Multisource precipitation data fusion: Generating high-quality precipitation estimates

Hanqing Chen,[1,2] Jiangyuan Zeng,[3,*] Yi Lyu,[1] and Bin Yong[1,*]
[1]National Key Laboratory of Water Disaster Prevention, Hohai University, Nanjing 210098, China
[2]College of Ocean Engineering and Energy, Guangdong Ocean University, Zhanjiang 524005, China
[3]State Key Laboratory of Remote Sensing and Digital Earth, Aerospace Information Research Institute, Chinese Academy of Sciences, Beijing 100101, China
*Correspondence: zengjy@radi.ac.cn (J.Z.); yongbin@hhu.edu.cn (B.Y.)
Received: October 21, 2025; Accepted: January 12, 2026; Published Online: January 14, 2026; https://doi.org/10.59717/j.xinn-geo.2026.100195

Citation: Chen H., Zeng J., Lyu Y., et al. (2026). Multisource precipitation data fusion: Generating high-quality precipitation estimates. *The Innovation Geoscience* **4**:100195.

Dear Editor,

As a key component of Earth's water cycle, precipitation intricately links the atmosphere, land, and ocean. This central role makes obtaining high-quality, global precipitation estimates imperative for advancing our understanding of hydrological processes. The primary methods for measuring precipitation include rain gauges, weather radars, satellite retrievals, and atmospheric reanalysis (Figure 1).[1] Rain gauges provide accurate point-scale precipitation measurements, but yield high uncertainties in sparsely instrumented regions. Weather radars offer area-averaged precipitation, yet their effectiveness is limited by spatial coverage constraints (~250 km in radius), sensitivity to topographic and structural interference, and high deployment costs, which hinder their scalability. Thus, achieving high-quality global precipitation remains challenging with these two methods.[2]

Satellite- and model-based techniques provide hydrometeorologists with spatiotemporally continuous precipitation across global scales, particularly in ungauged regions.[3] Nevertheless, these precipitation products often contain significant errors and uncertainties, making accurate regional and global precipitation estimation challenging. With the intensification of global change, many fields such as climate change, water cycle, and disaster monitoring demand increasingly accurate precipitation measurements. More importantly, global warming leads to higher atmospheric water content, resulting in more intense rainfall and frequent natural disasters, underscoring the need for high-precision precipitation data.[4] Multisource precipitation fusion, which leverages the complementary strengths from different datasets, is one of the most effective methods for generating high-quality precipitation estimates. In this paper, we briefly introduce main multisource precipitation fusion strategies, discuss their constraints, and offer perspectives on their future development (Figure 1).

STRATEGIES

Numerous studies have proposed various multisource precipitation fusion methods to meet the high-accuracy demands of different operational applications. The multisource precipitation fusion methods can be roughly divided into three categories: (1) point-grid two-source fusion, (2) weighting schemes, and (3) artificial intelligence (AI) techniques.[5] These approaches reliably produce high-quality precipitation estimates.

Among the three categories, point-grid two-source fusion approaches, such as bias correction and multiple regression analysis, leverage point-based ground precipitation observations in conjunction with gridded satellite/model data to fit statistical relationships or correction models. These fitted statistical relationships/models are used to enhance the quality of satellite/model data for specific areas. Classic bias correction methods are widely used to adjust satellite-only precipitation estimates, using point-based ground observations as a benchmark. Relying on these high-quality point-based ground precipitation observations, they produce research-quality global precipitation products such as Integrated Multi-satellitE Retrievals for Global Precipitation Measurement (GPM) Final Run (IMERG-Final), Global Satellite Mapping of Precipitation-Gauge Calibrated (GSMaP-Gauge), and Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks-Climate Data Record (PCDR). The performance of bias correction methods is influenced by the quality of ground data and the spatial heterogeneity of errors in the gridded precipitation estimates. Multiple regression analysis, such as geographically weighted regression and empirical statistical models, uses statistically explainable relationships between errors and multiple impacting factors (e.g., topography, climate type, precipitation intensity, seasonality, and distance) to correct precipitation estimates. While these methods partially account for error characteristics in precipitation estimates, they also highly depend on the quality of point-based ground precipitation data. In summary, the effectiveness of point-grid two-source fusion approaches depends heavily on the quality of point-based ground precipitation observations. Weighting schemes merge multiple precipitation datasets by assigning precise weights to each source, generating high-accuracy merged precipitation estimates. Since this approach does not correct errors in the input data, the resulting fusion quality is consequently contingent on the initial quality of the input sources. Although AI-based fusion techniques function as 'black boxes' with limited interpretability, they are powerful tools for merging various precipitation data sources, particularly in addressing complex nonlinear relationships without requiring explicit statistical models.[6-7]

CONSTRAINTS

While existing multisource precipitation fusion methods have markedly enhanced precipitation accuracy, several critical constraints remain, covering data sources and fusion methods (Figure 1).

