atmosphere

MDPI

Article

Bayesian Model Averaging Method for Merging Multiple Precipitation Products over the Arid Region of Northwest China

Yong Yang [1,*], Rensheng Chen [1], Xinyu Lu [2,3], Weiyi Mao [2,3], Zhangwen Liu [1] and Xueliang Wang [4]

[1] Qilian Alpine Ecology and Hydrology Research Station/State Key Laboratory of Ecological Safety and Sustainable Development in Arid Lands, Northwest Institute of Eco-Environment and Resources, Chinese Academy of Sciences, Lanzhou 730000, China; crs2008@lzb.ac.cn (R.C.); zwliu@lzb.ac.cn (Z.L.)
[2] Institute of Desert Meteorology, China Meteorological Administration, Urumqi 830002, China; luxy@idm.cn (X.L.); maowiyi@idm.cn (W.M.)
[3] Field Scientific Experiment Base of Akdala Atmospheric Background, Xinjiang Wulanwusu Ecological and Agrometeorological Field Scientific Observation and Research Station/Wulanwusu Agrometeorological Experimental Station, Urumqi 830002, China
[4] School of Hydraulic Engineering, Lanzhou Resource and Environment Vocational and Technical University, Lanzhou 730021, China; wangxueliang@nieer.ac.cn
* Correspondence: yy177@lzb.ac.cn

Abstract

Accurate precipitation estimation is essential for hydrological modeling and water resource management in arid regions; however, complex terrain and sparse meteorological station networks introduce substantial uncertainties into gridded precipitation datasets. This study evaluates the performance of nine widely used precipitation products in the arid region of Northwest China (ARNC) at both the meteorological station scale and the sub-basin scale, and applies the Bayesian Model Averaging (BMA) approach to merge multi-source precipitation estimates. The results reveal pronounced spatial heterogeneity and significant differences in performance among datasets, with the Integrated Multi-Satellite Retrievals for the Global Precipitation Measurement mission performing best at the station scale and the Famine Early Warning Systems Network Land Data Assimilation System performing best at the sub-basin scale. Compared with individual products, the BMA-merged precipitation demonstrates substantial improvements at both scales, providing higher coefficients of determination and agreement indices, and lower relative mean absolute error and relative root mean square error, indicating enhanced accuracy and robustness. The BMA-merged precipitation product generally exhibits superior and more spatially consistent performance than the individual datasets across the ARNC, thereby providing a more reliable basis for regional hydrological and climate-related applications. The merged dataset shows that the mean annual precipitation in the ARNC during 2000–2024 is approximately 230.4 mm, exhibiting a statistically significant increasing trend of 1.4 mm per year, with the strongest increases occurring in the Tianshan and Qilian Mountains. This study provides a reliable foundation for hydrological modeling and climate-change assessments in data-limited arid environments.

Keywords: precipitation; arid region of Northwest China; precipitation products; Bayesian model averaging; precipitation merging; evaluation

Academic Editor: Merhala Thurai

Received: 23 November 2025
Revised: 26 December 2025
Accepted: 14 January 2026
Published: 16 January 2026

1. Introduction

Arid and semi-arid regions account for approximately 41% of the global land surface and support the livelihoods of hundreds of millions of people worldwide [1,2]. Recent

global assessments emphasize the extensive distribution of drylands and their high ecological and socioeconomic vulnerability under accelerated climate warming [3]. In these regions, water resources form the foundation of ecosystem sustainability, agricultural production, and human wellbeing, where persistent water scarcity constitutes a critical constraint for development [4,5]. Precipitation represents the primary source of renewable water in arid environments and governs both the magnitude and temporal variability of water availability [6]. Even small variations in precipitation intensity, frequency, or phase can substantially alter runoff generation, soil moisture dynamics, and vegetation growth [7,8]. In many arid regions with limited storage capacity, spatial and temporal heterogeneity in precipitation directly triggers hydrological extremes such as droughts and flash floods [7,9]. Moreover, precipitation exhibits pronounced spatial heterogeneity due to complex terrain effects [10], making the development of high-accuracy precipitation datasets essential for hydrological modelling, water-balance assessment, and water resource management in arid regions [11,12].

Direct measurements from the gauge observations at meteorological stations remain one of the most reliable and widely accepted sources of precipitation data [13]. However, in arid regions with complex and highly variable terrain—such as the arid region of Northwest China (ARNC)—station networks are unevenly distributed, with most gauges located in low-elevation oasis areas while vast high-elevation and desert regions remain largely unmonitored, leading to limited representativeness of station data for the entire region [14]. In recent years, satellite retrievals and reanalysis products have provided gridded precipitation data that help overcome the sparsity of gauge networks and enable the characterization of spatial patterns and temporal variability across extensive arid regions [15,16]. Nevertheless, substantial spatial and temporal uncertainties persist in gridded precipitation datasets due to multiple sources of error. For satellite retrievals, uncertainties arise from retrieval algorithm limitations, assumptions in cloud microphysics, and sensor sampling deficiencies [16,17]. For reanalysis datasets, errors mainly stem from deficiencies in model parameterizations, biases in assimilated observations, and elevation-dependent representation errors [18,19]. Therefore, a comprehensive understanding and rigorous evaluation of these uncertainties are essential for the development of high-accuracy precipitation datasets that support hydrological modelling and water-resource management in arid regions [16,20].

In recent years, numerous studies have systematically evaluated gridded precipitation products across arid regions worldwide—including Australia [21], Pakistan [12], Iran [22], Saudi Arabia [23], and ARNC [24]—typically using in situ gauge observations as the benchmark in a point-to-pixel validation framework. However, due to sparse and unevenly distributed station networks in arid regions, where gauges are predominantly located in oasis or easily accessible low-elevation areas while vast high- elevation mountains, plateaus, and desert regions remain largely unobserved, gauge-based point validation often fails to adequately represent the regional performance of gridded precipitation datasets. An alternative evaluation strategy that has recently gained increasing attention is basin-scale water-balance assessment, which offers a complementary and spatially integrated perspective on precipitation product performance in arid regions [25]. Based on the principle of water balance, regional precipitation can be estimated using evapotranspiration, runoff, and total water storage (TWS), thereby enabling the assessment of gridded precipitation products at the regional scale [26,27].

Evaluation results across different regions reveal that the performance of precipitation products exhibits substantial spatial variability. For instance, Islam et al. [21] reported that Tropical Rainfall Measuring Mission (TRMM) Multi-satellite Precipitation Analysis outperforms Precipitation Estimation from Remotely Sensed Information using Artificial Neural

Networks (PERSIANN) in Australia, whereas Dehaghani et al. [22] found that PERSIANN-cloud classification system performs better than TRMM in Iran. Even at relatively small spatial scales, such as the Heihe River Basin in the ARNC (approximately 142,900 km^2), the performance of different precipitation products varies within the basin. The Multi-source weighted-ensemble precipitation (MSWEP) shows the highest overall accuracy across the basin, while PERSIANN, Climate Research Unit dataset, and the fifth generation of European ReAnalysis perform best in the upstream, midstream, and downstream regions, respectively [28]. Overall, no single precipitation product performs optimally across all regions, and each exhibits distinct advantages and limitations depending on regional climatic and geographic conditions [25].

To integrate the strengths of precipitation datasets from different sources, this study introduces the multi-model averaging (MMA) approach to merge multiple precipitation products. MMA is a statistical ensemble technique that combines outputs from multiple models or data sources to generate estimates that are generally more accurate and robust than those from any individual product [29,30]. Beyond the simplest equal-weight averaging approach, commonly used MMA techniques include information-theoretic methods such as Akaike and Bayesian Information Criterion averaging [31,32], as well as regression-based optimization approaches such as Bates and Granger Averaging and its Granger–Ramanathan A, B, and C variants [33,34]. Among the various multi-model averaging approaches, the Bayesian model averaging (BMA) has gained particular attention due to its solid probabilistic foundation and flexibility in integrating multiple data sources. The BMA framework assumes that the true state of a variable can be represented as a weighted combination of multiple model outputs, where the weights correspond to the posterior probabilities that each model is the most plausible representation given the observations [35,36].

Previous studies have demonstrated that BMA can effectively improve the accuracy of hydrological predictions, including streamflow forecasting, groundwater simulation, drought assessment, evapotranspiration estimation, and precipitation merging [37–40]. Comparative studies of various MMA methods have shown that the BMA approach is one of the most effective ensemble techniques for hydrological simulation [41–43]. Yang et al. [25] applied 8 MMA methods to the merging of multi-source precipitation products in the Qilian Mountains and ultimately recommended the BMA approach. Based on these findings, this study adopts the BMA method to merge multi-source precipitation products in the arid regions, with the aim of further expanding its application potential in hydrological research of arid environments.

The main objectives of this study are: (1) to evaluate the performance of multiple precipitation products at both the meteorological station scale and the sub-basin scale across the ARNC; (2) to merge multi-source precipitation products using the BMA method and evaluate its performance in the ARNC; and (3) to analyze the spatiotemporal characteristics of precipitation over the ARNC based on the merged dataset.

2. Materials and Methods

2.1. Study Area

The ARNC (Figure 1), which covers the Xinjiang Region, the Hexi Corridor of Gansu, the northeastern edge of Qinghai and western Inner Mongolia, spans approximately between 73 and 107° E and 35–50° N, accounts for about 25% of China's land area [44]. Located in the heart of the Eurasian continent, it is one of the driest regions at its latitude in the world. The landform exhibits a pronounced "mountain-basin interleaved" pattern, with east–west trending mountain ranges from north to south, including the Altay, Tianshan, and Karakoram–Kunlun–Altun–Qilian belts, interspersed with large inland basins

such as the Junggar, Tarim, and Qaidam basins. The region exhibits large elevation differences across short distances, ranging from −152.31 m asl (Aydingkol Lake) to 8611 m asl (Mount Qogir) [45]. Under a temperate continental climate, the region is extremely dry, with sparse vegetation and low annual precipitation exhibiting strong spatial variability. Precipitation generally increases with elevation, making high mountain regions act as "wet islands" and major sources of runoff, while most rivers are internally drained and have short flow paths [46]. As a key region for water security and ecological stability in western China [47], the ARNC, with its complex terrain and sparse observational coverage, serves as an ideal natural testbed for multisource precipitation data merging and provides valuable implications for improving regional hydrological modeling, enhancing water-resource management, and advancing understanding of water-cycle processes in global arid regions.

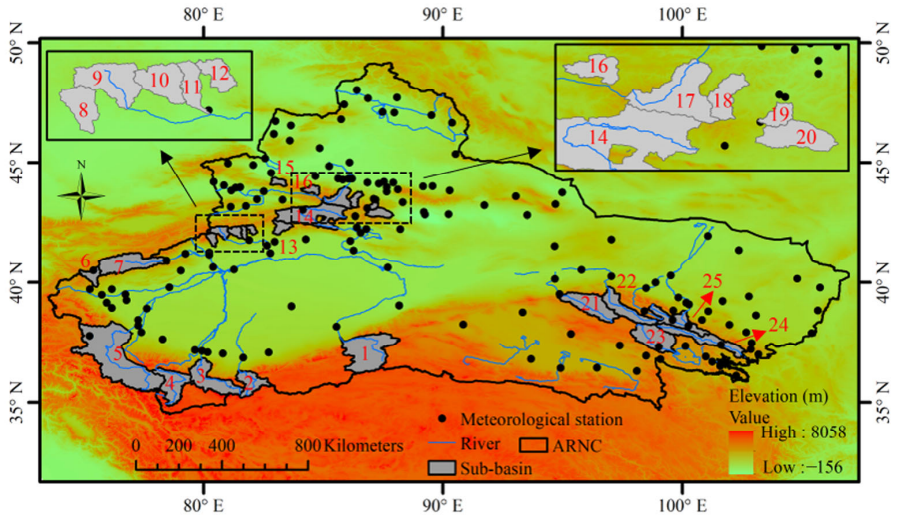

Figure 1. Spatial distribution of meteorological stations and sub-basins in the arid region of Northwest China (ARNC). (Numbers in the sub-basins are the serial number of sub-basins, see in Table 1 for details).

