Postprint of Applicability and Error Components of Multi-source Precipitation Products in Alpine Inland River Basins

Authors: Xu Liuxin, Wang Wenyu, Xiaoyan Wang, Wang Xueying, Gu Huanghe

Date: 2025-02-27T00:00:00+00:00

Abstract

The quality of precipitation data is a crucial factor affecting the accuracy of runoff simulation in alpine mountainous regions, which is essential for water resource management and ecological security. This study combines multiple statistical metrics and error decomposition models to conduct a comparative analysis of four precipitation products (satellite precipitation data GPM, the high-quality high spatiotemporal resolution precipitation dataset AIMERG for Asia, reanalysis data CMFD and ERA5) regarding their spatiotemporal distribution characteristics of precipitation in the upper Yarkant River basin, evaluates the accuracy of different products, and analyzes their error characteristics. The results show that: (1) The annual precipitation from CMFD and AIMERG exhibits a spatial pattern of high in the south and low in the north, consistent with the characteristics of the CN05.1 gridded dataset interpolated from Chinese ground stations, while ERA5 and GPM show the opposite spatial distribution. The high-resolution AIMERG and CMFD can capture the characteristic of high precipitation in the southwestern glacier region. (2) The interannual variation characteristics differ significantly among different precipitation products, and for most products, the proportion of precipitation in summer and autumn exceeds 60%. Comparative analysis reveals that only the AIMERG product can better represent the peak pattern and timing of intra-annual precipitation variation in the study area, demonstrates the strongest capability in capturing monthly precipitation at stations, showing higher correlation coefficients (>0.6) and smaller root mean square errors (8.45~11.57 mm); while the ERA5 product performs the worst. (3) The accuracy of different precipitation products at the daily scale all exhibits the characteristic of being higher during the rainy season (May–October) than during the dry season (November–April of the following year), with AIMERG showing higher Critical Success Index (CSI) for daily precipitation across different periods. (4) The dominant error for different precipitation products in summer is the hit error, while the dominant error in winter varies with the precipitation product. The research findings can provide valuable reference for runoff simulation in alpine regions and algorithm improvement of precipitation products.

Full Text

Arid Zone Research, Vol. 42 No. 1, January 2025

Applicability and Error Components of Multi-source Precipitation Products in Alpine Inland River Basins

XU Liuxin[12], WANG Wenyu[12], WANG Xiaoyan[12], WANG Xueying[12], GU Huanghe[12]

[1]. The National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing 210098, Jiangsu, China
[2]. College of Hydrology and Water Resources, Hohai University, Nanjing 210098, Jiangsu, China

Abstract

The quality of precipitation data is a critical factor influencing the accuracy of runoff simulation in high-cold mountainous regions, which is essential for water resource management and ecological security. This study combines multiple statistical metrics and an error decomposition model to compare and analyze the spatiotemporal distribution characteristics of precipitation in the upper Yarkant River basin using four precipitation products: satellite precipitation data (GPM IMERG), the Asian high-quality high spatiotemporal resolution precipitation dataset AIMERG, reanalysis data ERA5, and the China Meteorological Forcing Dataset (CMFD). The accuracy of different products was evaluated against observed precipitation data. The results indicate: (1) The spatial pattern of annual precipitation for CMFD and AIMERG shows higher values in the south and lower values in the north, consistent with the CN05.1 gridded dataset interpolated from surface meteorological stations, while ERA5 and GPM IMERG exhibit the opposite spatial distribution. High-resolution AIMERG and CMFD both capture the characteristic of higher precipitation in the southwestern glacier region. (2) The interannual variation characteristics differ significantly among precipitation products, with most products showing that summer and autumn precipitation accounts for over 60% of annual precipitation. Only AIMERG successfully reproduces the peak shape and timing of intra-annual precipitation variation at all stations, demonstrating the strongest capability to capture monthly precipitation at gauge stations with higher correlation coefficients (>0.6) and smaller root mean square errors (8.45–11.57 mm). (3) At the daily scale, all precipitation products exhibit higher accuracy during the wet period (May-October) than during the dry period (November–April), with AIMERG showing a consistently higher critical success index across different periods. (4) The dominant error for various precipitation products in summer is the hit error, while the dominant error in winter varies by product. These findings provide valuable references for runoff simulation and algorithm improvement of precipitation products in high-cold regions.

