5. Hou A Y, Kakar R K, Neeck S, et al. The global precipitation measurement mission[J]. Bulletin of the American Meteorological Society, 2014, 95(5): 701-722.
6. Huffman G J, Adler R F, Bolvin D T, et al. The TRMM multi-satellite precipitation analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scale[J]. Journal of Hydrometeorology, 2007, 8(1): 38-55.
7. Zhang, M.; Lu, N.; Gu, S.; Zhang, W., Temperature-sounding Microwave Channels for FY-3(02). Journal of Applied Meteorological Science 2012, 23 (2), 223-230.
8. Grody, N. C., Classification of snow cover and precipitation using the Special Sensor Microwave Imager. Journal of Geophysical Research: Atmospheres 1991, 96 (D4), 7423-7435.
9. Kummerow, C.; Barnes, W.; Kozu, T.; Shiue, J.; Simpson, J., The tropical rainfall measuring mission (TRMM) sensor package. Journal of atmospheric and oceanic technology 1998, 15 (3), 809-817.
10. Li, W.; Chen, Y.; Zhu, Y.; Zhao, B., Retrieval of rain over land by using TRMM/TMI measurements. Acta Meteorologica Sinica 2001, 59 (5), 591-601.
11. Li, X.; Yang, H.; You, R.; Zhao, F.; Qiao, Y., Remote sensing typhoon Songda's rainfall structure based on Microwave Radiation Imager of FY-3B satellite. Chinese Journal of Geophysics 2012, 55 (9), 2844-2853.
12. Li, P.; Xu, Z.; Ye, C.; Ren, M.; Chen, H.; Wang, J.; Song, S., Assessment on IMERG V06 Precipitation Products Using Rain Gauge Data in Jinan City, Shandong Province, China. Remote Sensing 2021, 13 (7), 1241.
13. 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. Bulletin of the American Meteorological Society 2014, 95 (5), 701-722.
14. Xie Shun, Sun Xiaogong, Zhang Suping, et al. Precipitation forecast correction in South China based on SVD and machine learning. J Appl Meteor Sci, 2022, 33(3): 293-304. DOI: 10.11898/1001-7313.20220304. (in Chinese)
15. ZHOU Kanghui, ZHENG Yongguang, HAN Lei, et al, 2021. Advances in Application of Machine Learning to Severe Convective Weather Monitoring and Forecasting. Meteorological Monthly, 47(3): 274-289. DOI: 10.7519/j.issn.1000-0526.2021.03.002.(in Chinese)
16. Li Dan, Lin Wen, Liu Qun, et al. Application of machine learning to statistical evaluation of artificial rainfall enhancement. J Appl Meteor Sci, 2024, 35(1): 118-128. DOI: 10.11898/1001-7313.20240110. (in Chinese)
17. ZHANG Yaping, ZHANG Yan, ZHAI Danhua, et al. 2022. Thoughts on application and improvement of deep learning in severe precipitation nowcasting technology [J]. Torrential Rain and Disasters, 41(5): 506-514. (in Chinese)
18. CHEN Yun, CAO Yong, SUN Jian, et al, 2021. Progress of Fine Gridded Quantitative Precipitation Forecast Technology of National Meteorological Centre. Meteorological Monthly, 47(6): 655-670. DOI: 10.7519/j.issn.1000-0526.2021.06.002.(in Chinese)

19. FANG Wei, SHEN Liang, ZOU Liyao, PANG Lin. 2023: Extrapolation method of precipitation nowcasting radar echo based on GCA-ConvLSTM prediction network. Torrential Rain and Disasters, 42(4): 427-436. DOI: 10.12406/byzh.2021-245. (in Chinese)
20. Moraux, A.; Dewitte, S.; Cornelis, B.; Munteanu, A., Deep learning for precipitation estimation from satellite and rain gauge measurements. Remote Sensing 2019, 11 (21), 2463.
21. Kidd, C.; Matsui, T.; Ringerud, S., Precipitation retrievals from passive microwave cross-track sensors: The precipitation retrieval and profiling scheme. Remote Sensing 2021, 13 (5), 947.
22. Liu, N., Ren, S., Jiang, J., Qin, D., & Han, B. (2023). AI-based estimation of precipitation in the Tibetan Plateau using multi-temporal FY-4A satellite data. International Journal of Remote Sensing, 44(20), 6523–6547. https://doi.org/10.1080/01431161.2023.2270112
23. Xian D. Fengyun-4B [J]. Application of satellite, 2021(7):4. DOI:10.3969/j.issn.1674-9030.2021.07.014.(in Chinese)
24. Huffman, G., GPM IMERG Late Precipitation L3 Half Hourly 0.1 degree x 0.1 degree V05. Goddard Earth Sciences Data and Information Services Center (GES DISC). 2015.
25. Wang J，Wang H J，Hong Y. 2016. Comparison of satellite-estimated and model-forecasted rainfall data during a deadly debris-flow event in Zhouqu, Northwest China. Atmos Oceanic Sci Lett，9（2）：139-145
26. He S S，Wang J，Wang H J. 2018. Hindcast study of "6.18" Mentougou debris flow event based on satellite rainfall and WRF forecasted rainfall. Chinese J Atmos Sci，42（3）：590-606 （in Chinese）
27. Zhang L L, Kang Y, Yue Q H, et al. 2021. Analysis of the applicability of various satellite-based precipitation in the source region of the Yellow River. Yellow River，43（3）：29-33 （in Chinese）
28. BO ZHAO, DAVID HUDAK et al. 2023, Assessment of IMERGv06 Satellite Precipitation Products in the Canadian Great Lakes Region. JOURNALOFHYDROMETEOROLOGY. 1017-1037. DOI: 10.1175/JHM-D-22-0214.1