Table 1. Features of 25 sub-basins in the arid region of Northwest China.

Number	Sub-Basin	Hydrologic Station	Basin Area (km^2)	Runoff Data Period
1	Cheercen	Qiemo	24,780.0	2000–2011
2	Keliya	Keliya	8141.9	2000–2011
3	Yulongkash	Tongguziluoke	14,923.6	2000–2011
4	Kalakash	Tuoman	12,484.2	2000–2011
5	Yarkand	Kaqun	48,624.7	2000–2011
6	Qiakemake	Qiaqiga	3333.5	2000–2011
7	Tuoshigan	Shaliguilanke	14,822.6	2000–2011
8	Tailan	Tailan	1547.3	2000–2011
9	Muzhati	Pochengzi	2604.0	2000–2011
10	Kamuslang	Kamuluk	1852.0	2000–2011
11	Taileweiqiuke	Baicheng	1137.3	2000–2011
12	Kalasu	Kalasu	1038.0	2000–2011
13	Kuqa	Langan	2625.0	2000–2011
14	Kaidu	Dashankou	17,191.8	2000–2011
15	Jinghe	Jinghe Shankou	1406.0	2000–2011
16	Kuitun	Jiangjunmiao	1749.4	2000–2011
17	Manas	Kensiwate	5183.4	2000–2011
18	Hutubi	Shimen	1827.9	2000–2020
19	Urumqi	Yingxiongqiao	922.1	2000–2011
20	Alagou	Alagou	2874.5	2000–2011
21	Danghe	Dangchengwan	14,374.6	2000–2020
22	Shule	Changmabao	10,948.5	2000–2020
23	Buha	Buha	14,469.0	2000–2020
24	Datong	Tiantang	12,488.1	2000–2020
25	Heihe	Yingluoxia	10,018.1	2000–2020

2.2. Data Collection

2.2.1. In Situ Observation Precipitation Data

According to the Specifications for Surface Meteorological Observation (2003) [48] issued by the China Meteorological Administration (CMA), precipitation is measured using a standard rain gauge installed 70 cm above the ground with an inner diameter of 20 cm. Observations are conducted at 08:00 and 20:00 each day, recording the precipitation amount for the preceding 12 h. For liquid precipitation, the water collected in the storage bottle is poured completely into a measuring cylinder for reading. During winter snowfall, the rain-receiving funnel is replaced with a snow-receiving funnel, and the storage bottle is removed so that solid precipitation is collected directly in the outer cylinder. During observation, the outer cylinder containing snow or ice is taken indoors, where the solid precipitation is melted and measured with a measuring cylinder, or the cylinder with snow is weighed using a designated balance, and the empty cylinder weight (or corresponding millimeters) is subtracted to obtain the snowfall amount.

The ground-based meteorological dataset used in this study was sourced from the CMA (http://data.cma.cn/, accessed on 11 July 2022). The precipitation observations provided by the CMA are compiled by the National Meteorological Information Center and undergo stringent quality-control procedures, including screening for extreme values, checking internal consistency, and removing suspicious records [49]. These gauge-based datasets have been extensively applied in evaluating precipitation products and in various hydrometeorological studies over China and its subregions [14,25,50,51]. Daily precipitation measurements for the period 2000–2020, were subsequently aggregated to a monthly scale for consistency in this study. In total, 152 meteorological stations distributed across the ARNC (Figure 1) were utilized to assess the accuracy of multiple gridded precipitation products.

2.2.2. Grided Precipitation Data

1. Climate Hazards group infrared precipitation with stations (CHIRPS; CHIRPS V2).

The CHIRPS dataset was jointly developed by the Climate Hazards Group and the U.S. Geological Survey to support drought monitoring and long-term trend analysis. It combines the Climate Hazards Group Precipitation Climatology, high-resolution infrared satellite imagery, satellite-only precipitation estimates, and multiple gauge-based observations [52]. CHIRPS V2 offers both daily and monthly data at a spatial resolution of $0.05^\circ \times 0.05^\circ$. In this study, the monthly dataset was used. A summary of CHIRPS and other gridded precipitation datasets is provided in Table 2.

2. Climate prediction center (CPC) morphing technique (CMORPH; CMORPH V1).

CMORPH precipitation estimates are bias-corrected and reprocessed on a global 8 km × 8 km grid (60° S–60° N) with a temporal resolution of 30 min. The product is generated through two main steps: initial recalibration of raw satellite CMORPH estimates followed by bias correction using probability density function matching against CPC daily gauge observations over land and GPCP pentad analyses over oceans [53]. In this work, the high-frequency data were aggregated to monthly totals for analysis.

3. Land component of the fifth generation of European ReAnalysis (ERA5-Land).

Produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) under the Copernicus Climate Change Service, ERA5-Land represents an enhanced land component of the fifth generation of European ReAnalysis (ERA5) reanalysis. It retains the main physical parameterizations of ERA5 but improves the global horizontal resolution to 0.1°, compared with 0.25° in ERA5 [54]. The monthly ERA5-Land reanalysis data were utilized in this study.

Table 2. Summary of datasets used in this study.

Name	Abbreviation	Resolution	Period	Data Access	Last Accessed
CHIRPS V2	CHIRPS	0.05°	1981–NRT	https://data.chc.ucsb.edu/products/CHIRPS-2.0	20 October 2025
CMORPH V1	CMORPH	8 km	1998–NRT	https://doi.org/10.5065/0EFN-KZ90	10 August 2025
ERA5-Land	ERA5-Land	0.1°	1950–NRT	https://doi.org/10.24381/cds.68d2bb30	25 October 2025
FLDAS_NOAH01_C_GL_M	FLDAS	0.1°	1982–NRT	https://doi.org/10.5067/5NHC22T9375G	8 July 2025
GLDAS_NOAH025_M	GLDAS	0.25°	2000–NRT	https://doi.org/10.5067/SXAVCZFAQLNO	25 September 2025
GPM_3IMERGM V07	IMERG	0.1°	2000–NRT	https://doi.org/10.5067/GPM/IMERG/3B-MONTH/07	27 August 2025
MERRA V5.12.4	MERRA	0.5°	1980–NRT	https://doi.org/10.5067/0JRLVL8YV2Y4	29 October 2025
MSWEP V3.15	MSWEP	0.1°	1979–NRT	http://www.gloh2o.org/mswep/	20 September 2025
WorldClim	WorldClim	2.5′	1960–NRT	https://www.worldclim.org/data/monthlywth.html	28 September 2025
GLEAM v4.2a	GLEAM	0.1°	1980–NRT	https://www.gleam.eu/	6 December 2025
GRACE-CSR RL06 V3		0.25°	2002–NRT	https://www2.csr.utexas.edu/grace/RL06_mascons.html	27 August 2025
GRACE-GSFC RL06 V2	GRACE	0.5°	2002–NRT	https://earth.gsfc.nasa.gov/geo/data/grace-mascons	27 August 2025
GRACE-JPL RL06 V4		0.5°	2002–NRT	https://grace.jpl.nasa.gov/data/get-data/jpl_global_mascons/	27 August 2025

Note. NRT stands for near-real-time.

4. Famine early warning systems network land data assimilation system (FLDAS; FLDAS_NOAH01_C_GL_M).

The FLDAS dataset provides global monthly land-surface variables simulated using the Noah and VIC land-surface models [55]. With a 0.1° spatial resolution and temporal coverage from 1982 to the present, the Noah-driven version (3.6.1) was selected for this analysis.

5. Global land data assimilation system (GLDAS; GLDAS_NOAH025_M_2.1).

GLDAS-2.1 employs the Princeton meteorological forcing dataset and produces a consistent record from 2000 to the present. Monthly precipitation data were derived by averaging 3-hourly simulations from the Noah Model (version 3.6) implemented within the Land Information System (LIS) version 7. The dataset has a spatial resolution of 0.25° × 0.25° [56].

6. Integrated multi-satellite retrievals for global precipitation measurement (IMERG; GPM_3IMERGM V7).

The Global Precipitation Measurement (GPM) mission is an international collaboration involving a core satellite and approximately ten partner satellites. IMERG integrates, intercalibrates, and merges all microwave and infrared satellite precipitation estimates, along with gauge analyses and other potential sources, to generate precipitation data at high spatial and temporal resolutions [57]. This study used the monthly GPM_3IMERGM V7 data with a 0.1° × 0.1° resolution.

7. Modern-Era Retrospective Analysis for Research and Applications (MERRA; MERRA-2).

MERRA-2 introduces several improvements over the original MERRA system, notably the assimilation of observation-based precipitation products to correct land surface precipitation outside high-elevation regions [58]. Produced by NASA's Global Modeling and Assimilation Office (GMAO) using the Goddard Earth Observing System (GEOS) Model version 5.12.4, MERRA-2 provides global atmospheric reanalysis data from 1980 to the present, with a spatial resolution of 0.5° × 0.625°.

8. Multi-Source Weighted-Ensemble Precipitation (MSWEP; MSWEP V3.15).

MSWEP is a globally consistent, machine learning–based precipitation dataset available from 1979 onward, with a 0.1° spatial resolution [59]. It integrates information from reanalysis outputs, satellite retrievals, and gauge measurements to ensure reliable precipitation estimates across diverse climatic regimes [59]. In this study, the latest version (V3.15) of the monthly MSWEP dataset was applied.

9. WorldClim (WorldClim V2.1).

Gridded Climatic Research Unit (CRU) Time-Series (TS) data, were produced by CRU at the University of East Anglia and funded by the UK National Centre for Atmospheric Science, was first published in 2000. CRU TS dataset is derived by the interpolation of monthly climate anomalies from extensive networks of weather station observations. WorldClim provides historical monthly weather data in 2.5′ × 2.5′ spatial resolution for 1950–2024, which are downscaled from CRU TS v4.03 using WorldClim 2.1 for bias correction [60].

2.2.3. Evapotranspiration, Total Water Storage, and Runoff Data

1. Global land evaporation amsterdam model (GLEAM; GLEAM V4.2a).

The GLEAM is a suite of algorithms designed to estimate the individual components of terrestrial evapotranspiration, including plant transpiration, bare-soil evaporation, canopy interception loss, open-water evaporation, and sublimation [61]. Numerous studies have demonstrated the robustness and reliability of the GLEAM dataset across diverse regions and research applications [62,63]. The dataset has been widely used for evapotranspiration quantification, hydrological model forcing, global climate variability assessment, and climate model evaluation [62,64,65]. In this study, the latest version, GLEAM V4.2a [64], with a spatial resolution of 0.1° × 0.1°, was employed.

2. Gravity recovery and climate experiment (GRACE; GRACE-Center for space research (CSR), GRACE-Goddard space flight center (GSFC), and GRACE-Jet propulsion laboratory (JPL)).

Since its launch in 2002, the Gravity Recovery and Climate Experiment (GRACE) mission, together with its follow-on (GRACE-FO), has provided invaluable observations of global variations in total water storage (TWS) [66]. Although GRACE data are characterized by relatively coarse spatial resolution, they remain highly effective for investigating basin-scale water balance dynamics [67–69]. The latest Release 06 (RL06) mascon solutions are available from three processing centers: the Center for Space Research (CSR) at the University of Texas, NASA's Goddard Space Flight Center (GSFC), and the Jet Propulsion Laboratory (JPL). To minimize product uncertainty, ensemble mean values from these three centers [25,70] were adopted in this study.

3. Measured river runoff data.

A total of 25 sub-basins were selected to assess the performance of different precipitation datasets at the basin scale (Figure 1; Table 1). Monthly runoff data from hydrological gauging stations were obtained from the Hydrographic Yearbooks of the People's Republic of China.

2.3. Water Balance

According to the principle of water balance, the basin precipitation can be expressed as:

$$P = ET + R + \Delta TWS \quad (1)$$

where P, ET, R represent precipitation, evapotranspiration and runoff, respectively (mm). ΔTWS denotes the change in the total water storage (mm).