Keywords: reanalysis data; satellite precipitation; AIMERG; error decomposition model; upper Yarkant River

Introduction

Precipitation is a crucial component of the water cycle, and its spatiotemporal distribution significantly impacts water resources and ecological environments. However, in the high-cold mountainous areas of Northwest China, meteorological observation stations are sparse and unevenly distributed due to climatic constraints, topographical influences, and limited observation methods, making it difficult to characterize regional precipitation patterns. Therefore, identifying high-quality precipitation datasets suitable for these regions is fundamental for studying regional climate change and is of great significance for promoting research in water resource management and ecological environmental protection.

Satellite remote sensing and reanalysis systems are the primary means of obtaining large-scale, high spatiotemporal resolution precipitation information. Numerous scholars have conducted applicability analyses of different precipitation products in the Northwest China and Tibetan Plateau regions. Previous studies have shown that the GPM IMERG satellite precipitation product performs well in China, with accuracy superior to other products such as CMORPH. However, GPM IMERG still exhibits significant estimation errors in high-altitude areas, indicating that the application potential of current mainstream satellite precipitation products in mountainous regions needs further improvement. To address this, the Asian high-quality high spatiotemporal resolution precipitation dataset AIMERG was developed by calibrating GPM IMERG with APHRODITE data at the daily scale. AIMERG has been preliminarily proven to have higher accuracy than GPM IMERG in Northeast China, Southeast humid regions, and the Tianshan area.

While accuracy evaluation of precipitation products helps identify overall errors, analysis of different error components provides important theoretical guidance for product algorithm improvement and error correction. Current research on precipitation product error decomposition has focused primarily on individual satellite precipitation datasets such as GSMaP, with fewer comprehensive error analyses comparing multiple precipitation sources. Therefore, this study selects the typical inland river basin of the upper Yarkant River as the research area to compare and analyze the spatiotemporal characteristic differences among multiple precipitation products, evaluate their accuracy, and identify error sources using an error decomposition model. The research results can provide data references for runoff simulation in high-cold regions and offer insights for precipitation product algorithm improvement.

1.2 Data Sources

Ground observation precipitation data (2001–2010) were obtained from one national meteorological station (Tashkurgan) and two hydrological stations (Kaqun and Kulukelangan) for point-scale accuracy evaluation and error component analysis of precipitation products.

The CN05.1 dataset, based on interpolation of observational data from multiple national meteorological stations across China, is widely used in satellite precipitation and climate simulation evaluations. This dataset has a spatial resolution of 0.25° and a daily temporal step, obtained from the China Meteorological Data Sharing Network. Due to the scarcity of meteorological stations in the study area, the reliability of CN05.1 data in the upper Yarkant River was further verified. Previous research evaluated CN05.1 against station observations and found that the relative error of multi-year average precipitation ranged from -3% to -6%, with monthly precipitation correlations showing good consistency and correlation coefficients exceeding 0.6 at most stations. Given that CN05.1 has a longer time series (71 years) compared to most experimental stations in the region, this study selected CN05.1 as the ground reference dataset for spatial accuracy assessment.

GPM IMERG (Global Precipitation Measurement Integrated Multi-satellite Retrievals for GPM) is a product of the Global Precipitation Measurement mission jointly implemented by NASA and JAXA. Benefiting from the fusion of spaceborne microwave, infrared, and precipitation radar sensors, GPM IMERG achieves complementary advantages among multiple data sources. This study used GPM IMERG Version 06 data with a spatial resolution of 0.1°$×$0.1° and daily temporal resolution for the period 2001–2010.

ERA5 is the fifth-generation atmospheric reanalysis product from the European Centre for Medium-Range Weather Forecasts (ECMWF), representing an advancement over the ERA-Interim reanalysis. ERA5 integrates advanced modeling and data assimilation systems, offering high spatiotemporal resolution, rapid updates, and multiple parameters. The data have a spatial resolution of 0.25°$×$0.25° and hourly temporal resolution.

CMFD (China Meteorological Forcing Dataset) is a reanalysis dataset generated by fusing ERA-Interim reanalysis data, GLDAS data, and Princeton meteorological data as background fields with Chinese meteorological station precipitation data. The precipitation data in this dataset have a temporal resolution of 3 hours and a spatial resolution of 0.1°$×$0.1°.

AIMERG is a high-quality, high spatiotemporal resolution precipitation dataset for Asia, generated by calibrating GPM IMERG with APHRODITE data. Sourced from the National Tibetan Plateau Data Center, this dataset has a temporal step of 30 minutes and a spatial resolution of 0.1°$×$0.1°.