Limitations of the data sources

Large errors and the scarcity of rain gauges. The quality of the inputs is key to that of the merged results. Low-quality inputs, especially when the quality between different inputs varies significantly, can distort or mask effective precipitation information, thereby undermining the effectiveness of fusion methods. Existing gridded precipitation products exhibit significant errors and uncertainties across most regions globally (Figure 1).[8] These limitations constitute a major obstacle to effectively improving the accuracy of fused precipitation products. Additionally, precipitation data from dense rain gauge networks are among the most reliable observations and consequently play a crucial role in improving fusion accuracy. However, station observation coverage remains scarce or inadequate across most global regions, particularly in remote oceans and complex mountainous areas (Figure 1).[9] The absence of gauge observations leads to limited fusion accuracy in these areas.

Limitations in the data types of inputs. Existing multisource precipitation fusion methods primarily merge three types of precipitation data (i.e., satellite-, model-, and ground-based products) and auxiliary variables (e.g., topographic indices, Normalized Differential Vegetation Index, soil moisture, temperature, and latitudinal/longitudinal data) to reduce various types of errors like random errors and systematic errors. Although most methods have attempted to improve the accuracy of precipitation, the quality of merged precipitation estimates still needs significant improvement, especially over ungauged areas and in reducing precipitation intensity-related errors, e.g., precipitation overestimation/underestimation for low/high rain rates, Root Mean Square Error (RMSE) increasing with the increase of precipitation intensity. Introducing new data sources could potentially resolve this issue.

Large uncertainties in fusion methods

Large uncertainties in the weights. Obtaining optimal weights is critical for multisource precipitation fusion. Each weighting method has its pros and cons, making it challenging to derive optimal weights through any single