2.4. Evaluation Criteria

To identify potential overestimation or underestimation in different precipitation products, the mean error (ME) was applied:

$$ME = \frac{1}{n}\sum_{i=1}^{n}(X_{si} - X_{oi}) \tag{2}$$

where X_s and X_o denote precipitation products data and observation data, respectively. n is the number of data points. Given large differences in precipitation magnitude across stations, the relative mean error (RME) was calculated to reduce scale effects:

$$RME = ME/\overline{X}_o \times 100\% \tag{3}$$

Positive RME values indicate overestimation, whereas negative values indicate underestimation.

To quantitatively assess the performance of precipitation products, several statistical indicators were employed, including the coefficient of determination (R^2), mean absolute error (MAE), root mean square error (RMSE), and index of agreement (d), which are defined as:

$$R^2 = [cov(X_s, X_o)/\sigma X_s \sigma X_o]^2 \tag{4}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|X_{si} - X_{oi}| \tag{5}$$

$$RMSE = \sqrt{\left(\sum_{i=1}^{n}(X_{si} - X_{oi})^2\right)\Big/ n} \tag{6}$$

$$d = 1 - \left[\frac{\sum_{i=1}^{n}(X_{si} - X_{oi})^2}{\sum_{i=1}^{n}\left(|X_{si} - \overline{X}_o| + |X_{oi} - \overline{X}_o|\right)^2}\right], 0 \le d \le 1 \tag{7}$$

where *cov* and σ are the covariance and standard deviation, respectively. Relative MAE (RMAE), and relative RMSE (RRMSE) were further calculated:

$$RMAE = MAE/\overline{X}_o \times 100\% \tag{8}$$

$$RRMSE = RMSE/\overline{X}_o \times 100\% \tag{9}$$

The rank score (RS) method [25,71] was used to evaluate overall performance:

$$S_i = \begin{cases} (x_i - x_{i,min})/(x_{i,max} - x_{i,min}) & for\ RMAE \text{ and } RRMSE \\ (x_{i,max} - x_i)/(x_{i,max} - x_{i,min}) & for\ R^2 \text{ and } d \end{cases} \tag{10}$$

where i represents the ith products, and $x_{i,min}$ and $x_{i,max}$ represent the products of minimum and maximum values of each evaluation criterion. The sum of RS values across the four criteria yields the total RS (range 0–4), where lower values indicate better performance.

2.5. Bayesian Model Averaging Method

In this study, the nine precipitation datasets were treated as an ensemble, and the BMA approach was applied to merge them into probabilistic precipitation estimates for the ARNC. The application of BMA in hydrological and meteorological studies have been extensively documented [25,38–40]. BMA determines the weight of each ensemble member based on its probability distribution function (PDF). Denote y as the variable to be predicted, $D = [y_1^o, y_2^o, ..., y_T^o]$ as the observed series with a length of T, and $f = [f_1, f_2, ..., f_N]$ as

the ensemble of series from precipitation products data. According to the law of total probability, the PDF of the BMA probabilistic prediction of y can be represented as:

$$P(y|D) = \sum_{i=1}^{N} P(f_i|D) \cdot P_i(y|f_i, D) \quad (11)$$

where $P(y|f_i, D)$ is the posterior predictive distribution of y on the condition of the observation data D and each individual precipitation product f_i. $P(f_i|D)$ is posterior probability, and represents the likelihood that a simulation is the right simulation. It can also be seen as the weight, $w_i = P(f_i|D)$. The posterior mean and variance of the BMA prediction can be expressed as:

$$E[y|D] = \sum_{i=1}^{N} P(f_i|D) \cdot E[P_i(y|f_i, D)] = \sum_{i=1}^{N} w_i f_i \quad (12)$$

$$Var[y|D] = \sum_{i=1}^{N} w_i \left(f_i - \sum_{k=1}^{N} w_k f_k\right)^2 + \sum_{i=1}^{N} w_i \sigma_i^2 \quad (13)$$

where E denotes the mathematical expectation operator, σ_i^2 is the variance associated with precipitation product f_i with respect to observation D. The EM algorithm was applied to estimate weights [37], and the Box–Cox transformation was applied to satisfy Gaussian assumptions [38,72].

2.6. Trend Analysis

The non-parametric Mann–Kendall test [73–75] was applied to assess the significance of precipitation trends at a confidence level of α = 0.05 (95% confidence level). Sen's slope method [76] was used to quantify the trend magnitude, indicating both direction and steepness.

To maintain consistent spatial resolution, all datasets were resampled to 2.5′ using bilinear interpolation. Considering data availability, monthly station-scale evaluations were conducted for 2000–2020, while sub-basin evaluations followed the hydrological record periods (Table 2). The BMA weights were calibrated independently at each meteorological station using observed precipitation during the training period (2000–2020). This station-wise calibration strategy enables the BMA framework to account for spatial heterogeneity in the performance of individual precipitation products across the arid and topographically complex ARNC. Subsequently, the station-based BMA weights were interpolated to continuous gridded fields using the inverse distance weighting method. A power parameter of 1 was adopted, such that the influence of a station decreases linearly with increasing distance. In the interpolation process, no strict distance threshold was imposed, allowing surrounding stations to jointly constrain the spatial distribution of BMA weights, which enhances spatial smoothness and robustness in data-sparse regions. The resulting gridded weights were then applied to generate the BMA-merged precipitation dataset over the ARNC for the period 2000–2024.

3. Results

3.1. The Precipitation Values from 9 Products

The spatial distribution characteristics of the multi-year mean precipitation derived from 9 different datasets indicate that low-precipitation areas in the ARNC are mainly located in the low-elevation desert regions, while high-precipitation areas are primarily distributed in the high-elevation mountainous regions, such as the Tianshan and Qilian Mountains (Figure 2). Although the spatial patterns of the products are generally consistent, substantial discrepancies exist in their estimated precipitation amounts. The maximum difference in area-averaged annual precipitation across the ARNC reached 2.37-fold, with the highest estimate from MERRA (286.2 mm) and the lowest from MSWEP (120.8 mm). Based on the magnitude of annual precipitation, the products are ranked as follows: MERRA >

ERA5-Land > FLDAS > GLDAS > CMORPH > IMERG > CHIRPS > WorldClim > MSWEP. Figure S1 shows that trends in area-averaged precipitation during 2000–2024 also varied among the nine datasets. CHIRPS, ERA5-Land, FLDAS, and WorldClim exhibited decreasing trends, none of which were statistically significant, while the remaining five products showed increasing trends, among which only CMORPH was not statistically significant ($p > 0.05$). Overall, substantial discrepancies exist among the products in estimating precipitation over the ARNC, underscoring the necessity of evaluating their applicability.

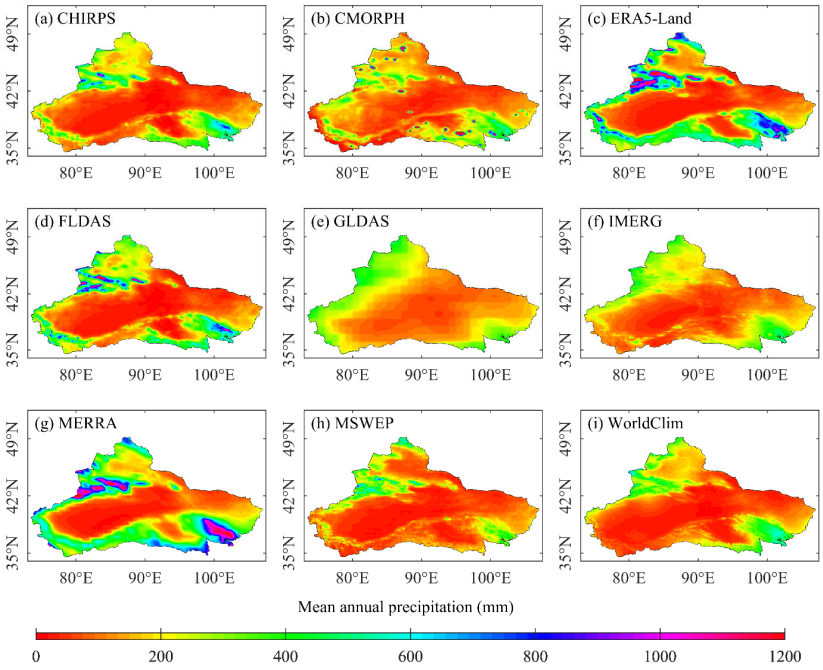

Figure 2. Spatial distribution of mean annual precipitation in the arid region of Northwest China for 2000–2024 determined from the data of 9 products.

3.2. Point-to-Pixel Evaluation Using Observational Data from Meteorological Stations

3.2.1. Evaluation of the Precipitation Products

The mean error (ME) analysis shows that none of the nine precipitation products consistently overestimate or underestimate precipitation across all 152 meteorological stations in the ARNC. Overall, most products exhibit a greater number of stations with precipitation overestimation than underestimation (Figure 3). Among them, ERA5-Land overestimates precipitation at 138 stations, with an average overestimation of 52%, representing the product with the largest number of overestimating stations. In contrast, MSWEP and WorldClim tend to underestimate precipitation, with 86 and 87 stations exhibiting underestimation, and mean values of −27% and −20%, respectively. The spatial pattern of ME also varies substantially among products; for instance, CMORPH and GLDAS overestimate precipitation around the Tarim Basin yet underestimate it in high-elevation regions such as the Tianshan and Qilian Mountains, whereas MERRA consistently overestimates precipitation in the Qilian Mountains.

To further evaluate the accuracy of 9 precipitation products, the R^2, RMAE, RRMSE and d values between each product and observations from 152 meteorological stations were calculated. Figure 4 presents the spatial distribution of RRMSE, while Figures S2–S4 show the spatial distributions of R^2, RMAE, and d. Spatially, most products perform better in northern Xinjiang and the Qilian Mountains, characterized by higher R^2 and d values and lower RMAE and RRMSE values, whereas performance is generally poorer in southern Xinjiang. As shown in Figure 4, MSWEP and WorldClim provide the highest overall performance, whereas CMORPH, GLDAS, and MERRA show the weakest performance. Based

on the four evaluation criteria, the comprehensive performance ranking of precipitation products at the meteorological station scale is as follows: IMERG > MSWEP > ERA5-Land > WorldClim > CHIRPS > FLDAS > GLDAS > MERRA > CMORPH.

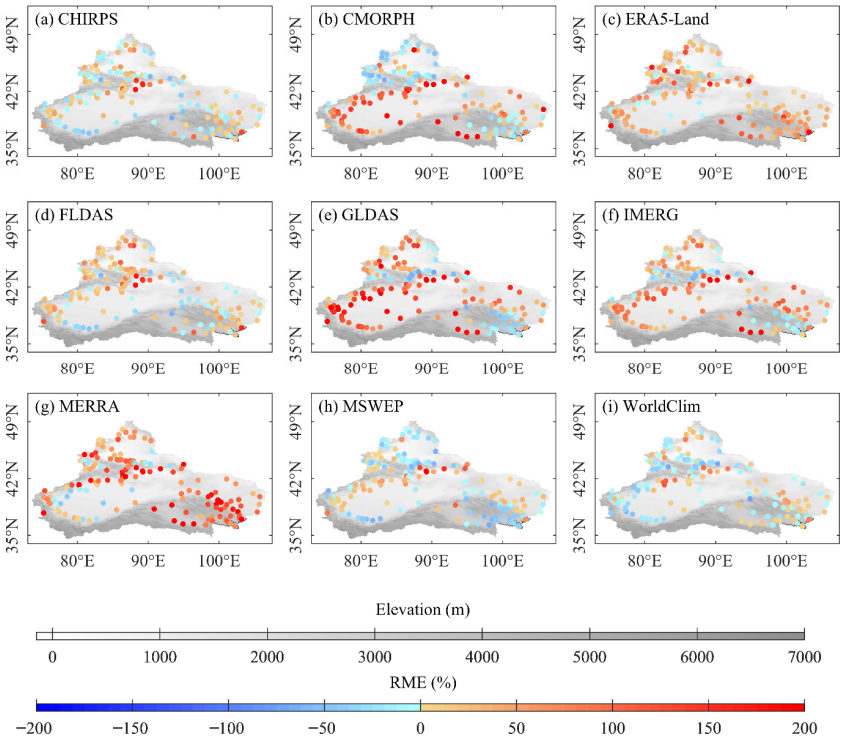

Figure 3. Spatial distribution of relative mean error (RME) between 9 products' monthly data and observational monthly data obtained at 152 meteorological stations in the arid region of Northwest China.