1.3.1 Evaluation Metrics

This study employs the correlation coefficient (CC), root mean square error (RMSE), bias (BIAS), probability of detection (POD), false alarm ratio (FAR), and critical success index (CSI) to evaluate the effectiveness of different precipitation products. The formulas for these metrics are:

$POD = \frac{H}{H+M}$

$FAR = \frac{F}{H+F}$

$CSI = \frac{H}{H+M+F}$

where H represents the number of precipitation events detected by both satellite products and ground stations, F represents events detected by satellite products but not by ground stations, and M represents events detected by ground stations but not by satellite products.

1.3.2 Error Decomposition Model

Error components of precipitation products can be assessed through the error decomposition model, which categorizes precipitation events into hit events, miss events, and false alarm events. Assuming these precipitation event types are independent, the total error (T) of a precipitation product can be decomposed into three error components: hit error (H), miss error (M), and false alarm error (F).

The definitions of different precipitation event types are summarized in .

The total error is calculated as:

$T = \frac{P_{ref} - P}{P_{ref}}$

where P_{ref} and P represent the total observed precipitation and precipitation product totals, respectively; P_H^{ref} and P_M^{ref} represent the observed precipitation totals for hit and miss events; and P_H and P_F represent the precipitation product totals for hit and false alarm events. The threshold value used in this study is 0.10 mm.

2.1 Spatiotemporal Distribution Characteristics of Precipitation Products

Research indicates that precipitation in mountainous areas exhibits strong seasonal variation. In the upper Yarkant River basin, precipitation from May to October accounts for over 90% of annual precipitation. Therefore, annual precipitation is divided into wet-period precipitation (May-October) and dry-period precipitation (November–April). The spatial distribution of different precipitation products reveals that CMFD and AIMERG show a south-high, north-low pattern, while ERA5 and GPM IMERG display the opposite spatial distribution. Most products show consistent distribution patterns between wet-period precipitation and annual precipitation, but dry-period precipitation characteristics vary considerably among products. Furthermore, different precipitation products exhibit varying abilities to describe the spatial heterogeneity of basin precipitation.

Lower spatial resolution products (ERA5 and CMFD) fail to reflect the higher precipitation amounts in glacier areas. In contrast, high-resolution AIMERG and GPM IMERG both capture the characteristic of higher precipitation in the southwestern glacier region, with AIMERG showing more pronounced spatial precipitation differences across the basin.

The interannual variation of annual precipitation time series (2001–2010) shows that most precipitation products exhibit relatively stable interannual changes. AIMERG shows an initial increase followed by a decreasing trend, while CN05.1 shows a decreasing then increasing trend. Significant differences exist in mean annual precipitation among the four products, with the highest from CMFD (456.02 mm) and the lowest from satellite precipitation data GPM IMERG (89.37 mm). Seasonal precipitation distribution reveals that most products have summer and autumn precipitation exceeding 60% of annual totals, with substantial differences in summer precipitation among products. CMFD reaches 190.40 mm, while GPM IMERG' s summer precipitation is only one-third of that amount.

2.2.1 Point-Scale Evaluation

Observed precipitation intra-annual patterns include unimodal distribution (Tashkurgan station) and bimodal distribution (Kaqun and Kulukelangan stations), with peak occurrence in July for the former and June–July for the latter. Among the products, only AIMERG successfully reproduces both the peak shape and timing of intra-annual precipitation variation across all stations. CMFD shows unimodal distribution consistent with Tashkurgan station but exhibits large fluctuations at the Kaqun and Kulukelangan hydrological stations, significantly differing from the observed bimodal characteristics. GPM IMERG and ERA5 show both unimodal and bimodal patterns across stations, with most products overestimating observed monthly precipitation, particularly ERA5.

To quantitatively reflect product performance, error metrics including correlation coefficient, root mean square error, and relative bias were calculated for 2001–2010. AIMERG shows higher correlation coefficients with observed monthly precipitation, while ERA5 significantly overestimates precipitation with relative errors ranging from 355.85% to 565.62% and substantially higher RMSE than other products. AIMERG demonstrates the smallest RMSE and, despite having fewer observation stations incorporated in its generation, shows the highest correlation coefficient and lowest RMSE at Tashkurgan station. Overall, AIMERG exhibits the strongest capability to reproduce monthly precipitation across stations.

2.2.2 Spatial Distribution Accuracy Assessment

In addition to point-scale evaluation, this study assesses product accuracy from a spatial distribution perspective. At the annual scale, AIMERG shows relative errors less than 20% at most stations, with the highest correlation coefficients exceeding 0.6 in most regions. CMFD exhibits the lowest spatial heterogeneity in monthly precipitation correlation coefficients. Other products show high spatial heterogeneity, with correlation coefficients reaching 0.4–0.6 in northern areas but less than 0.2 in high-elevation southwestern regions.