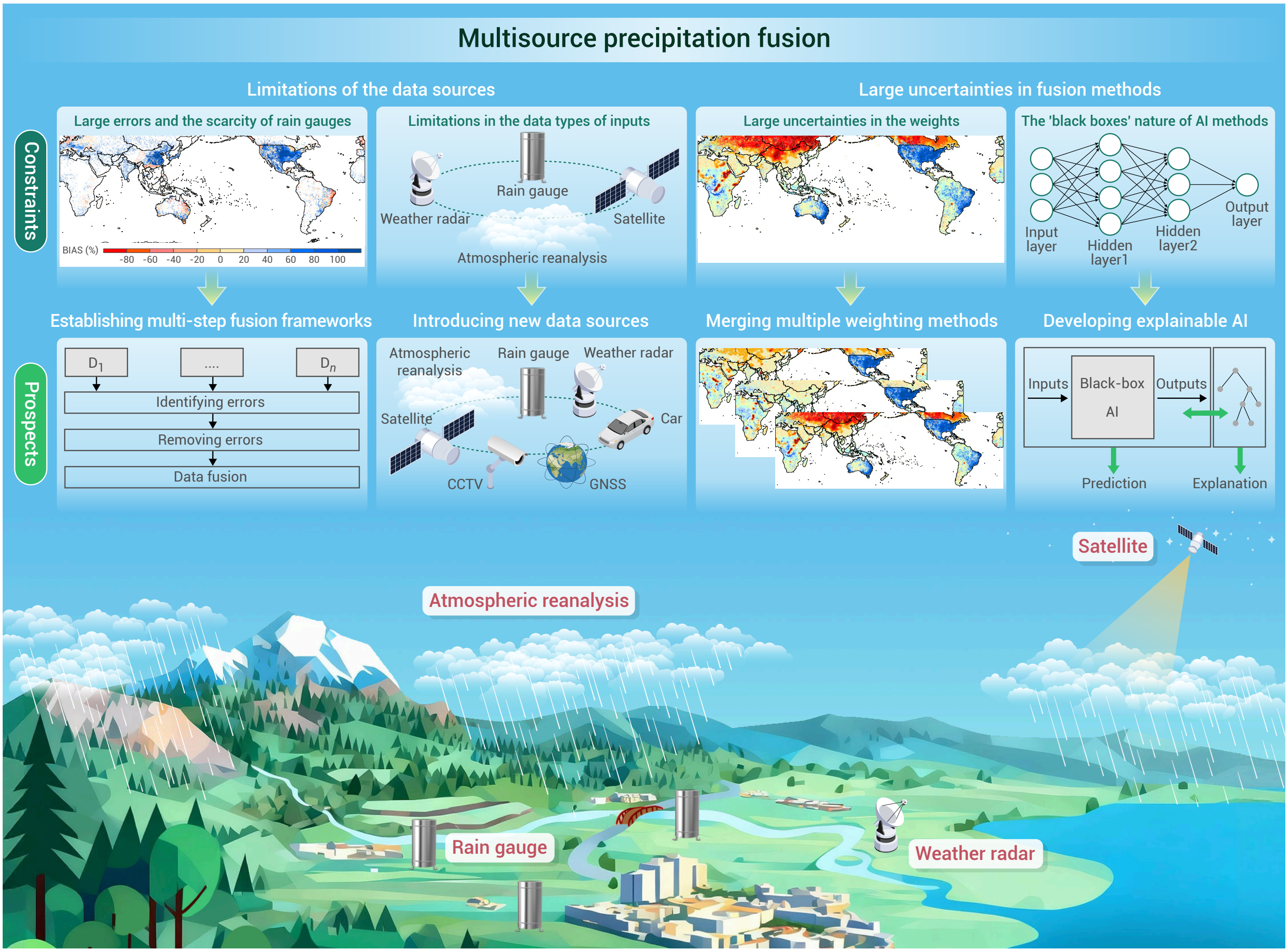

Figure 1. Multisource precipitation fusion: Measurements, Constraints and Prospects.

approach alone. In particular, without dense ground precipitation data for reference, any single weighting method fails to determine optimal weights for different input sources in sparse regions, which inevitably compromises the accuracy of the final fused precipitation product.

The 'black boxes' nature of AI methods. Although AI-based methods are powerful tools for merging different precipitation data sources, their limited interpretability poses a challenge. This lack of transparency makes it difficult for researchers to understand the relationships between inputs and merged results, as well as to identify and correct potential errors or biases in the model's output. The performance of AI-based methods also highly depends on the quality and quantity of training data. Insufficient or biased training data can lead to models that do not generalize well to new conditions, limiting their applicability in diverse and ungauged regions.

PROSPECTS

To address these constraints, we suggest the following directions for enhancing multisource precipitation fusion (Figure 1).

Establishing multi-step fusion frameworks. To reduce the errors in gridded input data prior to multi-source data fusion, future efforts should focus on developing multi-step fusion frameworks. These frameworks employ a step-by-step approach to detect and eliminate errors in gridded input data, followed by data fusion. By identifying and removing input errors (e.g., hit errors, false alarms, and missed precipitation) before fusion while retaining meaningful precipitation information, the negative impacts of low-quality inputs on the final merged results will be significantly reduced.

Introducing new data sources. Introducing new data sources is an effective way to further improve the accuracy of merged results. For instance, precipitable water vapor from global navigation satellite systems can provide critical input for precipitation fusion, particularly in ungauged regions. Additionally, the wiper frequency from moving cars can measure precipitation and reduce intensity-related errors through valid relationships between wiper frequency and rainfall intensity. More importantly, Closed-Circuit Television (CCTV) surveillance cameras also represent a viable technique for precipitation monitoring, particularly in data-sparse regions where conventional gauges are lacking. High-resolution spatiotemporal data from CCTV surveillance cameras and urban vehicle fleets enable the generation of accurate urban waterlogging maps.

Merging multiple weighting methods. A Multi-method fusion scheme can derive optimal weights by synergistically integrating the complementary advantages of distinct weighting approaches. The theoretically optimal weights for different precipitation products are defined as those that minimize the overall error. Using ground precipitation as the reference in gauged areas and the triple collocation/ generalized three-cornered hat method in ungauged areas, the RMSE is calculated. Subsequently, the optimal weights for various precipitation products are determined through an iterative optimization process aimed at minimizing the RMSE. This scheme is especially advantageous in ungauged regions, where the absence of ground-based references renders individual weighting methods prone to high uncertainty.