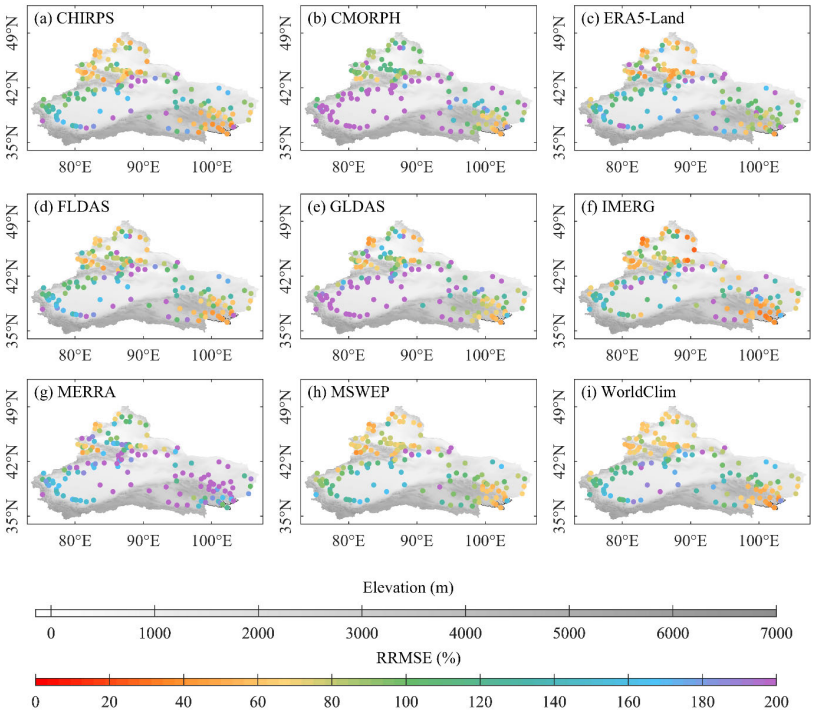

Figure 4. Spatial distribution of relative root mean square error (RRMSE) between 9 products' monthly data and observational monthly data obtained at 152 meteorological stations in the arid region of Northwest China.

3.2.2. Performance of the Bayesian Model Averaging Method

Using precipitation data from the nine products and observations from 152 meteorological stations, the Bayesian model averaging method was applied to generate BMA-merged precipitation. RME analysis indicates that the merged dataset overestimates precipitation at 131 stations and underestimates it at 21 stations, with a mean RME of 34.0%. Figure 5 shows the spatial distribution of differences in RRMSE between the nine individual products and the BMA-merged precipitation. Compared with each single product, the BMA-merged precipitation substantially reduces RRMSE at most stations, with IMERG showing the fewest improvements (103 stations) and CMORPH showing the most (151 stations). The evaluation metrics demonstrate that BMA-merged precipitation exhibits overall superior performance relative to individual products, evidenced by higher median R^2 and d values, lower median RMAE and RRMSE values, and reduced interquartile dispersion (Figure 6). Figure S5 shows that BMA performed better than IMERG at 88 stations, and outperformed CMORPH and MERRA at 151 stations, based on the RS ranking results across all stations. Overall, the BMA-generated precipitation demonstrates significantly higher accuracy at the station scale across the ARNC.

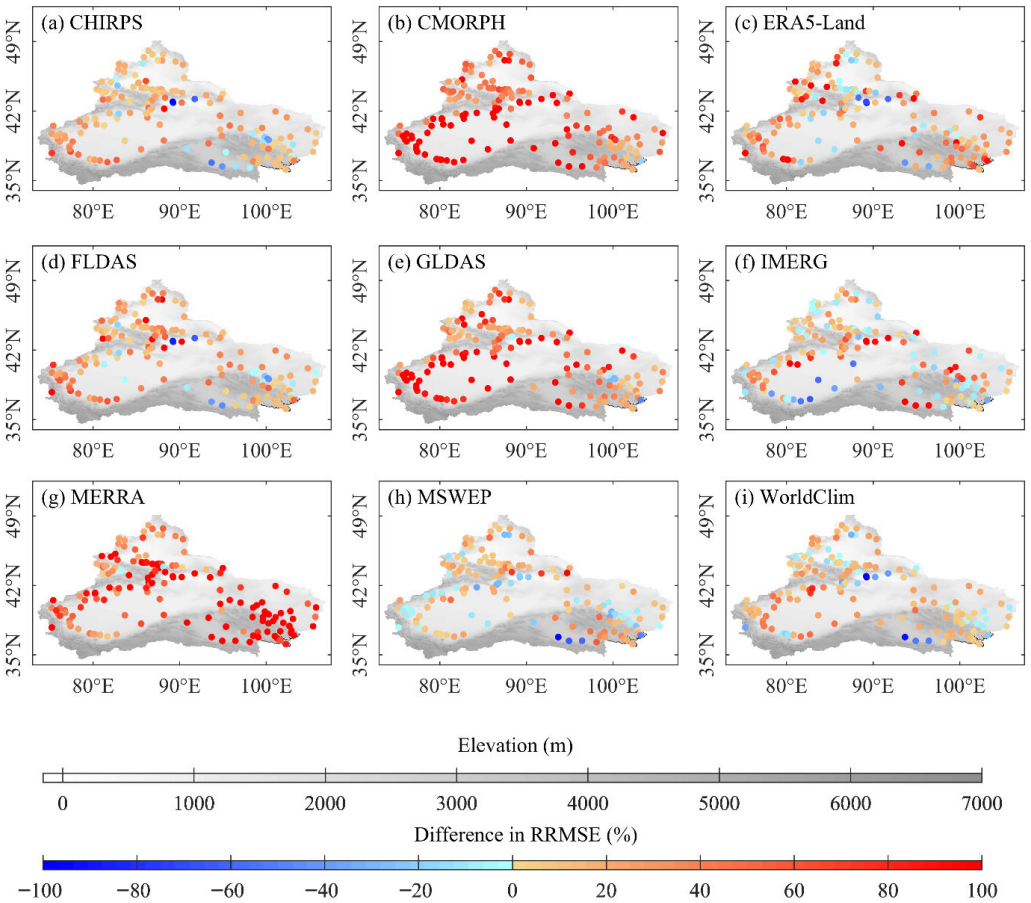

Figure 5. Spatial distribution of the differences in relative root mean square error (RRMSE) between the nine precipitation products and the merged precipitation from the Bayesian model averaging method, both evaluated against observations at 152 meteorological stations, in the arid region of Northwest China.

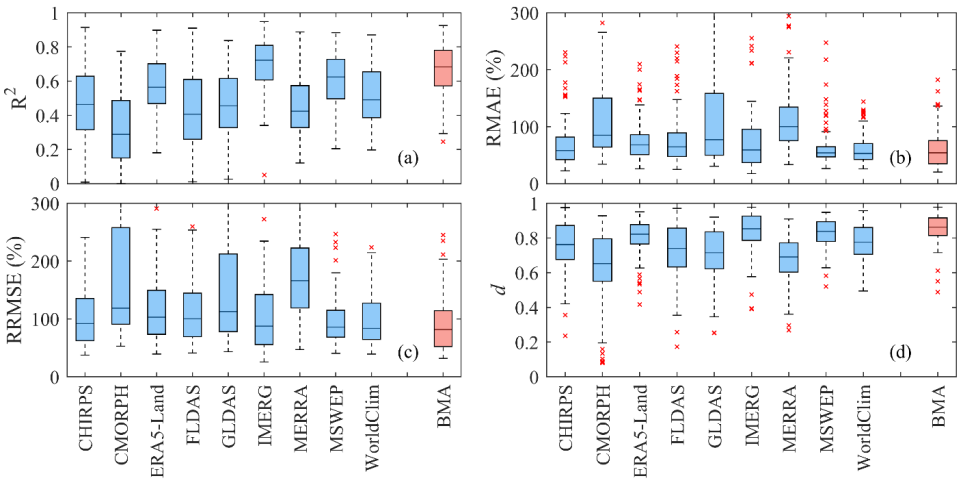

Figure 6. Box plots of R^2 (**a**), RMAE (**b**), RRMSE (**c**), and *d* (**d**) values over 152 meteorological stations for 9 products and Bayesian model averaging method. The black line in each boxplots represents the median value, and the bottom and top edges of each box indicate the 25th and 75th percentiles, respectively. The whiskers represent the maximum and minimum values that are not considered outliers, where outliers are more than 1.5 standard deviations from the mean and are shown as red crosses.

3.3. Sub-Basin Evaluation Using Water Balance Method

3.3.1. Evaluation of the Precipitation Products

Annual precipitation estimated for 25 sub-basins using the water balance method (Equation (1)) was employed to evaluate the performance of the nine precipitation products at the basin scale. The ME analysis indicates that none of the products consistently overestimate or underestimate precipitation across all sub-basins in the ARNC (Figure 7), although their bias patterns exhibit considerable consistency. ERA5-Land and MERRA underestimate precipitation in only one sub-basin but overestimate it in the remaining 24. Among the generally underestimating products, CMORPH, GLDAS, and IMERG underestimate precipitation in only one sub-basin each, whereas MSWEP and WorldClim underestimate precipitation in two and three sub-basins, respectively. CHIRPS and FLDAS show the largest spatial variability, underestimating precipitation in 7 and 14 sub-basins, respectively, while overestimating it elsewhere. After integrating results from all 25 sub-basins, the magnitude of overestimation follows the order MERRA > ERA5-Land > FLDAS, while the magnitude of underestimation follows IMERG > GLDAS > CMORPH > MSWEP > WorldClim > CHIRPS.

Further analysis using R^2 (Figure S6), RMAE (Figure S7), RRMSE (Figure 8), and *d* (Figure S8) reveals substantial spatial variation in product performance across the basins. CHIRPS and FLDAS perform best in the Tianshan region; CHIRPS, FLDAS, CMORPH, and WorldClim perform well in the Qilian Mountains; GLDAS performs best in the Karakoram region; all products show weaker performance in the Kunlun and Altun regions. Among the 25 sub-basins, FLDAS and CHIRPS achieve the best performance in 9 and 6 sub-basins, respectively, and neither product ranks worst in any basin. Conversely, MERRA performs best in only one sub-basin but ranks lowest in 11 sub-basins, and IMERG, while never ranking first, ranks worst in just two sub-basins (Figure 9). Considering all performance metrics, the overall ranking of precipitation products across the ARNC at the sub-basin scale is FLDAS > CHIRPS > WorldClim > MSWEP > GLDAS > IMERG > ERA5-Land > CMORPH > MERRA.