In the northern basin, most precipitation products show higher monthly correlation coefficients during the wet period than the dry period, while the southern region shows the opposite pattern. AIMERG consistently shows higher correlation coefficients across all periods compared to other products. Most products overestimate precipitation in more than half of the study area during the wet period, while AIMERG underestimates wet-period precipitation in most areas. Dry-period error characteristics are dominated by overestimation across most products.

2.3 Error Component Characteristics of Different Precipitation Products

The total and component errors of multi-source precipitation products across different seasons are illustrated in [Figure 8: see original paper]. Overall, GPM IMERG shows significantly high precipitation with relative errors exceeding those of other products. ERA5 shows generally low precipitation with mean errors around 1.96 mm, while other products exhibit maximum errors of 6.03 mm. AIMERG demonstrates lower errors with mean values of 2.88 mm.

The temporal variation characteristics of error components for the three precipitation products show obvious seasonal differences, particularly between summer and winter. To identify dominant errors, the most frequent maximum component error and its occurrence frequency were statistically analyzed for summer and winter. The results indicate that the dominant summer error for all products is the hit error, with frequencies between 0.51 and 0.62. In winter, dominant errors vary by product: global reanalysis data (ERA5 and CMFD) show false alarm errors as dominant, likely because reanalysis data contain many light rain events in winter, leading to poor detection of non-precipitation events. In contrast, the regionally enhanced AIMERG data show hit error as the dominant winter error.

3 Discussion

This study comprehensively evaluated four precipitation products (GPM IMERG, AIMERG, CMFD, and ERA5) in the upper Yarkant River basin across temporal and spatial scales. lists typical cases of applicability evaluation involving these products in alpine inland river basins. Numerous studies have focused on the applicability of GPM IMERG, finding its accuracy superior to other products (e.g., TRMM, CMORPH) but with underestimation in high-altitude areas. Comparative studies of different precipitation products in high-cold inland river source regions remain relatively scarce, with only a few conducted in the upper Ili River and Kaidu River basins.

Previous research found that CMFD overestimates precipitation at most stations in the Kaidu and Ili River headwaters, while GPM IMERG shows higher correlation and detection rates than CMORPH in the Ili River basin—consistent with our findings. In addition to the commonly evaluated products, this study included AIMERG for comparative analysis. Results demonstrate that AIMERG' s accuracy exceeds that of CMFD and ERA5 because AIMERG integrates more ground observation data. Similarly, other studies have shown AIMERG' s superior ability to reproduce precipitation characteristics over the Tibetan Plateau compared to other products.

AIMERG, a high-resolution daily gridded precipitation dataset for Asia generated by calibrating GPM IMERG with APHRODITE data, has proven superior to APHRODITE in the Tianshan region. Our study further focuses on the Yarkant River headwaters in the Kunlun Mountains and finds that AIMERG not only significantly improves daily precipitation detection capability during both wet and dry seasons compared to GPM IMERG but also better captures the intra-annual variation characteristics of monthly precipitation at stations than other products. This demonstrates AIMERG' s high potential for application in high-cold alpine regions at both daily and monthly scales.

4 Conclusions

Based on multiple statistical metrics and an error decomposition model, this study investigated the applicability and error characteristics of four precipitation products (GPM IMERG, AIMERG, CMFD, and ERA5) in the upper Yarkant River basin. The main conclusions are:

(1) The spatial patterns of annual precipitation for CMFD and AIMERG show higher values in the south and lower values in the north, consistent with CN05.1, while ERA5 and GPM IMERG exhibit the opposite distribution. Compared with other products, AIMERG and CMFD show higher spatial heterogeneity and can both capture the characteristic of higher precipitation in the southwestern glacier region.

(2) From a point-scale perspective, only AIMERG successfully reproduces the peak shape and timing of intra-annual precipitation variation at all sta-

tions and shows higher critical success indices for daily precipitation during both wet and dry periods. From a spatial perspective, AIMERG shows the highest monthly precipitation correlation coefficients, exceeding 0.6 in most regions.

(3) Overall, the temporal variation characteristics of error components for the four precipitation products show significant seasonal differences, particularly between summer and winter. The dominant error across products in summer is the hit error, while dominant errors in winter vary substantially among products.

These findings provide valuable references for runoff simulation in high-cold regions and for algorithm improvement of precipitation products.

References

[References are preserved as numbered citations throughout the text but not fully listed in the provided material.]

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.