Developing explainable AI. Explainable AI can help researchers understand the internal decision-making process of AI models, increasing confidence in their predictions and enabling the identification and correction of potential errors. Furthermore, integrating domain knowledge and physical constraints into AI models can substantially enhance model accuracy and physical consistency. For example, the water balance equation (Precipitation – Evapotranspiration – Runoff = Storage change) can be embedded in an AI architecture by introducing a physics-guided loss term that penalizes violations of mass conservation, or by designing intermediate network layers that explicitly predict evapotranspiration, runoff, and storage change before combining them to satisfy the balance constraint. With the increasing availability of big data from various sources, such as high-resolution satellite products, radar networks, and atmospheric reanalysis, future AI models should be designed to effectively leverage these massive datasets with diverse data types and scales, improving their applicability in diverse and ungauged regions.

SUMMARY AND OUTLOOKS

This paper provides an overview of the current status, constraints, and prospects of multisource precipitation fusion. We have identified two pressing challenges in this field: one pertaining to data sources, and the other to fusion methods. To address these challenges, future efforts will focus on: (1) establishing multi-step fusion frameworks to mitigate the adverse impacts of low-quality inputs on merged results; (2) introducing new data sources to enhance the accuracy of merged results, particularly in ungauged areas; (3) merging multiple weighting methods to determine optimal weights; (4) developing explainable AI models to improve transparency. The advancements are expected to thoroughly enhance precipitation accuracy, thereby benefiting various research fields such as hydrology, meteorology, and ecology.

REFERENCES

1. Sun Q., Miao C., Duan Q., et al. (2018). A review of global precipitation data sets: Data sources, estimation, and intercomparisons. *Rev. Geophys.* **56**:79–107. DOI:10.1002/2017RG000574
2. Hou A. Y., Kakar R. K., Neeck S., et al. (2014). The global precipitation measurement mission. *Bull. Am. Meteorol. Soc.* **95**:701–722. DOI:10.1175/BAMS-D-13-00164.1
3. You Y., Huffman G., Kidd C., et al. (2023). Evaluating and improving TROPICS millimeter - wave sounder's precipitation estimate over ocean. *J. Geophys. Res.-Atmos.* **128**:e2023JD038697. DOI:10.1029/2023JD038697
4. Salzmann M. (2016). Global warming without global mean precipitation increase. *Sci. Adv.* **2**:e1501572. DOI:10.1126/sciadv.1501572
5. Zhu S., Ma Z., Xu J., et al. (2022). A morphology-based adaptively spatio-temporal merging algorithm for optimally combining multisource gridded precipitation products with various resolutions. *IEEE Trans. Geosci. Remote Sens.* **60**:4103221. DOI:10.1109/TGRS.2021.3097336
6. Lei H., Zhao H., and Ao T. (2022). A two-step merging strategy for incorporating multi-source precipitation products and gauge observations using machine learning classification and regression over China. *Hydrol. Earth Syst. Sci.* **26**:2969–2995. DOI:10.5194/hess-26-2969-2022
7. Jiang Y., Yang K., Qi Y., et al. (2023). TPHiPr: A long-term (1979–2020) high-accuracy precipitation dataset (1/30°, daily) for the Third Pole region based on high-resolution atmospheric modeling and dense observations. *Earth Syst. Sci. Data* **15**:621–638. DOI:10.5194/essd-15-621-2023
8. Chen H., Yong B., Kirstetter P. E., et al. (2021). Global component analysis of errors in three satellite-only global precipitation estimates. *Hydrol. Earth Syst. Sci.* **25**:3087–3104. DOI:10.5194/hess-25-3087-2021
9. Kidd C., Becker A., Huffman G. J., et al. (2017). So, how much of the Earth's surface is covered by rain gauges. *Bull. Am. Meteorol. Soc.* **98**:69–78. DOI:10.1175/BAMS-D-14-00283.1

FUNDING AND ACKNOWLEDGMENTS

This research has been sponsored by the National Natural Science Foundation of China (grant no. 42422106, U2243229, 42201029) and the program for scientific research start-up funds of Guangdong Ocean University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

Hanqing Chen: Conceptualization, Funding acquisition, visualization, writing-original draft. Jiangyuan Zeng: Conceptualization, Funding acquisition, writing-original draft. Yi Lyu: Writing-review and editing. Bin Yong: Conceptualization, Writing-review and editing. All authors contributed to the manuscript and approved the final version.

DECLARATION OF INTERESTS

The authors declare no competing interests.