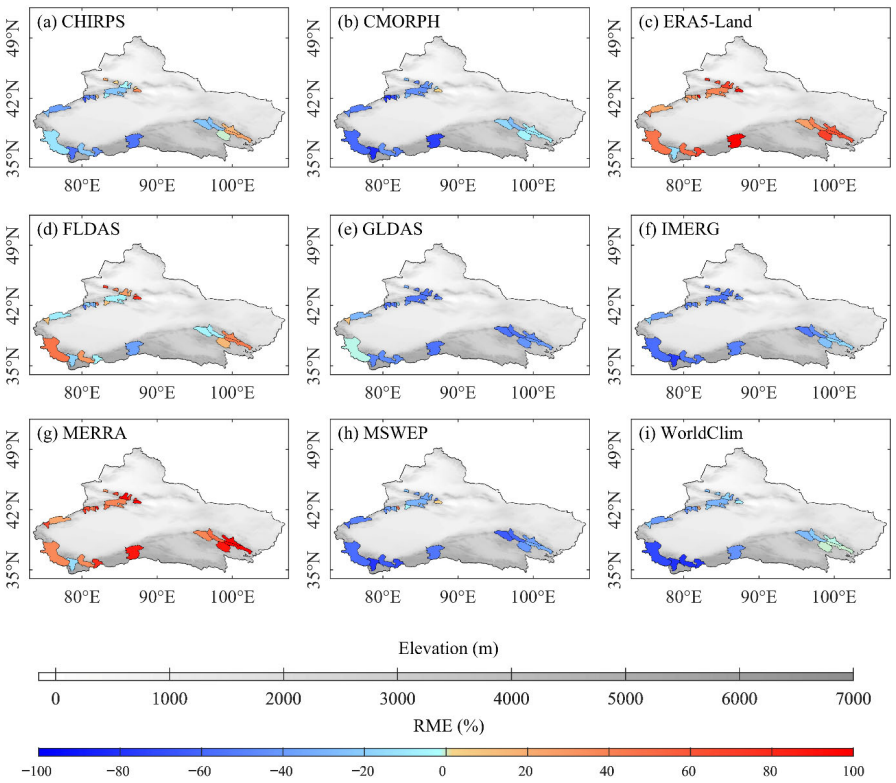

Figure 7. Spatial distribution of relative mean error (RME) between 9 products' data and mean annual precipitation calculated using the water balance method at 25 sub-basins in the arid region of Northwest China.

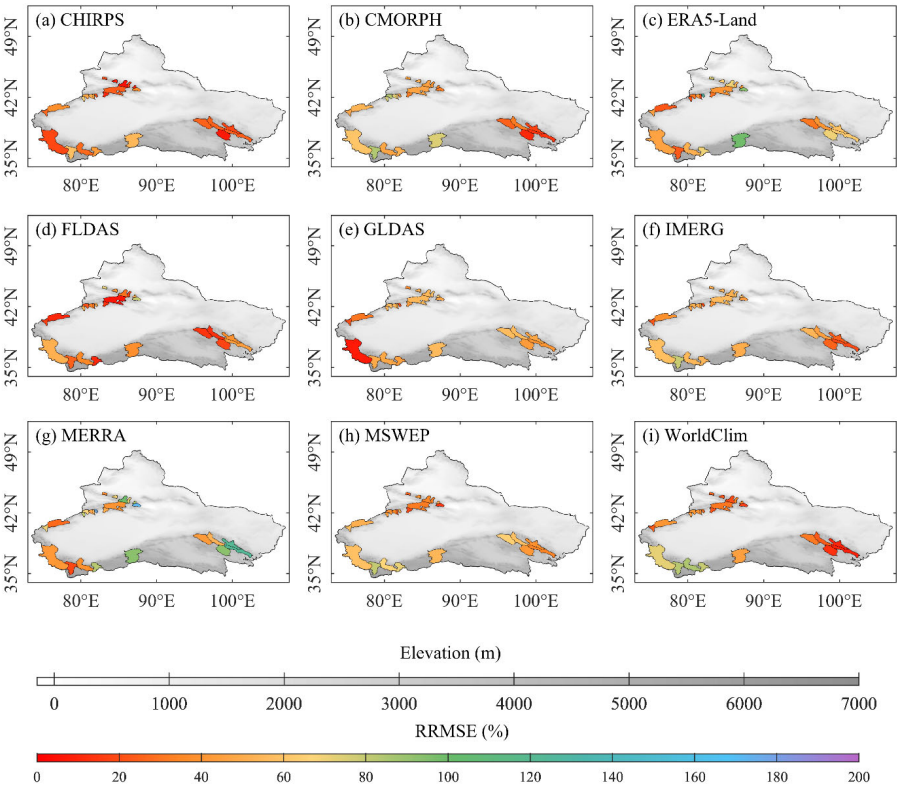

Figure 8. Spatial distribution of relative root mean square error (RRMSE) between 9 products' data and annual precipitation calculated using the water balance method at 25 sub-basins in the arid region of Northwest China.

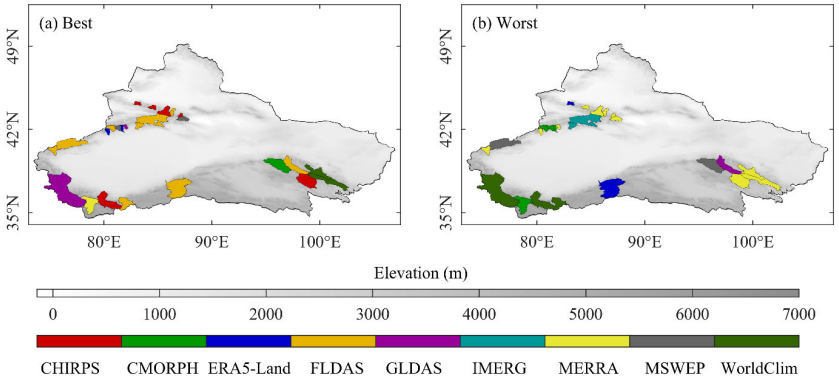

Figure 9. Best (**a**) and worst (**b**) performance of the precipitation products at 25 sub-basins in the arid region of Northwest China.

3.3.2. Performance of the Bayesian Model Averaging Method

Precipitation estimates obtained from the water balance method for 25 sub-basins were utilized to evaluate the areal performance of the BMA method across the ARNC. RME analysis shows that BMA-merged precipitation overestimates precipitation in 7 sub-basins and underestimates it in 18 sub-basins, with a mean RME of −8.6%. Consistent with the station-scale results, the BMA-merged precipitation demonstrates enhanced areal accuracy relative to the nine individual products. Compared with each product, the BMA-merged precipitation substantially reduces RRMSE in most sub-basins, with IMERG, MERRA and MSWEP showing the greatest improvement by reducing RRMSE in 24 sub-basins, and FLDAS showing the least improvement but still reducing RRMSE in 17 sub-basins (Figure 10).

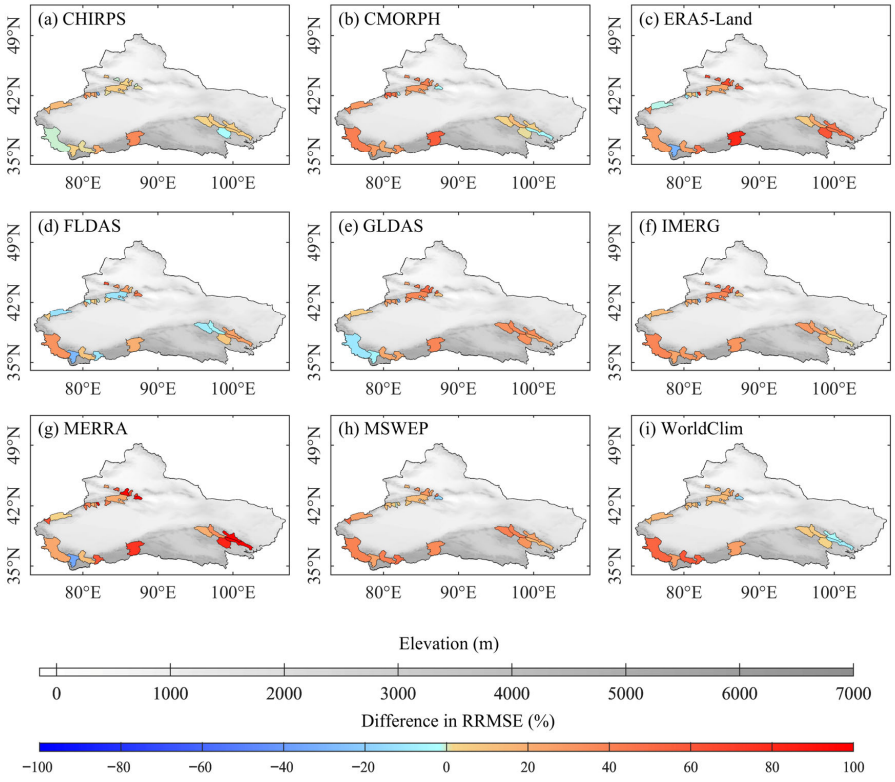

Figure 10. Spatial distribution of the differences in relative root mean square error (RRMSE) between the nine precipitation products and the merged precipitation from the Bayesian model averaging method, both evaluated against estimated precipitation at 25 sub-basins, in the arid region of Northwest China.

Evaluation metrics for the 25 sub-basins demonstrate that BMA-merged precipitation overall outperforms individual products. Higher median R^2 and d values, lower median RMAE and RRMSE values, and narrower interquartile ranges collectively indicate enhanced robustness and stability of the merged dataset (Figure 11). Figure S9 shows that the BMA dataset performed better than each individual product in a larger number of basins, with improvements ranging from 18 basins relative to FLDAS to 24 basins relative to IMERG and MERRA, based on the RS ranking results. Overall, the BMA-merged precipitation demonstrates significantly higher accuracy at the sub-basin scale across the ARNC compared to all single precipitation products.

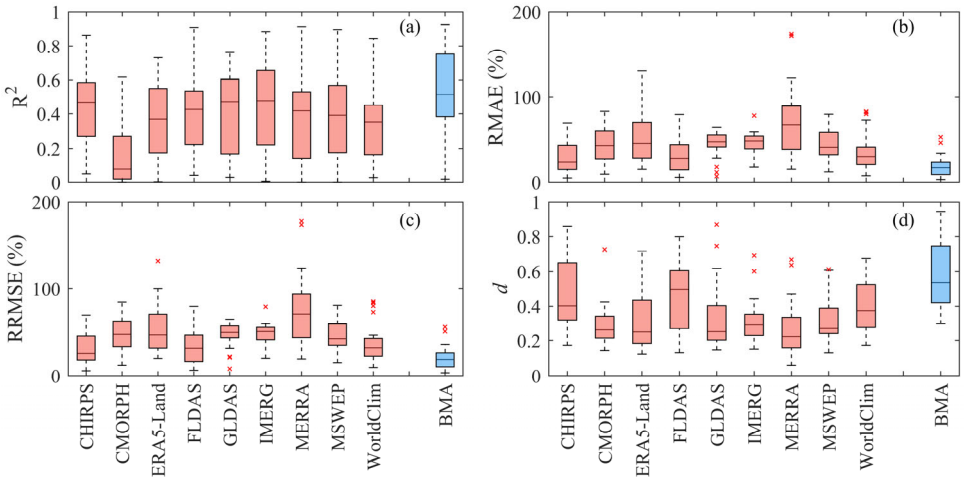

Figure 11. Box plots of R^2 (**a**), RMAE (**b**), RRMSE (**c**), and *d* (**d**) values at 25 sub-basins for 9 products and Bayesian model averaging method. The black line in each boxplots represents the median value, and the bottom and top edges of each box indicate the 25th and 75th percentiles, respectively. The whiskers represent the maximum and minimum values that are not considered outliers, where outliers are more than 1.5 standard deviations from the mean and are shown as red crosses.

3.4. Temporal and Spatial Variability in Precipitation in the Arid Region of Northwest China

Based on the evaluation results from both the meteorological station and sub-basin scales, the BMA-merged precipitation exhibits significantly higher accuracy than any single precipitation product. Therefore, the BMA-based dataset was used to analyze the temporal and spatial variability of precipitation across the ARNC from 2000 to 2024. The results indicate that the multi-year mean annual precipitation over the ARNC is approximately 230.4 mm. High-precipitation regions are primarily located in high-elevation mountainous areas such as the Qilian Mountains and the Tianshan Mountains, while low-precipitation regions are mainly distributed across low-elevation desert areas, including the Taklimakan Desert, the Kumtag Desert, and the Badain Jaran Desert (Figure 12a). Sen's slope analysis demonstrates that annual precipitation exhibited a predominantly increasing trend across most of the ARNC during 2000–2024, with the most pronounced positive trends occurring in the Qilian Mountains, where the increases were statistically significant at the 95% confidence level (Figure 12b). Only a few regions showed decreasing trends, primarily along the northern slopes of the Tianshan Mountains and within the Junggar Basin; however, none of these negative trends were statistically significant at the 95% confidence level (Figure 12c). Overall, annual precipitation in the ARNC increased significantly over the period 2000–2024, with a linear trend of 1.4 mm per year (Figure 12d), indicating an enhancement of water availability under ongoing climatic warming.

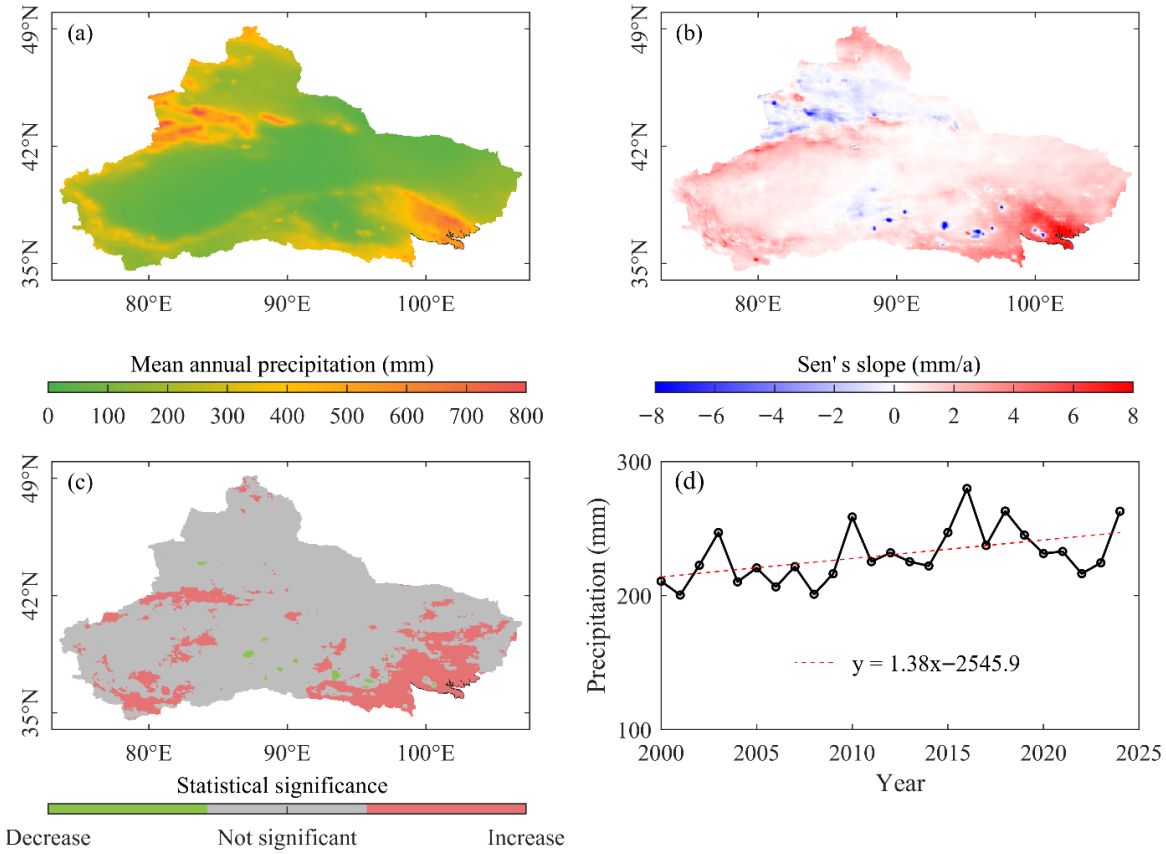

Figure 12. Spatial distribution of mean annual precipitation (**a**), Sen's slope of the annual precipitation (**b**), trends in annual precipitation based on the Mann–Kendall method (**c**) and inter-annual variation in the mean annual precipitation (**d**) in the arid region of Northwest China during the 2000–2024 period (the dotted line is the trend line).

4. Discussion

4.1. Differences Between Evaluation at Meteorological Station and Sub-Basin Scales

Evaluation of the nine precipitation products at meteorological station and sub-basin scales reveals notable scale-dependent discrepancies in performance. In the ARNC, IMERG and CMORPH were identified as the best- and worst-performing products at the station scale, whereas FLDAS and MERRA ranked highest and lowest at the sub-basin scale, respectively. Similar inconsistencies between scales have also been reported in the Qilian Mountains [25,77] and in global assessments [78]. These differences reflect the influence of scale effects: point-to-pixel validation is sensitive to representativeness errors, uneven gauge distribution, and steep topographical gradients, potentially obscuring the true regional performance of precipitation datasets [79,80]. The superior performance of IMERG at the point scale can be attributed to its incorporation of gauge observations within its satellite–gauge fusion algorithm, which enhances accuracy at individual station locations. In contrast, FLDAS emphasizes physically consistent water and energy balances at the basin scale. As a result, it provides more reliable precipitation estimates at sub-basin or regional scales, even if its point-scale accuracy is comparatively lower. Sub-basin evaluation relies on spatially aggregated precipitation, effectively reducing random uncertainties and providing more stable performance patterns. As demonstrated in Figure 7, several products display systematic regional biases, while their station-scale errors exhibit greater dispersion (Figure 3). These findings underscore the need to align validation strategies with application contexts and to adopt multi-scale evaluation frameworks when selecting precipitation datasets for hydrological applications.

4.2. Advantages of Bayesian Model Averaging for Precipitation Merging

Compared with individual precipitation products, the BMA-merged precipitation significantly improves estimation accuracy at both meteorological station and sub-basin scales in the ARNC. This improvement results from the probabilistic integration of model uncertainties and the optimization of weighting based on observational evidence rather than subjective assumptions [25,37]. Numerous studies have demonstrated the effectiveness of BMA in enhancing hydrometeorological predictions such as runoff forecasting, drought monitoring, and evapotranspiration simulation [38–40]. The advantages of BMA in precipitation merging have also been confirmed by Yang et al. [25] in the Qilian Mountains. Consistent with previous findings, our results show that BMA substantially reduces RRMSE relative to all individual products (Figures 5 and 10), and other evaluation criteria consistently demonstrate superior performance (Figures 6 and 11). Moreover, the method exhibits remarkable robustness across sub-basins with diverse hydro-climatic characteristics. Furthermore, the method exhibits notable robustness across diverse hydro-climatic environments. Therefore, BMA provides an effective strategy for producing reliable precipitation datasets in sparsely monitored and topographically complex arid regions, supporting hydrological modeling and climate assessment.

4.3. Uncertainties and Limitations

Despite the demonstrated improvements of the BMA-merged dataset, several uncertainties and limitations must be acknowledged. First, the accuracy of BMA is strongly dependent on the quality and spatial representativeness of gauge observations. In the ARNC, meteorological stations are predominantly concentrated in oasis and low-elevation areas, whereas high-elevation mountain regions and desert interiors remain sparsely monitored (Figure 1), introducing uncertainty in weight estimation. Second, structural differences among precipitation products—including spatial resolution, sensor characteristics, and retrieval algorithms—result in systematic biases that cannot be fully eliminated through statistical fusion techniques [81,82]. Third, basin-scale validation using the water balance method relies on evapotranspiration, TWS changes, and runoff data, which themselves contain uncertainties. In cryosphere-dominated basins in the ARNC influenced by glaciers, snow, and permafrost, hydrological processes are more complex [83,84], potentially affecting diagnostic accuracy. In addition, the basin-scale water balance validation is constrained by the availability of observed runoff data, which in most sub-basins ends around 2011. As a result, the hydrological validation does not fully cover the entire period (2000–2024) for which the final BMA-merged precipitation dataset and its long-term trends are presented. The extension of the merged precipitation product beyond the runoff validation period implicitly assumes that the relative performance of individual precipitation products remains temporally stable, and the trend analysis should therefore be interpreted with due consideration of this limitation. Furthermore, the BMA-merged precipitation exhibits a positive mean RME of +34.0%, indicating a systematic overestimation. This overestimation can be primarily attributed to the input datasets themselves: except for MSWEP and WorldClim, the remaining seven products tend to overestimate precipitation at most stations (Figure 3). Although the BMA-merged product improves correlation and spatial consistency compared with individual datasets, this positive bias should be considered when using absolute precipitation amounts for hydrological or climate applications. Future improvements should focus on enhancing high-elevation observations, adopting systematic uncertainty quantification frameworks, and developing fusion strategies tailored to coupled cryosphere–hydrology systems.

5. Conclusions

Based on a comprehensive evaluation of 9 precipitation products across the ARNC, this study employed the BMA method to generate a high-accuracy merged precipitation dataset. The main conclusions are as follows:

(1) Substantial spatial heterogeneity and uncertainty exist among all precipitation products, and no single dataset consistently performs best or worst at both meteorological-station and sub-basin scales. Among the 152 meteorological stations, IMERG, MSWEP, and ERA5-Land show the best performance, whereas at the sub-basin scale, FLDAS, CHIRPS, and WorldClim rank highest.

(2) The BMA-merged precipitation significantly improves estimation accuracy relative to all nine individual products. At both station and sub-basin scales, the merged dataset demonstrates higher R^2 and d values, lower RMAE and RRMSE values, and strong robustness and stability across diverse hydro-climatic environments.

(3) The BMA-merged dataset indicates that the mean annual precipitation in the ARNC during 2000–2024 is approximately 230.4 mm, exhibiting a statistically significant upward trend of 1.4 mm per year. The most pronounced increases are concentrated in high-elevation mountainous regions, including the Tianshan and Qilian Mountains.

Overall, the BMA-merged precipitation dataset provides a reliable foundation for hydrological modeling, cryosphere–hydrology interaction studies, and research on climate-change impacts and adaptation in arid regions. Continued advancements in data-fusion techniques and increased availability of high-resolution ground and remote-sensing observations will further enhance precipitation estimation accuracy in complex terrain such as the ARNC.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/atmos17010094/s1, Figure S1. Interannual variability in the area-averaged annual precipitation in the arid region of Northwest China for 2000–2024 determined from the data of 9 products; Figure S2. Spatial distribution of R^2 (coefficient of determination) between 9 products' data and observational data obtained at 152 meteorological stations in the arid region of Northwest China; Figure S3. Spatial distribution of RMAE (relative mean absolute error) between 9 products' data and observational data obtained at 152 meteorological stations in the arid region of Northwest China; Figure S4. Spatial distribution of d (index of agreement) between 9 products' data and observational data obtained at 152 meteorological stations in the arid region of Northwest China; Figure S5. Comparison of the performance between the merged precipitation from the Bayesian model averaging method and the 9 precipitation products at 152 meteorological stations in the arid region of Northwest China; Figure S6. Spatial distribution of R^2 (coefficient of determination) between 9 products' data and annual precipitation calculated using the water balance method at 25 sub-basins in the arid region of Northwest China; Figure S7. Spatial distribution of RMAE (relative mean absolute error) between 9 products' data and annual precipitation calculated using the water balance method at 25 sub-basins in the arid region of Northwest China; Figure S8. Spatial distribution of d (index of agreement) between 9 products' data and annual precipitation calculated using the water balance method at 25 sub-basins in the arid region of Northwest China; Figure S9. Comparison of the performance between the merged precipitation from the Bayesian model averaging method and the 9 precipitation products at 25 sub-basins in the arid region of Northwest China.

Author Contributions: Conceptualization, Y.Y. and R.C.; methodology, Y.Y., R.C. and X.L.; software, Y.Y., W.M. and Z.L.; validation, Y.Y. and X.W.; resources, X.L. and W.M.; data curation, Y.Y. and Z.L.; writing—original draft preparation, Y.Y.; visualization, Y.Y. and X.W.; supervision, Y.Y. and R.C.; funding acquisition, Y.Y. and R.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Desert Meteorological Science Research Foundation (Sqj2023018), the Gansu Provincial Science and Technology Planning Project (24ZD13FA004), the National Key Research and Development Project (2024YFF0808602), and the Natural Science Foundation of Gansu Province (24JRRA708).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The grid datasets used in this work are publicly available as detailed in Table 2. The precipitation observation data collected from meteorological stations were obtained from the China Meteorological Administration (http://data.cma.cn/, accessed on 11 July 2022). The runoff data from hydrological gauging stations were obtained from the Hydrographic Yearbooks of the People's Republic of China. The precipitation data in the arid region of Northwest China generated by the Bayesian model averaging method are available from the National Cryosphere Desert Data Center (http://www.ncdc.ac.cn/).

Acknowledgments: We sincerely thank the precipitation dataset producers used in this research.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Shen, Y.; Chen, Y. Global perspective on hydrology, water balance, and water resources management in arid basins. *Hydrol. Process.* **2009**, *24*, 129–135. [CrossRef]
2. Zuo, M.; Zhou, T.; Man, W. Wetter Global Arid Regions Driven by Volcanic Eruptions. *J. Geophys. Res. Atmos.* **2019**, *124*, 13648–13662. [CrossRef]
3. Lewin, A.; Murali, G.; Rachmilevitch, S.; Roll, U. Global evaluation of current and future threats to drylands and their vertebrate biodiversity. *Nat. Ecol. Evol.* **2024**, *8*, 1448–1458. [CrossRef] [PubMed]
4. Lu, J.; Jia, L.; Menenti, M.; Zheng, C.; Hu, G.; Ji, D. The impacts of drought on water availability: Spatial and temporal analysis in the Belt and Road region (2001–2020). *Int. J. Climatol.* **2025**, *18*, 2449706. [CrossRef]
5. Ershadfath, F.; Shahnazari, A.; Sarjaz, M.R.; Andaryani, S.; Trolle, D.; Olesen, J.E. Blue and green water availability under climate change in arid and semi-arid regions. *Ecol. Inform.* **2024**, *82*, 102743. [CrossRef]
6. Wang, L.; Collins, S.L. The complex relationship between precipitation and productivity in drylands. *Camb. Prism. Drylands* **2024**, *1*, e1. [CrossRef]
7. Sang, L.; Zhu, G.; Qiu, D.; Zhang, Z.; Liu, Y.; Zhao, K.; Wang, L.; Sun, Z. Spatial variability of runoff recharge sources and influence mechanisms in an arid mountain flow-producing zone. *Hydrol. Process.* **2022**, *36*, e14642. [CrossRef]
8. Zhang, Y.; Gentine, P.; Luo, X.; Lian, X.; Liu, Y.; Zhou, S.; Michalak, A.M.; Sun, W.; Fisher, J.B.; Piao, S.; et al. Increasing sensitivity of dryland vegetation greenness to precipitation due to rising atmospheric CO_2. *Nat. Commun.* **2022**, *13*, 4875. [CrossRef]
9. Xu, L.; Zheng, C.; Ma, Y. Variations in precipitation extremes in the arid and semi-arid regions of China. *Int. J. Climatol.* **2020**, *41*, 1542–1554. [CrossRef]
10. Saidaliyeva, Z.; Shahgedanova, M.; Yapiyev, V.; Wade, A.J.; Akbarov, F.; Esenaman uulu, M.; Kalashnikova, O.; Kapitsa, V.; Kasatkin, N.; Rakhimov, I.; et al. Precipitation in the mountains of Central Asia: Isotopic composition and source regions. *Atmos. Chem. Phys.* **2024**, *24*, 12203–12224. [CrossRef]
11. Chakri, A.; Abakarim, S.; Rodrigues, J.C.A.; Laftouhi, N.-E.; Ibouh, H.; Zouhri, L.; Zaitseva, E. Spatial Bias Correction of ERA5_Ag Reanalysis Precipitation Using Machine Learning Models in Semi-Arid Region of Morocco. *Atmosphere* **2025**, *16*, 1234. [CrossRef]
12. Elahi, E.; Abro, M.I.; Khaskheli, M.A.; Kandhro, G.A.; Zehra, T.; Ali, S.; Shaikh, M.N.; Laghari, B.A.; Hassan, M.; Memon, M.A. Long-term evaluation of rainfall in the arid region of Pakistan using multi-source data. *Theor. Appl. Climatol.* **2023**, *155*, 2819–2840. [CrossRef]
13. Han, J.; Miao, C.; Gou, J.; Zheng, H.; Zhang, Q.; Guo, X. A new daily gridded precipitation dataset for the Chinese mainland based on gauge observations. *Earth Syst. Sci. Data* **2023**, *15*, 3147–3161. [CrossRef]
14. Zhao, C.; Yao, S.; Ding, Y.; Zhao, Q. A Gridded Monthly Precipitation Merged Rain Gauge and Satellite Analysis Dataset for the Tian Shan Range between 1981 and 2019. *J. Appl. Meteorol. Clim.* **2023**, *62*, 691–708. [CrossRef]
15. Dou, Y.; Shi, K.; Cai, H.; Xie, M.; Liu, R. Review of Utilisation Methods of Multi-Source Precipitation Products for Flood Forecasting in Areas with Insufficient Rainfall Gauges. *Atmosphere* **2025**, *16*, 835. [CrossRef]
16. Liu, Y.; Zheng, Y.; Li, W.; Zhou, T. Evaluating the Performance of Satellite-Based Precipitation Products Using Gauge Measurement and Hydrological Modeling: A Case Study in a Dry Basin of Northwest China. *J. Hydrometeorol.* **2022**, *23*, 541–559. [CrossRef]

17. Miao, C.; Gou, J.; Hu, J.; Duan, Q. Impacts of Different Satellite-Based Precipitation Signature Errors on Hydrological Modeling Performance Across China. *Earth's Future* **2024**, *12*, e2024EF004954. [CrossRef]
18. Ferguglia, O.; Palazzi, E.; Arnone, E. Elevation dependent change in ERA5 precipitation and its extremes. *Clim. Dyn.* **2024**, *62*, 8137–8153. [CrossRef]
19. Li, Y.; Qin, X.; Liu, Y.; Jin, Z.; Liu, J.; Wang, L.; Chen, J. Evaluation of Long-Term and High-Resolution Gridded Precipitation and Temperature Products in the Qilian Mountains, Qinghai–Tibet Plateau. *Front. Environ. Sci.* **2022**, *10*, 906821. [CrossRef]
20. Gebrechorkos, S.H.; Leyland, J.; Dadson, S.J.; Cohen, S.; Slater, L.; Wortmann, M.; Ashworth, P.J.; Bennett, G.L.; Boothroyd, R.; Cloke, H.; et al. Global-scale evaluation of precipitation datasets for hydrological modelling. *Hydrol. Earth Syst. Sci.* **2024**, *28*, 3099–3118. [CrossRef]
21. Islam, M.A.; Yu, B.; Cartwright, N. Assessment and comparison of five satellite precipitation products in Australia. *J. Hydrol.* **2020**, *590*, 125474. [CrossRef]
22. Dehaghani, A.M.; Gohari, A.; Zareian, M.J.; Torabi Haghighi, A. A comprehensive evaluation of the satellite precipitation products across Iran. *J. Hydrol. Reg.* **2023**, *46*, 101360. [CrossRef]
23. Mahmoud, M.T.; Al-Zahrani, M.A.; Sharif, H.O. Assessment of global precipitation measurement satellite products over Saudi Arabia. *J. Hydrol.* **2018**, *559*, 1–12. [CrossRef]
24. Wang, X.; Xu, M.; Kang, S.; Li, X.; Han, H.; Li, X. Comprehensive applicability evaluation of four precipitation products at multiple spatiotemporal scales in Northwest China. *J. Arid Land* **2024**, *16*, 1232–1254. [CrossRef]
25. Yang, Y.; Chen, R.; Ding, Y.; Qing, W.; Li, H.; Han, C.; Liu, Z.; Liu, J. Evaluation of 12 precipitation products and comparison of 8 multi-model averaging methods for estimating precipitation in the Qilian Mountains, Northwest China. *Atmos. Res.* **2023**, *296*, 107075. [CrossRef]
26. Jia, Y.; Lei, H.; Yang, H.; Hu, Q. Terrestrial Water Storage Change Retrieved by GRACE and Its Implication in the Tibetan Plateau: Estimating Areal Precipitation in Ungauged Region. *Remote Sens.* **2020**, *12*, 3129. [CrossRef]
27. Koukoula, M.; Nikolopoulos, E.I.; Dokou, Z.; Anagnostou, E.N. Evaluation of Global Water Resources Reanalysis Products in the Upper Blue Nile River Basin. *J. Hydrometeorol.* **2020**, *21*, 935–952. [CrossRef]
28. Wang, Y.; Zhao, N. Evaluation of Eight High-Resolution Gridded Precipitation Products in the Heihe River Basin, Northwest China. *Remote Sens.* **2022**, *14*, 1458. [CrossRef]
29. Ajami, N.K.; Duan, Q.; Gao, X.; Sorooshian, S. Multimodel Combination Techniques for Analysis of Hydrological Simulations: Application to Distributed Model Intercomparison Project Results. *J. Hydrometeorol.* **2006**, *7*, 755–768. [CrossRef]
30. Moges, E.; Demissie, Y.; Larsen, L.; Yassin, F. Review: Sources of Hydrological Model Uncertainties and Advances in Their Analysis. *Water* **2020**, *13*, 28. [CrossRef]
31. Akaike, H. ANew Look at the Statistical Model Identification. *IEEE Trans. Automat. Contr.* **1974**, *19*, 716–723. [CrossRef]
32. Schwarz, G. Estimating the dimension of a model. *Ann. Statist.* **1978**, *6*, 461–464. [CrossRef]
33. Bates, J.M.; Granger, C.W.J. The Combination of Forecasts. *J. Oper. Res. Soc.* **1969**, *20*, 451–468. [CrossRef]
34. Granger, C.W.J.; Ramanathan, R. Improved methods of combining forecasts. *J. Forecast.* **1984**, *3*, 197–204. [CrossRef]
35. Chmielecki, R.M.; Raftery, A.E. Probabilistic Visibility Forecasting Using Bayesian Model Averaging. *Mon. Weather Rev.* **2011**, *139*, 1626–1636. [CrossRef]
36. Neuman, S.P. Maximum likelihood Bayesian averaging of uncertain model predictions. *Stoch. Environ. Res. Risk Assess.* **2003**, *17*, 291–305. [CrossRef]
37. Duan, Q.; Ajami, N.K.; Gao, X.; Sorooshian, S. Multi-model ensemble hydrologic prediction using Bayesian model averaging. *Adv. Water Resour.* **2007**, *30*, 1371–1386. [CrossRef]
38. Seifi, A.; Ehteram, M.; Soroush, F.; Torabi Haghighi, A. Multi-model ensemble prediction of pan evaporation based on the Copula Bayesian Model Averaging approach. *Eng. Appl. Artif. Intel.* **2022**, *114*, 105124. [CrossRef]
39. Wu, H.; Su, X.; Singh, V.P.; Zhang, T. Predicting Hydrological Drought With Bayesian Model Averaging Ensemble Vine Copula (BMAViC) Model. *Water Resour. Res.* **2022**, *58*, e2022WR033146. [CrossRef]
40. Ning, S.; Cheng, Y.; Zhou, Y.; Wang, J.; Zhang, Y.; Jin, J.; Thapa, B.R. Bayesian Model Averaging for Satellite Precipitation Data Fusion: From Accuracy Estimation to Runoff Simulation. *Remote Sens.* **2025**, *17*, 1154. [CrossRef]
41. Duan, K.; Wang, X.; Liu, B.; Zhao, T.; Chen, X. Comparing Bayesian Model Averaging and Reliability Ensemble Averaging in Post-Processing Runoff Projections under Climate Change. *Water* **2021**, *13*, 2124. [CrossRef]
42. Dusa, S.; Manikanta, V.; Das, J.; Umamahesh, N.V. Does the performance enhancement through multi-model averaging at the catchment outlet gets translated to the interior ungauged points? *J. Hydrol.* **2023**, *627*, 130389. [CrossRef]
43. Yang, X.; Yuan, F.; Liu, S.; Xu, C.-Y.; Ren, L.; Jiang, S. Quantifying multi-source uncertainties in multi-model predictions using the Bayesian model averaging scheme. *Hydrol. Res.* **2018**, *49*, 954–970.
44. Yang, Z.; Chen, R.; Liu, Z.; Zhao, Y.; Liu, Y.; Wu, W. Spatiotemporal variability of rain-on-snow events in the arid region of Northwest China. *J. Arid Land* **2024**, *16*, 483–499. [CrossRef]

45. Yang, Y.; Shen, L.; Wang, B. How is the precipitation distributed vertically in arid mountain region of Northwest China? *J. Geogr. Sci.* **2022**, *32*, 241–258. [CrossRef]
46. Weng, J.; Yand, Y.; Mu, Z.; Yang, L. Mechanisms and spatiotemporal variation characteristics of flooding over northwestern China. *Acta Geogr. Sin.* **2024**, *79*, 2768–2779. (In Chinese)
47. Chen, Y.; Li, B.; Li, Z.; Li, W. Water resource formation and conversion and water security in arid region of Northwest China. *J. Geogr. Sci.* **2016**, *26*, 939–952. [CrossRef]
48. China Meteorological Administration. *Specifications for Surface Meteorological Observation*; China Meteorological Press: Beijing, China, 2003.
49. Chen, F.; Li, X. Evaluation of IMERG and TRMM 3B43 Monthly Precipitation Products over Mainland China. *Remote Sens.* **2016**, *8*, 472. [CrossRef]
50. Qian, L.; Zhao, P. Assessment of ERA5-Land Reanalysis Precipitation Data in the Qilian Mountains of China. *Atmosphere* **2025**, *16*, 826. [CrossRef]
51. Wei, D.; Di, W.; Tian, W.; Cheng, S.; Xie, H.; Xie, L.; Jing, Z. Assessment of Integrated Multi-Satellite Retrievals for Global Precipitation Measurement (IMERG) Precipitation Products in Northwest China. *Remote Sens.* **2025**, *17*, 1364. [CrossRef]
52. Funk, C.; Peterson, P.; Landsfeld, M.; Pedreros, D.; Verdin, J.; Shukla, S.; Husak, G.; Rowland, J.; Harrison, L.; Hoell, A.; et al. The climate hazards infrared precipitation with stations--a new environmental record for monitoring extremes. *Sci. Data* **2015**, *2*, 150066. [CrossRef] [PubMed]
53. Xie, P.; Joyce, R.; Wu, S.; Yoo, S.; Yarosh, Y.; Sun, F.; Lin, R. Reprocessed, Bias-Corrected CMORPH Global High-Resolution Precipitation Estimates from 1998. *J. Hydrometeorol.* **2017**, *18*, 1617–1641. [CrossRef]
54. Muñoz-Sabater, J.; Dutra, E.; Agustí-Panareda, A.; Albergel, C.; Arduini, G.; Balsamo, G.; Boussetta, S.; Choulga, M.; Harrigan, S.; Hersbach, H.; et al. ERA5-Land: A state-of-the-art global reanalysis dataset for land applications. *Earth Syst. Sci. Data* **2021**, *13*, 4349–4383. [CrossRef]
55. McNally, A.; Jacob, J.; Arsenault, K.; Slinski, K.; Sarmiento, D.P.; Hoell, A.; Pervez, S.; Rowland, J.; Budde, M.; Kumar, S.; et al. A Central Asia hydrologic monitoring dataset for food and water security applications in Afghanistan. *Earth Syst. Sci. Data* **2022**, *14*, 3115–3135. [CrossRef]
56. Rui, H.; Beaudoing, H. *README Document for NASA GLDAS Version 2 Data Products, National Aeronautics and Space Administration Goddard Earth Science Data Information Center (GES DISC)*; NASA: Greenbelt, MD, USA, 2022.
57. Huffman, G.J.; Stocker, E.F.; Bolvin, D.T.; Nelkin, E.J.; Tan, J. *GPM IMERG Final Precipitation L3 1 Month 0.1 Degree × 0.1 Degree V07*; Goddard Earth Sciences Data and Information Services Center (GES DISC); NASA: Greenbelt, MD, USA, 2023.
58. Reichle, R.H.; Liu, Q.; Koster, R.D.; Draper, C.S.; Mahanama, S.P.P.; Partyka, G.S. Land Surface Precipitation in MERRA-2. *J. Clim.* **2017**, *30*, 1643–1664. [CrossRef]
59. Beck, H.E.; Wood, E.F.; Pan, M.; Fisher, C.K.; Miralles, D.G.; van Dijk, A.I.J.M.; McVicar, T.R.; Adler, R.F. MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bull. Amer. Meteor. Soc.* **2019**, *100*, 473–500. [CrossRef]
60. Fick, S.E.; Hijmans, R.J. WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *Int. J. Climatol.* **2017**, *37*, 4302–4315. [CrossRef]
61. Martens, B.; Miralles, D.G.; Lievens, H.; van der Schalie, R.; de Jeu, R.A.M.; Fernández-Prieto, D.; Beck, H.E.; Dorigo, W.A.; Verhoest, N.E.C. GLEAM v3: Satellite-based land evaporation and root-zone soil moisture. *Geosci. Model Dev.* **2017**, *10*, 1903–1925. [CrossRef]
62. Jahromi, M.N.; Miralles, D.; Koppa, A.; Rains, D.; Zand-Parsa, S.; Mosaffa, H.; Jamshidi, S. Ten Years of GLEAM: A Review of Scientific Advances and Applications. In *Computational Intelligence for Water and Environmental Sciences*; Bozorg-Haddad, O., Zolghadr-Asli, B., Eds.; Springer Nature: Singapore, 2022; pp. 525–540.
63. Michel, D.; Jiménez, C.; Miralles, D.G.; Jung, M.; Hirschi, M.; Ershadi, A.; Martens, B.; McCabe, M.F.; Fisher, J.B.; Mu, Q.; et al. The WACMOS-ET project—Part 1: Tower-scale evaluation of four remote-sensing-based evapotranspiration algorithms. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 803–822. [CrossRef]
64. Miralles, D.G.; Bonte, O.; Koppa, A.; Baez-Villanueva, O.M.; Tronquo, E.; Zhong, F.; Beck, H.E.; Hulsman, P.; Dorigo, W.; Verhoest, N.E.C.; et al. GLEAM4: Global land evaporation and soil moisture dataset at 0.1 resolution from 1980 to near present. *Sci. Data* **2025**, *12*, 416. [CrossRef]
65. Zhao, Y.; Hoeltgebaum, L.E.B.; Kukal, M.S.; Zhao, M. Leading Satellite-Based Evapotranspiration Products Insufficiently Capture Interannual Variability: Evidence from GRACE/FO and In Situ Observations. *Geophys. Res. Lett.* **2025**, *52*, e2025GL116784. [CrossRef]
66. Landerer, F.W.; Flechtner, F.M.; Save, H.; Webb, F.H.; Bandikova, T.; Bertiger, W.I.; Bettadpur, S.V.; Byun, S.H.; Dahle, C.; Dobslaw, H.; et al. Extending the Global Mass Change Data Record: GRACE Follow-On Instrument and Science Data Performance. *Geophys. Res. Lett.* **2020**, *47*, e2020GL088306. [CrossRef]

67. Luo, Z.; Fu, H.; Shao, Q.; Dong, W.; Chen, X.; Ding, X.; Wang, L.; Gu, X.; Sarukkalige, R.; Huang, H.; et al. A novel method for correcting water budget components and reducing their uncertainties by optimally distributing the imbalance residual without full closure. *Hydrol. Earth Syst. Sci.* **2025**, *29*, 4607–4635. [CrossRef]
68. Sattar, H.; Kinouchi, T. Streamflow estimation in the Indus River Basin using a water balance framework and artificial neural networks with satellite- and model-derived global hydro-climatic data sources. *J. Hydrol. Reg. Stud.* **2025**, *60*, 102510. [CrossRef]
69. Zhao, H.; Ma, J.; Zhang, L.; Zeng, H.; Huang, Q. Human activities have exacerbated groundwater depletion in arid regions: A case study of the Hexi Corridor, China. *J. Hydrol. Reg. Stud.* **2025**, *61*, 102649. [CrossRef]
70. Ouatiki, H.; Boudhar, A.; Leblanc, M.; Fakir, Y.; Chehbouni, A. When climate variability partly compensates for groundwater depletion: An analysis of the GRACE signal in Morocco. *J. Hydrol. Reg. Stud.* **2022**, *42*, 101177. [CrossRef]
71. Chen, A.; Chen, D.; Azorin-Molina, C. Assessing reliability of precipitation data over the Mekong River Basin: A comparison of ground-based, satellite, and reanalysis datasets. *Int. J. Climatol.* **2018**, *38*, 4314–4334. [CrossRef]
72. Baran, S.; Hemri, S.; El Ayari, M. Statistical Postprocessing of Water Level Forecasts Using Bayesian Model Averaging With Doubly Truncated Normal Components. *Water Resour. Res.* **2019**, *55*, 3997–4013. [CrossRef]
73. Kendall, M.G. *Rank Correlation Methods*, 4th ed.; Charles Griffin: London, UK, 1975.
74. Mann, H.B. Nonparametric tests against trend. *Econometrica* **1945**, *13*, 245–259. [CrossRef]
75. Gavrilov, M.B.; Marković, S.B.; Janc, N.; Nikolić, M.; Valjarević, A.; Komac, B.; ZorN, M.; PuNišić, M.; Bačević, N. Assessing average annual air temperature trends using the Mann–Kendall test in Kosovo. *Acta Geogr. Slov.* **2018**, *58*, 7–25. [CrossRef]
76. Sen, P.K. Estimates of the regression coefficient based on Kendall's tau. *J. Am. Stat. Assoc.* **1968**, *63*, 1379–1389. [CrossRef]
77. Li, Y.; Yang, M.; Wan, G.; Man, Y.; Wang, X. Coupled statistical analysis and hydrological simulation to evaluate the performance of satellite and reanalysis precipitation products in the Qilian Mountains, Northwest China. *Clim. Dyn.* **2025**, *63*, 134. [CrossRef]
78. Beck, H.E.; Vergopolan, N.; Pan, M.; Levizzani, V.; van Dijk, A.I.J.M.; Weedon, G.P.; Brocca, L.; Pappenberger, F.; Huffman, G.J.; Wood, E.F. Global-scale evaluation of 22 precipitation datasets using gauge observations and hydrological modeling. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 6201–6217. [CrossRef]
79. Mao, R.; Wang, L.; Zhou, J.; Li, X.; Qi, J.; Zhang, X. Evaluation of Various Precipitation Products Using Ground-Based Discharge Observation at the Nujiang River Basin, China. *Water* **2019**, *11*, 2308. [CrossRef]
80. Behrangi, A.; Gardner, A.S.; Reager, J.T.; Fisher, J.B. Using GRACE to constrain precipitation amount over cold mountainous basins. *Geophys. Res. Lett.* **2017**, *44*, 219–227. [CrossRef]
81. Nourani, V.; Tosan, M.; Huang, J.J.; Gebremichael, M.; Kantoush, S.A.; Dastourani, M. Advances in multi-source data fusion for precipitation estimation: Remote sensing and machine learning perspectives. *Earth-Sci. Rev.* **2025**, *270*, 105253. [CrossRef]
82. Zhang, H.; Wei, L.; Zhu, Y.; Zhou, J.; Liu, S.; Sun, X.; Kang, X.; Gao, M.; Duan, Z.; Crow, W.T.; et al. Comparison of multi-source merged precipitation products using independent gauge observations. *Atmos. Res.* **2026**, *328*, 108427. [CrossRef]
83. Chen, Y.; Li, W.; Fang, G.; Li, Z. Review article: Hydrological modeling in glacierized catchments of central Asia—Status and challenges. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 669–684. [CrossRef]
84. Cheng, W.; Feng, Q.; Xi, H.; Yin, X.; Cheng, L.; Sindikubwabo, C.; Zhang, B.; Chen, Y.; Zhao, X. Modeling and assessing the impacts of climate change on groundwater recharge in endorheic basins of Northwest China. *Sci. Total Environ.* **2024**, *918*, 170829. [CrossRef]